THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 1 Capitole

Présentée et soutenue par

TIM EDERER

Le 26 juin 2023

Essais en Economie du Travail: Modèles Empiriques d'Appariement et Allocation Spatiale des Enseignants

Ecole doctorale : **TSE - Toulouse Sciences Economiques**

Spécialité : **Sciences Economiques - Toulouse**

Unité de recherche :
TSE-R - Toulouse School of Economics - Recherche

Thèse dirigée par
Matteo BOBBA et Thierry MAGNAC

Jury

Mme Caterina CALSAMIGLIA, Rapporteure
M. YingHua HE, Rapporteur
Mme Gabrielle FACK, Examinatrice
M. Christopher NEILSON, Examinateur
M. Matteo BOBBA, Directeur de thèse
M. Thierry MAGNAC, Directeur de thèse

Essays in Labor Economics: Empirical Models of Two-Sided Matching and Teacher Spatial Sorting

Tim Ederer

May 3, 2023

Acknowledgement

I sincerely thank Matteo Bobba, Thierry Magnac and Cristina Gualdani for providing continuous guidance throughout the development of this thesis and for encouraging me to walk off the beaten path. I am especially indebted to Matteo for not only being an inspiring advisor and co-author, but also a friend who always managed to be generous with his time when I needed it the most. More generally, I am extremely grateful for the support of TSE faculty and the Econometrics and Empirical Economics group who continuously provided extensive feedback on my work and helped me shape the different chapters of this thesis.

I also thank my family Mélyssa, Yanna, Yannick and Yasmina for keeping me grounded and supporting me from the very beginning, as well as my fellow PhD students and friends Gökçe, Guillem, Hippolyte, José Alfonso, Juan, Léa, Lisa, Max, Moritz, Peter, Tanja, Tuuli, Valentina and Wenxuan whom filled the past six years with memories that I will always cherish. Finally, I want to thank Anaïs Fabre for always being by my side and for being such a brilliant economist. This thesis would not have come to light without your unfailing support, your continuous input, and your invaluable help. I hope to continue learning from you as much as I have for the past nine years.

Overview

This thesis is composed of three chapters that have two main objectives: (i) broaden our understanding of the empirical content of two-sided matching models by developing novel methodological tools and (ii) apply these tools to get a thorough understanding of the causes and consequences of spatial inequalities in access to skilled teachers and provide concrete policy recommendations.

The research methods used in this thesis build on the empirical industrial organization literature by studying how data on observed choices can be used to recover the fundamental primitives governing agents' decisions. More specifically, I study how labor supply and demand map into equilibrium sorting on the labor market through the lens of two-sided matching models. I then explore the empirical content of these models and investigate whether the preferences of participating agents can be identified and estimated from data on observed matches. When the data permits, I also leverage research designs relying on less stringent assumptions in complement with model-based methods to identify the key primitive parameters of interest.

In the first chapter of this thesis titled "*Two-Sided Matching Without Transfers: A Unifying Empirical Framework*", I develop a unifying static framework of one-to-one and many-to-one matching without transfers and investigates how data on realized matches can be leveraged to identify and estimate preferences of participating agents. I find that, under parsimonious assumptions on preferences, one can only identify the joint surplus function both in the one-to-one and many-to-one case. I reconcile this finding with seemingly contradictory results from the literature. I then propose ways to overcome this negative result both in the one-to-one and many-to-one matching case making these tools applicable to a wider range of settings, such as marriage or labor markets.

In the second chapter titled "*Teacher Compensation and Structural Inequality: Evidence from Centralized School Choice in Peru*" and co-authored with Matteo Bobba, Gianmarco León-Ciliotta, Christopher Neilson and Marco Nieddu, we show that increasing teacher compensation in remote schools is effective at reducing spatial inequalities in student achievement and provide tools to design such policies in a cost effective way. Leveraging an unconditional change in the structure of teacher compensation in Perú, we first provide causal evidence that a 30% increase in salaries in rural locations attracted higher quality teachers which translated into an average increase in student test scores of 0.33-0.38 standard deviations.

We then use detailed data on job postings and applications to identify and estimate teachers' labor supply elasticities under minimal assumptions. This allows us to design a procedure that systematically delivers the wage schedule that shifts labor supply such as to reach a given social objective at a minimal cost. We use these tools to design two cost-effective wage bonus policies that would either (i) attract at least one certified teacher in each school or (ii) close the urban-rural gap in teacher quality.

Finally, in the third and last chapter titled "*Labor Market Dynamics and Teacher Spatial Sorting*", I provide a unifying explanation for the lack of supply of skilled teachers in remote locations. To do so, I build an empirical model of dynamic two-sided matching to link teachers' and schools' preferences with equilibrium sorting and job-to-job flows. I show that this mapping is invertible such that preferences can be identified and estimated from observed matches. Taking these tools to panel data on the assignment of public teachers in Peru, I show that the spatial disaggregation of labor demand coupled with the concentration of labor supply in cities imply the existence of a spatial job ladder. As a result, low quality teachers get displaced in remote schools and move toward urban schools by climbing up the ladder once they have accumulated experience and skills. Labor mobility thus magnifies the urban-rural gap in teacher quality by one third. I then show that dynamic wage contracts can largely mitigate this effect by fostering teacher retention.

Contents

Chapter 1

Two-Sided Matching Without Transfers: A Unifying Empirical Framework

Abstract. This paper provides a unifying framework of one-to-one and many-to-one matching without transfers and investigates how data on realized matches can be leveraged to identify preferences of participating agents. I find that, under parsimonious assumptions on preferences, one can only identify the joint surplus function both in the one-to-one and many-to-one case. While this negative identification result was already established for the one-to-one case, I reconcile this finding with the recent literature showing that preferences are separately identified when having data on many-to-one matchings. I find that these positive identification results are mostly driven by restrictions imposed on preferences rather than the additional identification power made available through the many-to-one structure of the data. I then show that by imposing similar restrictions on preferences, one can recover identification of preferences both in the one-to-one and many-to-one case. Finally, I show that the additional data brought by many-to-one matchings can alternatively be used to estimate more precisely the distribution of unobserved preference heterogeneity.

1.1 Introduction

Two-sided matching models with non-transferable utility are key to understand how centralized clearinghouses allocating jobs, college seats, public housing or deceased-donor kidneys are organized and how one should design them (Roth (2018), Agarwal and Budish (2021)). They are also essential tools to predict the impact of policies aiming at affecting how agents sort in such markets (Agarwal (2017)). However, this requires to know ex-ante the preferences of participating agents which are difficult to infer from observed sorting patterns only. In matching markets, agents' opportunities depend on preferences of agents from the other side. Thus, developing a revealed preference approach based on realized matches is not straightforward in the absence of prior information about preferences of one side of the market.[1]

The goal of this paper is to provide a unifying framework of one-to-one and many-to-one matching without transfers and investigate what can be identified from data on realized matches, when preferences of both sides of the market are unknown. I show that, under parsimonious assumptions on preferences, one can identify the joint surplus function both in the one-to-one and many-to-one case. However, I find that preferences of participating agents cannot be separately identified from the surplus function. Knowing the joint surplus is enough to simulate matching outcomes under various counterfactual scenarios. However, it does not allow us to characterize key objects, such as labor supply elasticities, that solely depend on individual preferences. While this negative identification result was already known for the one-to-one case (Menzel (2015)), it is at odds with the recent literature which highlights that data on many-to-one matching brings additional information that can separately identify preferences (Diamond and Agarwal (2017), He et al. (2021)). This suggests that these positive identification results are mostly driven by other restrictions imposed on preferences, rather than the additional identification power made available through the many-to-one structure of the data. In light of this result, I show that by imposing similar restrictions on preferences one can recover identification of preferences both in the one-to-one and many-to-one case, expanding the scope of what can be learned from data on one-to-one matches. Finally, I show that the additional data brought by many-to-one matchings can still be useful to estimate more precisely the distribution of unobserved preference heterogeneity.

[1]While such information is sometimes available in college admissions or school choice mechanisms (Agarwal and Somaini (2020)), preferences of both sides of the market are usually unknown.

To perform this analysis, I build on Menzel (2015) to develop a model of two-sided matching where one side is composed of firms and the other side is composed of workers. Each side is characterized by a large set of observed and unobserved attributes. I embed both the one-to-one and many-to-one framework in this model by assuming that each firm has an exogenous finite number of open vacancies, which is larger or equal than one. I impose three assumptions on the payoff functions and the equilibrium: (i) the systematic and unobserved part of the payoff functions are additively separable, (ii) the unobserved taste shocks are *iid* with type-I upper tail and (iii) the observed matching is stable. While (i) and (iii) are commonly used in the literature, (ii) departs from Diamond and Agarwal (2017) and He et al. (2021) by restricting the class of distributions taste shocks can follow for tractability purposes. However, (ii) remains nested in the broader classes they consider implying that the generality of the non identification result derived in this paper is not affected. On the other hand, I do not restrict preferences to be homogeneous and the number of agents on one side of the market to be fixed at the cost of allowing for multiple equilibria.

As in Menzel (2015), I consider that we observe a random sample of realized matches from a single large market where the number of participating firm and workers grows to infinity. Sorting patterns are thus collapsed into the limit of the joint distribution of matched characteristics. Under the assumptions described above, I characterize the mapping between this limit joint distribution function and agents' payoff functions in four steps. First, I show that stability implies that each worker is matched to its preferred firm among the set of firms that would be willing to hire her. Similarly, each firm is matched to its preferred group of workers among the set of workers that would be willing to work there. This implies that we can reinterpret the realized matches as the outcome of two discrete choice models with unobserved and endogenous choice sets, and where firms choose many alternatives. Second, I abstract away from this complexity and derive the limit of workers and firms' conditional choice probabilities under arbitrary exogenous choice sets. Third, I introduce choice sets' endogeneity and show that the information necessary to characterize conditional choice probabilities can be summarized into sufficient statistics called *inclusive values*. Finally, I show that these sufficient statistics converge to the unique solution of a fixed point problem which explicitly links agents' preferences and choice sets. This implies that all stable matches are observationally equivalent and that the limit joint distribution of matched characteristics can be expressed as a function of agents' payoff functions and inclusive values.

By inverting the mapping between the observed sorting and agents' preferences, I find that, without additional data or restrictions on preferences, one can only identify the joint surplus from data on realized matches. This shows that the additional data brought by many-to-one matchings does not help to separately identify agents' preferences. I then show that when the systematic part of the payoff functions is common to all workers/firms (as in Diamond and Agarwal (2017)), one can separately identify preferences from the joint surplus both in the one-to-one case and many-to-one case. Similarly, I find that under appropriate exclusion restrictions (as in He et al. (2021) and Agarwal and Somaini (2022)), one can also recover preferences in the one-to-one case and many-to-one case. I then propose a maximum likelihood estimator that can be tractably used for a parametric version of this framework. Finally, I validate the theoretical limiting results and test the performance of the estimation procedure proposed through Monte Carlo simulations. I find that having data on many-to-one matches allows to estimate more precisely the distribution of random coefficients, mirroring a similar result found for discrete choice models in Berry et al. (2004).

This paper contributes to the literature on empirical models of two-sided matching. One strand of this literature investigates what can be inferred from data on reported preferences within centralized allocation mechanisms (see Agarwal and Somaini (2020) for a review). These methods allowed, for example, to make progress in understanding how school choice mechanisms should be designed (Abdulkadiroğlu et al. (2017), Kapor et al. (2020)). However, in many instances, such data is not available and the econometrician can only rely on realized matches to learn about participating agents' preferences. A large literature examines what can be identified from sorting patterns in models of matching with transferable utility (TU) (Choo and Siow (2006), Fox (2010), Gualdani and Sinha (2019), Galichon and Salanié (Forthcoming)). However, only a handful of papers consider the same problem in the non-transferable utility (NTU) case (see Agarwal and Somaini (Forthcoming) for a review). Menzel (2015) shows that, under parsimonious assumptions on preferences and when matching is one-to-one, only the joint surplus is identified. To circumvent this negative result, Diamond and Agarwal (2017) find that, by restricting preferences to be common to all agents from the same side, one can separately identify preferences with data on many-to-one matches. He et al. (2021) and Agarwal and Somaini (2022) show that, by instead considering a many-to-one matching market where the number of agents on one side is fixed while the other side grows large, exclusion restrictions are sufficient and necessary in order to identify preferences from realized matches. This paper contributes to this literature by providing a unifying empirical framework of one-to-one and many-to-one matching and reconciling the results previously derived in the literature. I find that these recent positive identification results are mostly driven by the extra structure imposed on preferences and not by the inherent additional information brought by having data on many-to-one matches. This means that such methods would also work when having data on one-to-one matches which expands the scope of what can be learned from these models by making them more broadly applicable.[2]

The rest of the paper is organized as follows. Section 2.2 introduces the preference model along with the equilibrium concept. Section 1.3 defines the objects that are observed in the data and the sampling process that identifies them. Section 1.4 establishes the link between the limit joint distribution of matched characteristics and the primitives of the model. Section 1.5 discusses identification and estimation in the base model, as well as under a various set of additional restrictions on preferences. Section 1.6 displays results from Monte Carlo simulations.

1.2 Model

I consider a large two-sided matching market where the number of agents on both sides grows to infinity. I start by introducing the relevant parts of the model in the finite economy before defining the asymptotic sequence that characterizes the limit economy.

Throughout this section, I refer to one side of the market as workers and the other side as firms. Workers are indexed by $i \in \mathcal{I}$ where $\mathcal{I} = \{1, ..., n_w\}$ and firms are indexed by $j \in \mathcal{J}$ where $\mathcal{J} = \{1, ..., n_m\}$. I nest both the one-to-one and many-to-one matching framework by allowing each firm j to have a finite and exogenous number $q \geq 1$ of open vacancies.[3] I define the matching function μ_w which maps the set of available workers to their matching outcome, which is either their matched employer or the option to remain unmatched. Similarly, μ_m maps the set of available firms to their matching outcome, which is a set of length q including their matched employees as well as the option to leave any open vacancy unfilled.

For instance, consider a given worker i and firm j with $q = 2$. $\mu_w(i) = j$ means that

[2] In many empirical settings, such as centralized labor clearinghouses or college admissions, firms often open only one vacancy, making Diamond and Agarwal (2017) unapplicable, and the number of agents on both sides of the market is large, making He et al. (2021) and Agarwal and Somaini (2022) unapplicable.

[3] Allowing for each firm j to open a different number of vacancies q_j does not affect the main results of the analysis.

worker i is matched with school j whereas $\mu_m(j) = \{i, l\}$ means that firm j is matched with workers i and l. Similarly, $\mu_w(i) = 0$ means that worker i chooses to stay unmatched, while $\mu_m(j) = \{l, 0\}$ means that firm j is matched with worker l but leaves one of its vacancies unfilled. Note that all elements of the model nest Menzel (2015), which corresponds to the one-to-one case $q = 1$.

1.2.1 Preferences

Firms and workers are characterized by their observed attributes which collapse into two vectors of random variables $\boldsymbol{x}_i$ and $\boldsymbol{z}_j$. I define their probability distribution functions as $w(\boldsymbol{x})$ and $m(\boldsymbol{z})$ which have support $\mathcal{X}$ and $\mathcal{Z}$, respectively. I specify the utility that worker i gets from being matched with firm j as:

$$U_{ij} = U(\boldsymbol{x}_i, \boldsymbol{z}_j) + \sigma\eta_{ij}$$

whereas the utility that firm j gets from being matched with worker i is defined as:

$$V_{ij} = V(\boldsymbol{x}_i, \boldsymbol{z}_j) + \sigma\epsilon_{ij}$$

ϵ_{ij} and η_{ij} are worker-firm specific unobserved preference shocks and are assumed to be additively separable from the systematic part of the payoffs. I also assume that firms' preferences over groups of workers are responsive (Roth and Sotomayor (1992)). This implies that knowing firms' preferences over individual workers is enough to infer firms' preferences over groups of workers. Under this assumption, the preferred group of q workers for a given firm is composed of its q individually preferred workers.[4] I impose the following restrictions on the unknown functions U and V and the distribution of unobserved taste shocks.

Assumption 1 *(i). U and V are uniformly bounded in absolute value and $p \geq 1$ times differentiable with uniformly bounded partial derivatives in $\mathcal{X} \times \mathcal{Z}$.*
(ii). ϵ_{ij} and η_{ij} are iid and drawn independently from $\boldsymbol{x}_i$ and $\boldsymbol{z}_j$ from a distribution with absolutely continuous c.d.f. $G(s)$ and density $g(s)$. The upper tail of the distribution $G(s)$ is of type I with auxiliary function $a(s) = \frac{1-G(s)}{g(s)}$.

[4] Note that this rules out potential complementarities in preferences over workers. Relaxing this assumption would substantially complicate the analysis given that a stable equilibrium might not even exist in this case.

Assumption 1.(i) is a standard regularity condition which ensures that the functions U and V are well behaved. Assumption 1.(ii) deserves more discussion. It first assumes that observables are independent of unobserved preference shocks. This is usual in discrete choice models but might be particularly strong if we consider a market where prices are set endogenously. However, validity of this assumption can be restored through a control function approach, conditional on having an exogenous price shifter available.[5] Assumption 1.(ii) also imposes restrictions on the upper tail of the distribution of ϵ and η but leaves the lower tail unrestricted. As the number of workers and firms will grow to infinity, the number of independent draws of ϵ and η will also grow. All values of ϵ and η located in the lower tail of their distribution will thus be inconsequential in determining which alternative is the most preferred. As in Menzel (2015), I thus assume that G belongs to a class of distributions which might have different lower tails but for which the upper tail is type I extreme value distributed.[6] Note that this class of functions encompasses most of the parametric distributions traditionally used in discrete choice models. For the Gamma distribution or the Gumbel distribution, this assumption holds for $a(s) = 1$. For the standard normal distribution, this holds for $a(s) = \frac{1}{s}$.

1.2.2 Normalizations

For the limit economy to predict sorting patterns that are consistent with the finite economy, I make several additional assumptions. First, I specify the utility of the outside option as:

$$U_{i0} = \sigma \max_{k=1,\dots,J} \eta_{i0,k}$$

$$V_{0j} = \sigma \max_{k=1,\dots,J} \epsilon_{0j,k}$$

As in Menzel (2015), I then impose the following normalizations on the asymptotic sequence:

Assumption 2 *The asymptotic sequence is controlled by $n = 1, 2, \dots$ and we define:*

(i). $n_w = [\exp(\gamma_w)n]$, $n_m = [\exp(\gamma_m)n]$

(ii). $J = [n^{1/2}]$

(iii). $\sigma = \frac{1}{a(b_n)}$ *where* $b_n = G^{-1}(1 - n^{-1/2})$

[5] For example, Agarwal (2015) uses competing hospitals' Medicare reimbursements to instrument for wages in the labor market for medical residents.

[6] This class of distribution is also called the domain of attraction of the Gumbel distribution (Resnick (1987))

Assumption 2.(i) allows to control flexibly the relative sizes of each side of the market through the parameters γ_w and γ_m. Even if the size of each side converges to infinity, we can still allow for the mass of agents on one side to be larger than the other side and vice versa. Assumption 2.(ii) makes sure that the probability that workers stay unmatched or that firms keep one vacancy empty does not become degenerate in the limit. If the size of the outside option does not grow with the size of the market, the probability that it becomes dominated by an alternative option will tend to one given that taste shocks have unbounded support. Assumption 2.(iii) controls the scale of the unobserved shocks such that both unobserved and systematic parts of the payoffs jointly determine agents choices in the limit. Given that U and V are bounded and that the support of the taste shocks is unbounded, U and V would become irrelevant in the limit without this restriction. More specifically, if G is Gumbel, then $b_n \asymp \frac{1}{2}\log(n)$ and $\sigma_n = 1$. If taste shocks are standard normal, $b_n \asymp \sqrt{\log n}$ and $\sigma_n \asymp b_n$ and for Gamma distributed taste shocks, $b_n \asymp \log(n)$ and $\sigma_n = 1$.

1.2.3 Equilibrium

For the remainder of the paper, I refer to a matching as μ which collects μ_m and μ_w and summarizes the matching outcome of each agent. To rationalize the matching we observe and link it to the primitives of our model, I assume that the match is stable.

Definition 1 *For a given $q \geq 1$, a matching μ is stable if and only if for all $i = 1, ..., n_w$ and $j = 1, ..., n_m$:*

(i) Individual rationality: $U_{i\mu_w(i)} \geq U_{i0}$ and $V_{lj} \geq V_{0j}$ for all $l \in \mu_m(j)$.

(ii) No blocking pairs: There exist no pair i, j such that $U_{ij} > U_{i\mu_w(i)}$ and $V_{ij} > \min_{i' \in \mu_m(j)} V_{i'j}$.

A match is stable if agents weakly prefer their match rather than staying unmatched and if there is no worker-firm pair that would prefer be matched together instead of their current match partners. This assumption is typically used in centralized matching markets as it rules out the presence of mismatches due to frictions. Note that, for $q > 1$, this definition is valid only under the assumption that firms' preferences over groups of workers are responsive.[7] Responsiveness also ensures the existence of a stable match and of the worker-optimal/firm-optimal stable matches for $q > 1$ (Roth and Sotomayor (1992)).[8] However, when firms'

[7]For $q = 1$, $\mu_m(j)$ is a singleton for all j such that $\min_{i' \in \mu_m(j)} V_{i'j} = V_{\mu_m(j)j}$. We thus recover the same definition as in Menzel (2015).

[8]The worker-optimal stable match is the most preferred stable outcome from the workers' perspective and the least preferred stable outcome from the firms' perspective. On the contrary, the firm-optimal stable

preferences are heterogeneous, many stable matches can exist and their number grow with the size of the market. I impose no restrictions on which stable outcome is reached in the data. Throughout the rest of the paper, I thus refer to any arbitrary stable match as μ^*. I also define the worker-optimal stable match as μ^W and the firm-optimal stable match as μ^M.

1.3 Data and Sampling Process

I assume that we observe a sample of realized matches randomly drawn from the limit economy. Observed sorting patterns collapse into the matching frequency distribution function. I define this distribution in the finite economy as the function F_n which gives the expected number of groups of q workers with observable characteristics $(x_1, x_2, ..., x_q)$ matched with firms with observable characteristics z:

$$F_n(x_1, ..., x_q, z; \mu) = \frac{1}{J^{q+1}} \frac{1}{q!} \sum_{i_1=1}^{n_w} ... \sum_{i_q=1}^{n_w} \sum_{j=1}^{n_m} P(x_{i_1} \leq x_1, ..., x_{i_q} \leq x_q, z_j \leq z, \mu_m(j) = \{i_1, ..., i_q\})$$

Normalizing by $q!$ avoids counting the same matched group several times. Alternatively, for firms with observable characteristics z matched to $k < q$ workers with observable characteristics $(x_1, x_2, ..., x_k)$, F_n is defined as:

$$F_n(x_1, ..., x_k, *, z; \mu) =$$
$$\frac{1}{J^{k+1}} \frac{1}{k!} \sum_{i_1=1}^{n_w} ... \sum_{i_k=1}^{n_w} \sum_{j=1}^{n_m} P(x_{i_1} \leq x_1, ..., x_{i_k} \leq x_k, z_j \leq z, \mu_m(j) = \{i_1, ..., i_k\} \cup \{0\}^{q-k})$$

Finally, for firms with observable characteristics z leaving all their vacancies empty and unmatched workers with observable characteristics x, F_n is defined as:

$$F_n(*, z; \mu) = \frac{1}{J^2} \sum_{j=1}^{n_m} P(z_j \leq z, \mu_m(j) = \{0\}^q)$$

$$F_n(x, *; \mu) = \frac{1}{J^2} \sum_{i=1}^{n_w} P(x_i \leq x, \mu_w(i) = 0)$$

match is the most preferred stable outcome from the firms' perspective and the least preferred stable outcome from the workers' perspective.

I then denote F the limit of the distribution function F_n as the size of the market n grows to infinity. I also define the joint density of matched characteristics f which is the Radon-Nikodym derivative of the limiting measure F.

From there, I link this limiting joint density f to the density of matched characteristics that would arise under various sampling schemes. I assume that the sampling process draws individuals from the population regardless of whether they are firms or workers. One observation is thus composed of this individual alone, if it is unmatched, or along with its matched partners otherwise. Assuming that $q = 1$, the probability that a matched individual is selected by this sampling process is thus twice the probability that an unmatched individual is selected. Indeed, a matched pair could be selected either by drawing the corresponding firm or worker. For any $q \geq 1$, the probability that a matched individual is selected will thus depend on the number of other workers matched to the same firm. Indeed, if a firm is matched with three employees, the probability that any of them is selected is four times the probability that a single agent is selected. I thus define the joint density function arising from this sampling process as:

$$h(x_1, ..., x_q, z) = \frac{(q+1)f(x_1, ..., x_q, z)}{\exp\{\gamma_w\} + \exp\{\gamma_m\}}$$

where $h(x_1, ..., x_q, z)$ is the mass of firms with observable z matched with q workers with observed characteristics $(x_1, ..., x_q)$ arising from the sampling scheme defined above and $\exp\{\gamma_w\}+\exp\{\gamma_m\}$ is the total mass of workers and firms available in this economy. Similarly, I define:

$$h(x_1, ..., x_k, *, z) = \frac{(k+1)f(x_1, ..., x_k, *, z)}{\exp\{\gamma_w\} + \exp\{\gamma_m\}}$$

$$h(x, *) = \frac{f(x, *)}{\exp\{\gamma_w\} + \exp\{\gamma_m\}}$$

$$h(*, z) = \frac{f(*, z)}{\exp\{\gamma_w\} + \exp\{\gamma_m\}}$$

where $h(x_1, ..., x_k, *, z)$ is the mass of firms with observable z matched with k workers with observed characteristics $(x_1, ..., x_k)$, and $h(x, *)$ with $h(*, z)$ are the mass of unmatched workers and firms. This establishes a direct link between f and h. The next section focuses on linking f with agents' payoff functions.

1.4 Characterization of the Limit Economy

This section characterizes f, the limiting joint distribution of matched characteristics, as a function of the primitives of the model. The proof follows the same steps as Menzel (2015) and shows how each intermediary result generalizes to $q > 1$. First, I show that stability implies that the realized matches can be interpreted as the outcome of two discrete choice models with endogenous and unobserved choice sets. These choice sets are called opportunity sets and depend on preferences of the other side of the market and which stable match is selected. Second, I consider a simplified economy where opportunity sets would be observed and exogenous and derive the limit of the conditional matching probabilities. Third, I show that, the assumption imposed on the distribution of the tails of the unobserved preference shocks implies that we can use *inclusive values* as sufficient statistics to simplify the problem. These inclusive values collapse all the information contained in opportunity sets needed to characterize conditional matching probabilities. Finally, I show that these inclusive values can be represented as the approximate solution of a fixed point problem making explicit the relationship between agents' opportunity sets and preferences. This fixed point problem has a unique solution in the limit, which implies that all stable matches are observationally equivalent. I then characterize f as a function of agents' payoff functions and inclusive values.

1.4.1 Opportunity Sets

Given a match μ^*, I define the opportunity set of a worker as the set of firms that would be willing to hire her instead of one of its current matched employees. Similarly, the opportunity set of a firm is the set of workers that would be willing to quit its current employer to accept a position there. Formally, I define the opportunity set faced by a given worker $i \in \mathcal{I}$ under a match μ^* as:

$$M_i(\mu^*) = \{j \in \mathcal{J} : V_{ij} \geq \min_{i' \in \mu^*_m(j)} V_{i'j}\}$$

Similarly, I define the opportunity set of firm $j \in \mathcal{J}$ as:

$$W_j(\mu^*) = \{i \in \mathcal{I} : U_{ij} \geq U_{i\mu^*_w(i)}\}$$

I then define:

$$U_{i,(k)}(M_i(\mu^*)) = \max\{\min\{U_{ij} : j \in \mathcal{K}\} : \mathcal{K} \subset M_i(\mu^*) \cup \{0\} \text{ and } |\mathcal{K}| = k\}$$

$$V_{j,(k)}(W_j(\mu^*)) = \max\{\min\{V_{ij} : i \in \mathcal{K}\} : \mathcal{K} \subset W_j(\mu^*) \cup \{0\}^k \text{ and } |\mathcal{K}| = k\}$$

where $U_{i,(k)}(M_i(\mu^*))$ denotes the k^{th} highest element of $\{U_{ij'} : j' \in M_i(\mu^*) \cup \{0\}\}$. Note that $U_{i,(1)}(M_i(\mu^*)) = \max_{j' \in M_i(\mu^*) \cup \{0\}} U_{ij'}$. The first important result follows:

Proposition 1 *For any given $q \geq 1$, a match μ^* is stable if and only if for all $i = 1, ..., n_w$ and $j = 1, ..., n_m$:*

$$U_{i\mu_w^*(i)} = U_{i,(1)}(M_i(\mu^*)) \quad \textit{and} \quad \forall l \in \mu_m^*(j), V_{lj} \geq V_{j,(q)}(W_j(\mu^*))$$

See Appendix 1.A.1 for a proof of this result. Proposition 1 states that a match μ^* is stable if and only if each worker $i = 1, ..., n_w$ is matched to her preferred alternative among her opportunity set and each firm $j = 1, ..., n_m$ is matched to the q^{th} highest ranked alternatives among its opportunity set. This implies the following corollary:

Corollary 1 *For a given stable match μ^* and any worker i and firm j:*
(i). $j = \mu_w^*(i) \iff i \in \mu_m^*(j) \iff U_{ij} \geq U_{i,(1)}(M_i(\mu^*)) \quad \textit{and} \quad V_{ij} \geq V_{j,(q)}(W_j(\mu^*))$
(ii). $0 \in \mu_m^*(j) \iff V_{0j} \geq V_{j,(q)}(W_j(\mu^*))$
(iii). $\mu_w^*(i) = 0 \iff U_{i0} \geq U_{j,(1)}(M_i(\mu^*))$

This corollary states that a stable match μ^* can be rewritten as the outcome of two discrete choice models where each agent's choice set is its opportunity set. This equivalence establishes a link between the observed matching and the primitives of the model. However, opportunity sets are unobserved and endogenous objects as they depend on μ^* and on the preferences of agents from the other side of the market. Additionally, characterizing the probability of being among a given firm's q^{th} most preferred workers is not standard when $q > 1$. Deriving the limit of conditional matching probabilities is thus not straightforward.

1.4.2 Limit of Conditional Choice Probabilities

To simplify the analysis, I consider here arbitrary exogenous opportunity sets $M_i = \{1, ..., J\}$ and $W_j = \{1, ..., J\}$. From Corollary 1, we know that conditional matching probabilities can be characterized as two-sided conditional choice probabilities:

$$\begin{aligned}\mathbb{P}(j = \mu_w(i)|x_i, z_j) &= \mathbb{P}(U_{ij} \geq U_{i,(1)}(M_i) \quad \text{and} \quad V_{ij} \geq V_{j,(q)}(W_j)|x_i, z_j) \\ &= \mathbb{P}(U_{ij} \geq U_{i,(1)}(M_i)|x_i, z_j) \times \mathbb{P}(V_{ij} \geq V_{j,(q)}(W_j)|x_i, z_j)\end{aligned}$$

The limit of these conditional choice probabilities have the following expression:

Proposition 2 *Under Assumption 1 and 2, as $J \to \infty$ for a given finite $q \geq 1$ and for all i and j:*

$$J\mathbb{P}(U_{ij} \geq U_{i,(q)}(M_i)|x_i, z_j) \longrightarrow \exp(U(x_i, z_j)) \times \left[1 - \left(\frac{\int \exp(U(x_i, s))m(s)ds}{1 + \int \exp(U(x_i, s))m(s)ds}\right)^q\right]$$

$$\mathbb{P}(U_{i0} \geq U_{i,(q)}(M_i)|x_i) \longrightarrow \left[1 - \left(\frac{\int \exp(U(x_i, s))m(s)ds}{1 + \int \exp(U(x_i, s))m(s)ds}\right)^q\right]$$

See Appendix 1.A.2 for a proof of this result. Proposition 2 generalizes the Logit formula to cases where agents make many unranked choices among an infinite number of alternatives. For $q = 1$, we recover the usual Logit formula derived in Menzel (2015). Note that the Independence of Irrelevant Alternatives (IIA) property still holds for $q > 1$. This has several implications regarding how one can allow for realistic substitution patterns. Imposing the distribution of ϵ and η to be normal would not change this result as the normal distribution has a type-I upper tail. To avoid this problem, one can alternatively introduce unobserved preference heterogeneity for observed characteristics through the use of random coefficients. Note that the CCP of choosing a particular alternative j would converge to zero if we do not weight it by J, the rate at which the total number of alternatives increases. Lemma 1 in Appendix 1.A.3 establishes that the size of opportunity sets increases at a rate $\sqrt{n}$ which justifies Assumption 2.(ii).

1.4.3 Inclusive Values

I now introduce that opportunity sets are actually endogenous and unobserved. Endogeneity arises as shifting worker i's taste shocks could make her prefer another feasible firm to its current match. This could then trigger a chain of rematches that could potentially affect her own opportunity set. This problem is even more salient in the context of many-to-one matching as changing firm j's taste shocks could trigger at most q chains of rematches, which increases the probability that this ends up changing firm j's opportunity set. However, as in Menzel (2015), I find that, as the size of the market increases, the probability for such an event to occur vanishes to zero. This result stems mostly from two implications of Proposition 2: (i) the probability that firm j rematches with a specific worker i vanishes to zero as the size of opportunity sets increase to infinity and (ii) the probability of choosing the outside option instead, which would terminate such a chain of rematches, is non degenerate in the limit. This result is formalized in Lemma 2 in Appendix 1.A.3 where a more detailed discussion and proof can be found.

From this, I then show that the dependence between taste shocks and opportunity sets vanishes in the limit. This means that the distribution of taste shocks conditional on opportunity sets converges to their marginal distribution g. However, this claim can only be proven for the opportunity sets derived from the extremal matchings. The distribution of taste shocks conditional on opportunity sets is only well defined for the extremal matchings, given that they are the only stable matchings that always exist irrespective of the size of the market. Again, this result is formalized in Lemma 3 in Appendix 1.A.3. This means that we can use this result along with Proposition 2 to bound[9] the CCPs conditional on the opportunity sets that would arise under the firm-optimal stable match μ^M as follows:

$$\begin{aligned} n^{1/2}\mathbb{P}(U_{ij} \geq U_{i,(1)}(M_i(\mu^M))|x_i, z_j, (z_k)_{k\in M_i(\mu^M)}, M_i(\mu^M)) \\ \leq \frac{\exp\{U(x_i, z_j)\}}{1 + n^{-1/2}\sum_{k\in M_i(\mu^M)}\exp\{U(x_i, z_k)\}} + o(1) \end{aligned} \tag{1.1}$$

$$\begin{aligned} n^{1/2}\mathbb{P}(V_{ij} \geq V_{j,(q)}(W_j(\mu^M))|x_i, z_j, (x)_{l\in W_j(\mu^M)}, W_j(\mu^M)) \\ \geq \exp(V(x_i, z_j)) \times \left[1 - \left(\frac{n^{-1/2}\sum_{l\in W_j(\mu^M)}\exp\{V(x_l, z_j)\}}{1 + n^{-1/2}\sum_{l\in W_j(\mu^M)}\exp\{V(x_l, z_j)\}}\right)^q\right] + o(1) \end{aligned} \tag{1.2}$$

Similar bounds can be computed for the worker-optimal stable match μ^W where the direction of the inequalities is reversed. In Equation 1.1 and 1.2, $n^{-1/2}\sum_{k\in M_i(\mu^W)}\exp\{U(x_i, z_k)\}$ and

[9]Note that we only provide bounds given that there are several potential stable matches μ^* such that $M_i(\mu^*) = M_i(\mu^M)$ and $W_j(\mu^*) = W_j(\mu^M)$.

$n^{-1/2}\sum_{l\in W_j(\mu^M)}\exp\{V(x_l,z_j)\}$ serve as sufficient statistics that collapse all the information contained in opportunity sets which is needed to approximate CCPs. These objects are called *inclusive values.*

More generally, I define worker i's inclusive value given a realized stable match μ^* as:

$$I^*_{wi} = n^{-1/2}\sum_{j\in M_i(\mu^*)}\exp(U(x_i,z_j))$$

Similarly, I define firm j's inclusive value given μ^* as:

$$I^*_{mj} = n^{-1/2}\sum_{i\in W_j(\mu^*)}\exp(V(x_i,z_j))$$

I also define I^M_{wi} and I^M_{mj} as the inclusive values that would arise under the firm-optimal stable match and I^W_{wi} and I^W_{mj} as the inclusive values that would arise under the worker-optimal stable match.

Of course, in practice, inclusive values are unobserved and we do not know which stable match is selected. The rest of this section shows that the inclusive values arising from any stable match μ^* can be approximated by the solution of a fixed point problem which has a unique solution in the limit.

1.4.4 Fixed Point Characterization for Inclusive Values

I first show that, for any $q\geq 1$, inclusive values arising from the firm-optimal and worker-optimal stable match can be approximated by *expected inclusive value functions* (Menzel (2015)). I first rewrite I^M_{wi} as:

$$\begin{aligned} I^M_{wi} &= \frac{1}{n}\sum_{k=1}^{n_m}\exp\{U(x_i,z_k)\}\times\sqrt{n}\mathbb{1}\{k\in M_i(\mu^M)\} \\ &= \frac{1}{n}\sum_{k=1}^{n_m}\exp\{U(x_i,z_k)\}\times\sqrt{n}\mathbb{1}\{V_{ik}\geq V_{k,(q)}(W_k(\mu^M))\} \end{aligned}$$

The inclusive value of a given worker is determined by the set of firms that would accept her, which in turn depends on the preferences of all firms as well as their opportunity sets. Using

Equation 1.2, I then show that:

$$I_{wi}^M \geq \hat{\Gamma}_w^M(x_i) + o_p(1)$$

where $\hat{\Gamma}_w^M$ is the firm-optimal *expected inclusive value function* of workers which is defined as:

$$\hat{\Gamma}_w^M(x_i) = \frac{1}{n}\sum_{k=1}^{n_m} \exp\{U(x_i, z_k) + V(x_i, z_k)\} \times \left[1 - \left(\frac{I_{mk}^M}{1 + I_{mk}^M}\right)^q\right]$$

Similarly, using Equation 1.1, I show that we can approximate I_{mj}^M as follows:

$$I_{mj}^M \leq \hat{\Gamma}_m^M(z_j) + o_p(1)$$

where $\hat{\Gamma}_m^M$ is the firm-optimal *expected inclusive value function* of firms which is defined as:

$$\hat{\Gamma}_m^M(z_j) = \frac{1}{n}\sum_{l=1}^{n_w} \frac{\exp\{U(x_l, z_j) + V(x_l, z_j)\}}{1 + I_{wl}^M}$$

Note that similar bounds can be established for the inclusive values that would arise under the worker-optimal stable match:

$$I_{wi}^W \leq \hat{\Gamma}_w^W(x_i) + o_p(1) \quad \text{and} \quad I_{mj}^W \geq \hat{\Gamma}_m^W(z_j) + o_p(1)$$

A formal exposition and proof of this result can be found in Lemma 4 in Appendix 1.A.3. The inclusive value of a given worker can be approximated by a function of firms' preferences and inclusive values. Similarly, the inclusive value of a given firm can be approximated by a function of workers' preferences and inclusive values. Hence, the two-sided nature of the problem gives rise naturally to a fixed point problem characterizing these inclusive values. I define the fixed point mappings as follows:

$$\hat{\Psi}_w[\Gamma_m](x) = \frac{1}{n}\sum_{k=1}^{n_m} \exp\{U(x, z_k) + V(x, z_k)\} \times \left[1 - \left(\frac{\Gamma_m(z_k)}{1 + \Gamma_m(z_k)}\right)^q\right]$$

$$\hat{\Psi}_m[\Gamma_w](z) = \frac{1}{n}\sum_{l=1}^{n_w} \frac{\exp\{U(x_l, z) + V(x_l, z)\}}{1 + \Gamma_w(x_l)}$$

From there, I show, using Lemma 4, that for any $x \in \mathcal{X}$ and $z \in \mathcal{Z}$:

$$\hat{\Gamma}_w^M(x) \geq \hat{\Psi}_w[\hat{\Gamma}_m^M](x) + o_p(1) \quad \text{and} \quad \hat{\Gamma}_m^M(z) \leq \hat{\Psi}_m[\hat{\Gamma}_w^M](z) + o_p(1) \tag{1.3}$$

$$\hat{\Gamma}_w^W(x) \leq \hat{\Psi}_w[\hat{\Gamma}_m^W](x) + o_p(1) \quad \text{and} \quad \hat{\Gamma}_m^W(z) \geq \hat{\Psi}_m[\hat{\Gamma}_w^W](z) + o_p(1) \tag{1.4}$$

In addition, the firm-optimal stable match is unanimously preferred by firms while the worker-optimal stable match is unanimously preferred by workers (Roth and Sotomayor (1992)). This implies that $M_i(\mu^M) \subset M_i(\mu^*) \subset M_i(\mu^W)$ and $W_i(\mu^W) \subset W_i(\mu^*) \subset W_i(\mu^M)$ which means that for all i and j:

$$I_{wi}^M \leq I_{wi}^* \leq I_{wi}^W \quad \text{and} \quad I_{mj}^W \leq I_{mj}^* \leq I_{mj}^M$$

This in turn implies that for all (x, z):

$$\hat{\Gamma}_w^M(x) \leq \hat{\Gamma}_w^*(x) \leq \hat{\Gamma}_w^W(x) \quad \text{and} \quad \hat{\Gamma}_m^W(z) \leq \hat{\Gamma}_m^*(z) \leq \hat{\Gamma}_m^M(z)$$

Using Equation 1.3 and 1.4, we can thus show that, for any stable matching μ^*:

$$\hat{\Gamma}_w^*(x) = \hat{\Psi}_w[\hat{\Gamma}_m^*](x) + o_p(1) \quad \text{and} \quad \hat{\Gamma}_m^*(z) = \hat{\Psi}_m[\hat{\Gamma}_w^*](z) + o_p(1) \tag{1.5}$$

which concludes the proof that inclusive values arising from any given stable match μ^* can be approximated by the solution of a fixed point problem.

I now introduce the population equivalent of the fixed point problem described in Equation 3.3:

$$\Gamma_w^* = \Psi_w[\Gamma_m^*] \quad \text{and} \quad \Gamma_m^* = \Psi_m[\Gamma_w^*] \tag{1.6}$$

where

$$\Psi_w[\Gamma_m](x) = \int \exp(U(x,s) + V(x,s) + \gamma_m) \times \left[1 - \left(\frac{\Gamma_m(s)}{1 + \Gamma_m(s)}\right)^q\right] m(s)ds$$

$$\Psi_m[\Gamma_w](z) = \int \frac{\exp(U(s,z) + V(s,z) + \gamma_w)}{1 + \Gamma_w(s)} w(s)ds$$

This population fixed point problem has a unique solution and the approximate solution of the finite sample fixed point problem converges to it. This is stated in the following result:

Theorem 1 *Under Assumption 1 and 2 and for any $q \geq 1$:*

(i). The mapping $(\log\Gamma_w, \log\Gamma_w) \mapsto (\log\Psi_m[\Gamma_w], \log\Psi_w[\Gamma_m])$ *is a contraction.*

(ii). The fixed point problem described in Equation 1.6 always has a unique solution Γ_m^*, Γ_w^*.

(iii). For any given μ^*, $I_{wi}^* \longrightarrow \Gamma_w^*(x_i)$ *and* $I_{mj}^* \longrightarrow \Gamma_m^*(z_j)$ *for all* i *and* j.

A proof of this result can be found in Appendix 1.A.3. Theorem 1 has several implications. First, it implies that for any $q \geq 1$ and for any arbitrary stable match μ^*, inclusive values converge to the same limit. This means that all stable matches are observationally equivalent in the limit both in the one-to-one and many-to-one case. This implies that we do not need to have any information about the equilibrium selection mechanism nor do we need to impose restrictions on preferences to ensure that there is a unique stable match[10] to infer preferences from observed sorting. Second, it implies that, for any $q \geq 1$, we can use inclusive value functions as sufficient statistics to characterize CCPs as functions which only depend on agents' observable characteristics. Additionally, we know that the fixed point mappings are contractions which means that solving for inclusive value functions is computationally feasible.

1.4.5 Limit of Distribution of Matched Characteristics

Finally, using Theorem 1, I characterize the limit of the conditional matching probabilities as follows.

Proposition 3 *(i) For any firm* j *with* $q \geq 1$ *vacancies and any group of workers* $i = 1, ..., k$ *where* $k < q$*:*

$$J^{k+1}\mathbb{P}(\mu_m(j) = \{1,...,k\} \cup \{0\}^{q-k} | (x_i)_{i=1}^k, z_j) \longrightarrow \frac{k! \exp\left\{\sum_{i=1}^k U(x_i, z_j) + V(x_i, z_j)\right\}}{\prod_{i=1}^k (1 + \Gamma_w^*(x_i))\,(1 + \Gamma_m^*(z_j))^{k+1}}$$

(ii) For any firm j *with* $q \geq 1$ *vacancies:*

$$J^{q+1}\mathbb{P}(\mu_m(j) = \{1,...,q\} | (x_i)_{i=1}^q, z_j) \longrightarrow \frac{q! \exp\left\{\sum_{i=1}^q U(x_i, z_j) + V(x_i, z_j)\right\}}{\prod_{i=1}^q (1 + \Gamma_w^*(x_i))(1 + \Gamma_m^*(z_j))^q}$$

[10] Assuming that firms' preferences are homogenous, as in Diamond and Agarwal (2017), makes the stable match unique (Roth and Sotomayor (1992)). Similarly, when we assume that there is a continuum of students matching with a fixed number of colleges, as in He et al. (2021) and Agarwal and Somaini (2022), there exists a unique stable match (Azevedo and Leshno (2016)).

$$\mathbb{P}(\mu_m(j) = \{0\}^q | z_j) \longrightarrow \frac{1}{1 + \Gamma_m^*(z_j)}$$

A proof of this result can be found in Appendix 1.A.4. The probability that a given match is formed is thus positively correlated with the total match surplus $\sum_{i=1}^q U(x_i, z_j) + V(x_i, z_j)$. However, it is negatively correlated with inclusive values as they grow with the size of the set of other potential matching opportunities. It can also be noted that the rate at which these quantities converge to their limits depend on q. The larger is q the slower convergence is. This introduces a trade-off as increasing q might bring additional identification power at the cost of introducing bias due to approximation errors. From this, I characterize the limit joint distribution of matched characteristics:

$$f(x_1, ..., x_k, *, z) = \frac{\exp\left\{ \sum_{l=1}^k U(x_l, z) + V(x_l, z) + k\gamma_w + \gamma_m \right\}}{\Pi_{l=1}^k (1 + \Gamma_w^*(x_l))(1 + \Gamma_m^*(z))^{k+1}} m(z) \Pi_{l=1}^k w(x_l)$$

$$f(x_1, ..., x_q, z) = \frac{\exp\left\{ \sum_{l=1}^q U(x_l, z) + V(x_l, z) + q\gamma_w + \gamma_m \right\}}{\Pi_{l=1}^q (1 + \Gamma_w^*(x_l))(1 + \Gamma_m^*(z))^{q}} m(z) \Pi_{l=1}^q w(x_l)$$

$$f(x, *) = \frac{\exp(\gamma_w) w(x)}{1 + \Gamma_w^*(x)}$$

$$f(*, z) = \frac{\exp(\gamma_m) m(z)}{1 + \Gamma_m^*(z)}$$

Where $f(x_1, ..., x_k, *, z)$ is the mass of firms with observable z matched with k workers with characteristics $(x_1, ..., x_k)$, $f(x_1, ..., x_q, z)$ is the mass of firms with observable z matched with q workers with characteristics $(x_1, ..., x_q)$, $f(x, *)$ is the mass of unmatched workers with characteristic x and $f(*, z)$ is the mass of unmatched firms with observable z.

1.5 Identification and Estimation

1.5.1 Identification Joint Surplus

From the expression of f derived in the previous section, we can show that:

$$\frac{f(x_1, ..., x_k, *, z)}{f(x_1, ..., x_{k-1}, *, z)} = \frac{\exp\left\{ U(x_k, z) + V(x_k, z) + \gamma_w \right\}}{(1 + \Gamma_w^*(x_k))(1 + \Gamma_m^*(z))} w(x_k)$$

Inverting this mapping finally gives us:

$$U(x_k, z)+V(x_k, z) = \log f(x_1, ..., x_k, *, z)-\log f(x_1, ..., x_{k-1}, *, z)-\log f(x_k, *)-\log \frac{f(*, z)}{\exp(\gamma_m)m(z)}$$

Given that we can identify f directly from the data, as was discussed in Section 1.3, this implies that we can identify the surplus function $U + V$. Similarly, we can express inclusive values as functions of the distribution of the characteristics of unmatched individuals:

$$\Gamma^*_w(x) = \frac{\exp(\gamma_w)w(x)}{f(x, *)} - 1$$

$$\Gamma^*_m(z) = \frac{\exp(\gamma_m)m(z)}{f(*, z)} - 1$$

However, we cannot express U as a function of f separately from V and vice versa. This result is formalized in the following proposition.

Proposition 4 *Under Assumption 1 and 2 and for any $q \geq 1$:*
(i) The joint surplus function $U+V$ and the inclusive value functions Γ^*_w and Γ^*_m are identified from the limiting joint distribution of matched characteristics f.
(ii) Without further restrictions, we cannot separately identify U and V.

This means that the additional data available when $q > 1$ does not bring any additional information which would be useful to separately identify individual preferences from the joint surplus. This is in sharp contrast with Diamond and Agarwal (2017) and He et al. (2021) which find that preferences can be separately identified with data on many-to-one matching. This suggests that these positive identification results mostly rely on the extra assumptions they impose on preferences rather than the additional information made available by the many-to-one structure of the data. This would mean that, by using similar restrictions, we could thus achieve similar positive identification results even for $q = 1$. The goal of the remainder of this section is to verify this claim.

1.5.2 Homogeneous preferences

I mimic the framework developed in Diamond and Agarwal (2017) by assuming that the systematic part of the payoff functions is homogeneous across individuals. I thus define the utility that worker i gets from being matched with school j as:

$$u_{ij} = U(\boldsymbol{z}_j) + \sigma\eta_{ij}$$

whereas the utility that firm j gets from being matched with worker i is defined as:

$$v_{ij} = V(\boldsymbol{x}_i) + \sigma\epsilon_{ij}$$

Additionally, I assume that there exists $\bar{\boldsymbol{x}}$ such that $V(\bar{\boldsymbol{x}}) = 0$. Note that this framework differs from Diamond and Agarwal (2017) on two dimensions. Taste shocks are heterogeneous and *iid* over i, j and the class of distribution to which they belong is more restrictive. Under these assumptions, it is immediate to see that we can recover U and V from the joint surplus as $U(\boldsymbol{z}) + V(\bar{\boldsymbol{x}}) = U(\boldsymbol{z})$. I state the following result:

Proposition 5 *Under Assumptions 1 and 2 and for any $q \geq 1$, the payoff functions U and V are identified from the limiting joint distribution of matched characteristics f.*

This shows that a similar positive identification result as the one derived in Diamond and Agarwal (2017) can actually be achieved for both $q > 1$ and $q = 1$ by using similar restrictions on preferences. In fact, this suggests that their non identification result for $q = 1$ is mostly driven by the assumption they impose on the correlation structure of the unobserved taste shocks. As is pointed out by the authors, assuming that taste shocks are common to all agents from the same side makes the unique stable match perfectly assortative along these unobserved tastes. This creates an endogeneity problem. It thus becomes necessary to have data on at least two-to-one matching in order to have an additional measurement of these sorting patterns that would allow to disentangle the effect of observed and unobserved preferences. In the framework developed in this paper, this problem does not exist given that taste shocks are *iid* across individuals. In the limit, conditional matching probabilities are uniquely determined by observable characteristics even when $q = 1$.

1.5.3 Exclusion restrictions

As in He et al. (2021) and Agarwal and Somaini (2022), I assume that a set of variables affecting the utility of one side can be excluded from the utility of the other side. I define the utility that worker i gets from being matched with firm j as:

$$U_{ij} = U(\boldsymbol{x}_i, \boldsymbol{z}_j) + \sigma\eta_{ij}$$

whereas the utility that firm j gets from being matched with worker i is defined as:

$$V_{ij} = V(\boldsymbol{x}_i, \boldsymbol{z}_j) + g(w_i) + \sigma\epsilon_{ij}$$

I additionally assume that g is increasing in w and that $\lim_{w\to\infty} g(w) = \infty$.[11] I also assume that there exists $\bar{w}$ such that $g(\bar{w}) = 0$. Under these assumptions, we can state the following result:

Proposition 6 *Under Assumptions 1 and 2 and for any $q \geq 1$, the payoff functions U, V and g are identified from the limiting joint distribution of matched characteristics f.*

A proof of this result can be found in Appendix 1.A.5. Similarly to the argument used in He et al. (2021) and Agarwal and Somaini (2022), increasing w shifts the probability that a given firm becomes available which allows us to disentangle the role of firms' and workers' preferences in determining the sorting patterns we observe.[12] This argument also holds for $q = 1$ and the many-to-one structure of the data does not help in making this additional source of identification more salient. Note that we do not need here to have preference shifters for both sides of the market as in Agarwal and Somaini (2022) and He et al. (2021). As the joint surplus is already identified in the absence of exclusion restrictions, we only need to identify preferences of workers to recover preferences of firms from the surplus.[13]

1.5.4 Unobserved Preference Heterogeneity

In light of the previous results, one can wonder how we could use the additional information made available by data on many-to-one matching, if not for disentangling preferences from the joint surplus. Using a similar argument as what is used in the discrete choice literature, I

[11] This is similar to Assumption 2 in Agarwal and Somaini (2022).

[12] The identification argument only works at infinity as the match has not a fixed cutoff structure as in He et al. (2021) and Agarwal and Somaini (2022). As both the number of firms and workers grow to infinity in our case, the cutoffs grow to infinity as the size of the market grows.

[13] In Agarwal and Somaini (2022) and He et al. (2021) it is not clear whether the joint surplus is identified in the absence of exclusion restrictions. Further work could determine whether it is only the case when taste shocks have type-I upper tails.

claim that having data on several decisions made by the same firm allows to know more about the unobserved "type" this firm belongs to. More specifically, if we were to assume that firms have an unobserved individual and heterogeneous taste for a given worker characteristic x_i, having more than one measurement of given firm j's choice would be useful to pin it down. This is analogous to what is argued by Berry et al. (2004) who show that estimating random coefficients from a cross section of observed choices often fails when having only the first ranked choice of each consumer. Having at least the second choice of each consumer allows to disentangle what drives observed choices between random coefficients and unobserved taste shocks. A similar argument could apply with data on many-to-one matching given that we observe several workers matched to the same firm. I investigate in Section 1.6, whether such gains could also be achieved in a matching market setting thanks to the many-to-one structure of the data.

1.5.5 Estimation

Given that the identification proof is constructive, one could construct naturally a non-parametric estimator for the joint surplus function $U + V$ and the inclusive value functions Γ_w^* and Γ_m^*. However, this would quickly become intractable as the dimensionality of x and z increases.

I instead consider a parametric version of this framework where I define the payoff functions as $U(x, z; \boldsymbol{\theta})$ and $V(x, z; \boldsymbol{\theta})$. I assume that U and V are known for all (x, z) up to a vector of unknown parameters $\boldsymbol{\theta}$. Assume that we observe a random sample of K individuals, drawn from the sampling scheme described in Section 1.3, along with their respective matches. For a given observation k, we observe a vector $(x_1(k), ..., x_q(k), z(k))$ which has a different structure depending on the type of match we observe. For an unmatched worker, which is indexed by $w(k) = 0$, I record its characteristics in $x_1(k)$ and encode the other variables as missing. For an unmatched firm, which is indexed by $m(k) = 0$, I record its characteristics in $z(k)$. And for a firm matched with a group of workers of size n, indexed by $m(k) = n$, I record the characteristics of all matched workers along with the characteristic of the matched firm in $(x_1(k), ..., x_n(k), z(k))$ and encode the rest as missing. We can then construct the following sample average log-likelihood:

$$\begin{aligned}
L(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}) = \frac{1}{K} \sum_{k=1}^{K} & \mathbb{1}\{w(k) = 0\} h(x(k), *, \boldsymbol{\theta}) + \mathbb{1}\{m(k) = 0\} h(*, z(k), \boldsymbol{\theta}) \\
& + \mathbb{1}\{m(k) = 1\} h(x_1(k), *, z(k), \boldsymbol{\theta}) \\
& + \mathbb{1}\{m(k) = 2\} h(x_1(k), x_2(k), *, z(k), \boldsymbol{\theta}) \\
& + ... \\
& + \mathbb{1}\{m(k) = q\} h(x_1(k), ..., x_q(k), z(k), \boldsymbol{\theta})
\end{aligned}$$

where h is the joint density of matched characteristics under the sampling scheme described in Section 1.3. Of course, calculating the likelihood function for a given parameter $\boldsymbol{\theta}$ first involves solving for the fixed point problem described in Equation 1.6 to derive the inclusive value functions. This can be achieved by setting up an inner loop that will apply the contraction mapping until convergence. The estimator proposed is then defined as:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \Theta}{\arg\max}\, L(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta})$$

Asymptotic inference for $\hat{\boldsymbol{\theta}}$ is then standard as long as the size of the sample is not too large relative to the size of the overall economy. As is pointed out in Menzel (2015) and Diamond and Agarwal (2017), the inherent structure of matching markets could introduce dependence between observations. A bootstrap procedure could then be used for inference (Diamond and Agarwal (2017), Menzel (2021)).

1.6 Monte Carlo Simulations

In this section, I perform several Monte Carlo simulations in order to assess: (i) the validity of the convergence results derived in Section 1.4 and (ii) the validity of the estimation strategy proposed in Section 1.5.

1.6.1 Convergence of Conditional Match Probabilities

Consider a simple model where $q = 2$ and $U(\boldsymbol{x}, \boldsymbol{z}) = V(\boldsymbol{x}, \boldsymbol{z}) = 0$ for all $(x, z) \in \mathcal{X} \times \mathcal{Z}$. In this example, we can easily solve for the fixed point problem described in Equation 1.6 given

Table 1.1: Monte Carlo: Convergence of Matching Frequencies and Inclusive Values

n	Unmatched Workers	Firms with One Unfilled Vacancy	Firms with Two Unfilled Vacancies	I_w	I_m
20	0.5894	0.6778	0.2339	0.7617	0.5343
50	0.5714	0.6712	0.2290	0.8007	0.5365
100	0.5608	0.6661	0.2285	0.8245	0.5350
200	0.5521	0.6621	0.2279	0.8406	0.5344
500	0.5449	0.6588	0.2273	0.8546	0.5313
1000	0.5418	0.6577	0.2263	0.8611	0.5320
2000	0.5389	0.6561	0.2268	0.8662	0.5335
Model	0.5321	0.6527	0.2267	0.8794	0.5321

NOTES. This table reports the average share of unmatched firms and workers in each period taken over 200 sample draws for different sample sizes n.

that inclusive value functions collapse to a fixed number which does not vary with $(\boldsymbol{x}, \boldsymbol{z})$. This results in $\Gamma_w^* = 0.8794$ and $\Gamma_m^* = 0.5321$. I also compute the limit matching frequencies:

$$\mathbb{P}\left(U_{i0} \geq U_{i,(1)}^*\right) \longrightarrow 0.5321$$

$$\mathbb{P}\left(V_{0j} \geq V_{i,(1)}^*\right) \longrightarrow 0.2267$$

$$\mathbb{P}\left(V_{i,(1)}^* > V_{0j} \geq V_{i,(2)}^*\right) \longrightarrow 0.6527$$

To verify the validity of the large market approximation, I first simulate n individuals along with their taste shocks over the individuals from the other side of the market ϵ_{ij} and η_{ij} for all (i, j). I then use the worker-proposing Deferred Acceptance algorithm to get the worker-optimal stable match. Finally, I compute the empirical matching frequencies and the inclusive values under this stable match and check whether they converge to their theoretical limits as n grows large. Table 1.1 displays the result of this exercise. Both the inclusive values and the matching frequencies converge to their theoretical limits. This table also shows that the limit economy is a relatively good approximation even when the size of the market is moderately large.

Table 1.2: Monte Carlo: Estimation without Random Coefficient

	$q = 1$			$q = 2$			$q = 3$		
n	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$
100	0.989	1.048	0.511	0.942	1.053	0.494	0.944	1.051	0.457
	(0.309)	(0.374)	(0.231)	(0.202)	(0.334)	(0.173)	(0.159)	(0.372)	(0.152)
200	0.999	0.962	0.487	0.957	1.047	0.486	0.951	1.045	0.468
	(0.223)	(0.236)	(0.150)	(0.155)	(0.229)	(0.121)	(0.125)	(0.236)	(0.113)
500	0.991	0.997	0.497	0.976	1.018	0.479	0.966	1.013	0.476
	(0.153)	(0.155)	(0.101)	(0.101)	(0.153)	(0.082)	(0.082)	(0.143)	(0.071)
1000	0.989	0.992	0.500	0.987	1.009	0.486	0.974	1.005	0.485
	(0.108)	(0.111)	(0.069)	(0.066)	(0.102)	(0.055)	(0.060)	(0.097)	(0.047)
Model	1	1	0.5	1	1	0.5	1	1	0.5

NOTES. This table reports the average and standard deviation of the ML estimator of $\boldsymbol{\theta}$ for different values of q taken over 200 sample draws for different sample sizes n.

1.6.2 Convergence of ML Estimator

I now evaluate the performance of the estimator proposed in Section 1.5 through two Monte Carlo exercises. In the first exercise, I consider the following simple parametric framework:

$$U_{ij} = \theta_1 z_j + \eta_{ij} \quad \text{and} \quad V_{ij} = \theta_2 x_i + \theta_3 x_i z_j + \epsilon_{ij}$$

and estimate $\boldsymbol{\theta}$ on simulated data. To do so, I draw n individuals along with their observed characteristics x_i and z_j drawn from a standard normal distribution. I draw the taste shocks ϵ_{ij} and η_{ij} from the Gumbel distribution and set $\boldsymbol{\theta} = (1, 1, 0.5)$ to compute U_{ij} and V_{ij} for all (i, j). I then derive the worker-optimal stable match using the Deferred Acceptance algorithm. Finally, I estimate $\boldsymbol{\theta}$ and repeat this process 200 times to report the mean and standard deviation of $\hat{\boldsymbol{\theta}}$ over the sample draws. Table 1.2 shows that the estimator seems to converge to its true value as the size of the market increases given that the mean converges to the true value while the standard deviation vanishes. We can also see that, while increasing q lowers the variance of the estimator, it also seems to introduce bias. This is consistent with Proposition 3, given that the joint conditional matching probabilities converge to their theoretical limits at a slower rate when q increases. There is thus a trade off involved as increasing q allows to have a more precise estimator, given that we are using more information, but might introduce distortions as conditional matching frequencies converge to their limit at a slower rate.

Table 1.3: Monte Carlo: Estimation with Random Coefficient

	$q=1$				$q=2$				$q=3$			
n	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\mathbb{P}(\hat{\theta}_3=0)$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\mathbb{P}(\hat{\theta}_3=0)$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\mathbb{P}(\hat{\theta}_3=0)$
100	1.006	0.997	0.300	0.485	0.995	0.992	0.311	0.280	0.976	0.990	0.267	0.315
	(0.334)	(0.352)	(0.363)	-	(0.217)	(0.307)	(0.260)	-	(0.168)	(0.265)	(0.234)	-
200	1.037	0.969	0.286	0.425	0.993	0.965	0.317	0.210	0.964	0.960	0.327	0.110
	(0.191)	(0.206)	(0.301)	-	(0.152)	(0.208)	(0.224)	-	(0.113)	(0.195)	(0.169)	-
500	0.995	0.974	0.294	0.310	0.992	0.970	0.350	0.065	0.991	0.977	0.389	0.025
	(0.125)	(0.130)	(0.256)	-	(0.086)	(0.122)	(0.146)	-	(0.068)	(0.128)	(0.116)	-
1000	1.007	0.963	0.348	0.115	0.995	0.981	0.405	0.005	0.995	0.981	0.414	0
	(0.100)	(0.101)	(0.190)	-	(0.060)	(0.086)	(0.105)	-	(0.056)	(0.077)	(0.073)	-
Model	1	1	0.5	-	1	1	0.5	-	1	1	0.5	-

NOTES. This table reports the average and standard deviation of the ML estimator of $\boldsymbol{\theta}$ for different values of q taken over 200 sample draws for different sample sizes n.

In a second exercise, I now consider the following parametric framework:

$$U_{ij} = \theta_1 z_j + \eta_{ij} \quad \text{and} \quad V_{ij} = \theta_2 x_i + \theta_3 x_i \nu_j + \epsilon_{ij}$$

where ν_j is unobserved and follows a $\mathcal{N}(0,1)$. In this example, I assume that there is unobserved heterogeneity in schools' tastes over x_i which is parametrized through a normal distributed random coefficient with mean θ_2 and standard deviation θ_3. I then follow similar steps as for the first exercise to get the mean and the standard deviation of the ML estimator of $\boldsymbol{\theta}$ for different values of q and different sizes of the economy n. Note that to approximate the integral over ν_i to compute the conditional matching probabilities, I use a Gaussian-Hermite quadrature (Judd (1998)). Table 1.3 shows that $\hat{\boldsymbol{\theta}}$ converges to its true value as n increases. However, the standard deviation of the random coefficient θ_3 is poorly estimated when $q=1$. Even with large n, $\hat{\theta}_3$ is equal to 0 in 11% of the cases. Given that the log-likelihood function is symmetric around $\theta_3 = 0$, this indicates that it is maximized while being not differentiable at this point. In this case, traditional inference breaks down and this estimator is not informative. Although this issue also arises for $q=2$ and $q=3$ when n is small, we can see that as q increases this is less likely to happen. In fact, increasing q from 1 to 2 is already enough to drastically reduce the probability of estimating θ_3 to 0. This indicates that having data on two-to-one matching is already enough to bring additional identification power necessary to pin down the distribution of random coefficients. This mirrors the result found in Berry et al. (2004) which shows that having data on consumers' second choices allows to estimate random coefficients more easily.

1.7 Conclusion

This paper develops a unifying empirical framework of one-to-one and many-to-one matching without transfers to understand what can be inferred on agents' preferences from observed sorting in such markets. I impose few restrictions on preferences and assume that the observed matching is stable. Stability allows me to rewrite the model as a two-sided discrete choice model with endogenous and unobserved choice sets. I use a sufficient statistics approach to take into account choice sets' endogeneity and characterize agents' conditional choice probabilities. This allows me to form a clear mapping between the joint distribution of matched characteristics and agents' payoff functions.

I then show that we can identify the joint surplus from both one-to-one and many-to-one matching data. However, without further restrictions, individual preferences are not identified. While this negative identification result was already established in the one-to-one case, the literature has argued that many-to-one matchings can bring additional information which would allow to separately identify preferences from the joint surplus. I find that these positive identification results are ultimately not driven by the availability of such additional information but mostly by the extra assumptions imposed on preferences.

I then argue that, by imposing similar restrictions on preferences, one can extend these positive identification result to the one-to-one matching case. More specifically, by either assuming that the systematic parts of the payoffs is homogenous across individuals (as in Diamond and Agarwal (2017)) or under appropriate exclusion restrictions (as in He et al. (2021) and Agarwal and Somaini (2022)), one can separately identify preferences from the joint surplus both in the one-to-one and many-to-one case. Finally, I show that the additional information brought by the many-to-one structure of the data can instead be used to estimate more precisely the distribution of random coefficients in a parametric framework.

Appendices

1.A Proofs

1.A.1 Proof of Proposition 1

Suppose first that μ is not stable. This could imply first, by definition of stability, that there exists a pair (i,j) such that $U_{ij} > U_{i\mu_w(i)}$ and $V_{ij} > \min_{i' \in \mu_m(j)} V_{i'j}$. This would mean that there exists a pair (i,j) such that $j \in M_i(\mu)$ and $U_{ij} > U_{i\mu_w(i)}$ which contradicts that $U_{i\mu_w(i)} = \max_{j' \in M_i(\mu) \cup \{0\}} U_{ij'}$. This could also imply that $U_{i0} > U_{i\mu_w(i)}$ or $V_{0j} > \min_{i' \in \mu_m(j)} V_{i'j}$. In the first case, this would contradict that $U_{i\mu_w(i)} = \max_{j' \in M_i(\mu) \cup \{0\}} U_{ij'}$. In the second case, this would mean that there exist a $l \in \mu_m(j)$ such that $V_{0j} > V_{lj}$ which contradicts that $V_{lj} \geq V_{j,(q)}(W_j(\mu))$.

Now, suppose that for a given i, $U_{i\mu_w(i)} < \max_{j' \in M_i(\mu) \cup \{0\}} U_{ij'}$. This means that there exist a firm $k \in M_i(\mu) \cup \{0\}$ such that $U_{ik} > U_{i\mu_w(i)}$. If $k = 0$, this immediately contradicts stability. If $k \in M_i(\mu)$, this implies that there exist a firm k such that $V_{ik} \geq \min_{i' \in \mu_m(k)} V_{i'k}$ and $U_{ik} > U_{i\mu_w(i)}$. If $V_{ik} = \min_{i' \in \mu_m(k)} V_{i'k}$, this implies that $k = \mu_w(i)$ and we reach a contradiction. Otherwise, we have that $U_{ik} > U_{i\mu_w(i)}$ and $V_{ik} > \min_{i' \in \mu_m(k)} V_{i'k}$ which contradicts stability.

Finally, suppose that for a given j and for a given $l \in \mu_m(j)$, $V_{lj} < V_{j,(q)}(W_j(\mu))$. This means that there exist a worker s such that $s \in W_j(\mu) \cup \{0\}$ and $V_{sj} > V_{lj}$. If $s = 0$, this contradicts stability. If $s \in W_j(\mu)$, this implies that $U_{sj} \geq U_{s\mu_w(s)}$ and $V_{sj} > V_{lj}$. Again, we restrict ourselves to the case where $j \neq \mu_w(s)$ which implies that $U_{sj} > U_{s\mu_w(s)}$ and $V_{sj} > V_{lj}$ which contradicts stability. This concludes the proof.

1.A.2 Proof of Proposition 2

I first consider the case $q = 2$. The proof for $q = 1$ can be found in Menzel (2015). I start by decomposing in two terms the conditional probability that U_{ij} is above or equal $U_{i,(2)}(M_i)$ where $M_i = \{0, ..., J\}$. I remove the dependence on M_i for simplicity such that $U_{i,(q)}(M_i) = U_{i,(q)}$ for all q. I also rewrite $U_{ij} = u_{ij} + \sigma\eta_{ij}$ for simplicity.

$$\begin{aligned}\mathbb{P}(U_{ij} \geq U_{i,(2)}|(u_{ik})_{k=1}^J) = \mathbb{P}(U_{ij} \geq U_{i,(1)}|(u_{ik})_{k=1}^J) \\ + \mathbb{P}(U_{i,(1)} > U_{ij} \geq U_{i,(2)}|(u_{ik})_{k=1}^J)\end{aligned}$$

The first term is known already and is the conditional choice probability for $q = 1$. The second term can be expressed as the probability that there exists one alternative preferred to j but that j is preferred to the rest:

$$\begin{aligned}\mathbb{P}(U_{i,(1)} > U_{ij} \geq U_{i,(2)}|(u_{ik})_{k=1}^J) &= \int \sum_{k=1}^J \mathbb{P}(U_{ik} > U_{ij}, U_{ij} \geq U_{il}, l \in \mathcal{I} - \{k, j\}|(u_{ik})_{k=1}^J, \eta_{ij} = s)g(s)ds \\ &= \int \sum_{k=1}^J (1 - G(\sigma^{-1}(u_{ij} - u_{ik}) + s)) \prod_{l \in \mathcal{I}-\{k,j\}} G(\sigma^{-1}(u_{ij} - u_{il}) + s)g(s)ds \\ &= \int \sum_{k=1}^J \frac{1 - G(\sigma^{-1}(u_{ij} - u_{ik}) + s)}{G(\sigma^{-1}(u_{ij} - u_{ik}) + s)} \prod_{l=1}^{2J} G(\sigma^{-1}(u_{ij} - u_{il}) + s)\frac{g(s)}{G(s)}ds\end{aligned}$$

As in Menzel (2015), I then do the change of variables $s = a_J t + b_J$ where $a_J = a(b_J)$ and $b_J = G^{-1}(1 - J^{-1/2})$ and multiply by J on both sides:

$$\begin{aligned}J\mathbb{P}(U_{i,(1)} > U_{ij} \geq U_{i,(2)}|(u_{ik})_{k=1}^J) = \int \frac{1}{J} \sum_{k=1}^J \frac{J(1 - G(a_J(u_{ij} - u_{ik} + t) + b_J))}{G(a_J(u_{ij} - u_{ik} + t) + b_J)} \\ \times \exp\left(\frac{1}{J}\sum_{l=1}^{2J} J \log G(a_J(u_{ij} - u_{il} + t) + b_J)\right) \frac{Ja_J g(a_J t + b_J)}{G(a_J t + b_J)}dt\end{aligned}$$

Following Resnick (1987) and under Assumption 1 we can show that:

$$J(1 - G(a_J(u_{ij} - u_{ik} + t) + b_J)) \to e^{-(u_{ij} - u_{ik} + t)}$$

$$G(a_J(u_{ij} - u_{ik} + t) + b_J) \to 1$$

$$J \log G(a_J(u_{ij} - u_{il} + t) + b_J) \to -e^{-(u_{ij}-u_{ik}+t)}$$

$$\frac{J a_J g(a_J t + b_J)}{G(a_J t + b_J)} \to e^{-t}$$

We thus have under Assumption 1:

$$\begin{aligned} J\mathbb{P}(U_{i,(1)} > U_{ij} \geq U_{i,(2)} | (u_{ik})_{k=1}^J) &= \int \frac{1}{J} \sum_{k=1}^J e^{-(u_{ij}-u_{ik}+t)} \exp\left(-\frac{1}{J} \sum_{l=1}^{2J} e^{-(u_{ij}-u_{ik}+t)} \right) e^{-t} dt + o(1) \\ &= \int \frac{1}{J} \sum_{k=1}^J e^{(u_{ik}-u_{ij})} \exp\left(-\frac{1}{J} \sum_{l=1}^{2J} e^{-t} e^{(u_{ik}-u_{ij})} \right) e^{-t} e^{-t} dt + o(1) \end{aligned}$$

I then do a final change of variable $s = e^{-t}$ such that we get:

$$\begin{aligned} J\mathbb{P}(U_{i,(1)} > U_{ij} \geq U_{i,(2)} | (u_{ik})_{k=1}^J) &= \int_0^{+\infty} \frac{1}{J} \sum_{k=1}^J e^{(u_{ik}-u_{ij})} \exp\left(-\frac{1}{J} \sum_{l=1}^{2J} s e^{(u_{ik}-u_{ij})} \right) s ds + o(1) \\ &= \frac{1}{J} \sum_{k=1}^J e^{(u_{ik}-u_{ij})} \left(\frac{1}{J} \sum_{k=1}^{2J} e^{(u_{ik}-u_{ij})} \right)^{-2} + o(1) \\ &= \frac{1}{J} \sum_{k=1}^J e^{u_{ik}} \frac{\exp(u_{ij})}{\left(\frac{1}{J} \sum_{k=1}^{2J} \exp(u_{ik}) \right)^2} + o(1) \\ &= \frac{\exp(u_{ij})}{1 + \frac{1}{J} \sum_{k=1}^J \exp(u_{ik})} \times \frac{\frac{1}{J} \sum_{k=1}^J \exp(u_{ik})}{1 + \frac{1}{J} \sum_{k=1}^J \exp(u_{ik})} + o(1) \end{aligned}$$

From this we can finally show that:

$$J\mathbb{P}(U_{i,(1)} > U_{ij} \geq U_{i,(2)} | x_i, (z_k)_{k=1}^J) = \frac{\exp(U(x_i, z_j))}{1 + \frac{1}{J} \sum_{k=1}^J \exp(U(x_i, z_k))} \times \frac{\frac{1}{J} \sum_{k=1}^J \exp(U(x_i, z_k))}{1 + \frac{1}{J} \sum_{k=1}^J \exp(U(x_i, z_k))} + o(1)$$

which implies that:

$$J\mathbb{P}(U_{i,(1)} > U_{ij} \geq U_{i,(2)} | x_i, z_j) \longrightarrow \frac{\exp(U(x_i, z_j))}{1 + \int \exp(U(x_i, s)) m(s) ds} \times \frac{\int \exp(U(x_i, s)) m(s) ds}{1 + \int \exp(U(x_i, s)) m(s) ds}$$

We know from Menzel (2015) that:

$$J\mathbb{P}(U_{ij} \geq U_{i,(1)} | x_i, z_j) \longrightarrow \frac{\exp(U(x_i, z_j))}{1 + \int \exp(U(x_i, s)) m(s) ds}$$

So we can conlude that:

$$J\mathbb{P}(U_{ij} \geq U_{i,(2)}|x_i, z_j) \longrightarrow \frac{\exp(U(x_i, z_j))}{1 + \int \exp(U(x_i, s))m(s)ds} \times \left(1 + \frac{\int \exp(U(x_i, s))m(s)ds}{1 + \int \exp(U(x_i, s))m(s)ds}\right)$$

Following similar steps, we can prove that:

$$\mathbb{P}(U_{i0} \geq U_{i,(2)}|x_i) \longrightarrow \frac{1}{1 + \int \exp(U(x_i, s))m(s)ds} \times \left(1 + \frac{\int \exp(U(x_i, s))m(s)ds}{1 + \int \exp(U(x_i, s))m(s)ds}\right)$$

To illustrate how this iterates to any q, I write here the proof for $q = 3$. Similarly, we want to characterize the probability that there exists two alternatives preferred to j but that j is preferred to the rest:

$$\begin{aligned}
&\mathbb{P}(U_{i,(2)} > U_{ij} \geq U_{i,(3)}|(u_{ik})_{k=1}^J) \\
&= \int \frac{1}{2} \sum_{k=1}^J \sum_{\substack{m=1 \\ m \neq k}}^J \mathbb{P}(U_{ik} > U_{ij}, U_{im} > U_{ij}, U_{ij} \geq U_{il}, l \in \mathcal{I} \setminus \{k, m, j\}|(u_{ik})_{k=1}^J, \eta_{ij} = s)f(s)ds \\
&= \int \frac{1}{2} \sum_{k=1}^J \frac{1 - G(\sigma^{-1}(u_{ij} - u_{ik}) + s)}{G(\sigma^{-1}(u_{ij} - u_{ik}) + s)} \sum_{\substack{m=1 \\ m \neq k}}^J \frac{1 - G(\sigma^{-1}(u_{ij} - u_{im}) + s)}{G(\sigma^{-1}(u_{ij} - u_{im}) + s)} \prod_{l=1}^{2J} G(\sigma^{-1}(u_{ij} - u_{im}) + s) \frac{g(s)}{G(s)} ds \\
&= \int \frac{1}{2} \frac{1}{J} \sum_{k=1}^J \frac{J(1 - G(\sigma^{-1}(u_{ij} - u_{ik}) + s))}{G(\sigma^{-1}(u_{ij} - u_{ik}) + s)} \frac{1}{J-1} \sum_{\substack{m=1 \\ m \neq k}}^J \frac{(J-1)(1 - G(\sigma^{-1}(u_{ij} - u_{im}) + s))}{G(\sigma^{-1}(u_{ij} - u_{im}) + s)} \\
&\times \exp\left(\frac{1}{J} \sum_{l=1}^J J \log G(\sigma^{-1}(u_{ij} - u_{il}) + s)\right) \frac{Jg(s)}{G(s)} ds \\
&= \int \frac{1}{2} \frac{1}{J} \sum_{k=1}^J e^{-(u_{ij} + t - u_{ik})} \frac{1}{J-1} \sum_{\substack{m=1 \\ m \neq k}}^J e^{-(u_{ij} + t - u_{im})} \exp\left(-\frac{1}{J} \sum_{l=1}^{2J} \exp -(u_{ij} + s - u_{il})\right) e^{-t} dt + o(1) \\
&= \int \frac{1}{2} \frac{1}{J} \sum_{k=1}^J e^{(u_{ik} - u_{ij})} \frac{1}{J-1} \sum_{\substack{m=1 \\ m \neq k}}^J e^{(u_{im} - u_{ij})} \exp\left(-e^{-t} \frac{1}{J} \sum_{l=1}^{2J} \exp(u_{il} - u_{ij})\right) e^{-t} e^{-t} e^{-t} dt + o(1) \\
&= \int_0^{+\infty} \frac{1}{2} \frac{1}{J} \sum_{k=1}^J e^{(u_{ik} - u_{ij})} \frac{1}{J-1} \sum_{\substack{m=1 \\ m \neq k}}^J e^{(u_{im} - u_{ij})} \exp\left(-s \frac{1}{J} \sum_{l=1}^{2J} \exp(u_{il} - u_{ij})\right) s^2 ds + o(1) \\
&= \frac{1}{J} \sum_{k=1}^J e^{(u_{ik} - u_{ij})} \frac{1}{J-1} \sum_{\substack{m=1 \\ m \neq k}}^J e^{(u_{im} - u_{ij})} \times \left(\frac{1}{J} \sum_{l=1}^{2J} \exp(u_{il} - u_{ij})\right)^{-3} + o(1)
\end{aligned}$$

$$= \frac{\exp(u_{ij})}{\frac{1}{J}\sum_{k=1}^{2J}\exp(u_{ik})} \times \frac{\frac{1}{J}\sum_{k=1}^{J}\exp(u_{ik})\frac{1}{J-1}\sum_{\substack{m=1\\m\neq k}}^{J}\exp(u_{im})}{\left(\frac{1}{J}\sum_{l=1}^{2J}\exp(u_{il})\right)^2} + o(1)$$

$$= \frac{\exp(u_{ij})}{1+\frac{1}{J}\sum_{k=1}^{J}\exp(u_{ik})} \times \left(\frac{\frac{1}{J}\sum_{k=1}^{J}\exp(u_{ik})}{1+\frac{1}{J}\sum_{k=1}^{J}\exp(u_{ik})}\right)^2 + o(1)$$

From this we then have that:

$$J\mathbb{P}(U_{i,(2)} > U_{ij} \geq U_{i,(3)}|x_i, z_j) \longrightarrow \frac{\exp(U(x_i, z_j))}{1+\int \exp(U(x_i, s))m(s)ds} \times \left(\frac{\int \exp(U(x_i, s))m(s)ds}{1+\int \exp(U(x_i, s))m(s)ds}\right)^2$$

We can thus conclude that:

$$J\mathbb{P}(U_{ij} \geq U_{i,(3)}|x_i, z_j) \longrightarrow \frac{\exp(U(x_i, z_j))}{1+\int \exp(U(x_i, s))m(s)ds} \times \sum_{k=1}^{3}\left(\frac{\int \exp(U(x_i, s))m(s)ds}{1+\int \exp(U(x_i, s))m(s)ds}\right)^{k-1}$$

$$\mathbb{P}(U_{i0} \geq U_{i,(3)}|x_i, z_j) \longrightarrow \frac{1}{1+\int \exp(U(x_i, s))m(s)ds} \times \sum_{k=1}^{3}\left(\frac{\int \exp(U(x_i, s))m(s)ds}{1+\int \exp(U(x_i, s))m(s)ds}\right)^{k-1}$$

To prove this result for any q, I derive the limit of the following conditional probability:

$$\mathbb{P}(U_{i,(q-1)} > U_{ij} \geq U_{i,(q)}|(u_{ik})_{k=1}^{J})$$

Following the same steps, the probability that there exists $q-1$ alternatives preferred to j but that j is preferred to the rest can be expressed as:

$$\mathbb{P}(U_{i,(q-1)} > U_{ij} \geq U_{i,(q)}|(u_{ik})_{k=1}^{J})$$
$$= \int \frac{1}{(q-1)!}\sum_{j_1=1}^{J} \cdots \sum_{\substack{j_{q-1}=1\\ j_{q-1}\notin\{j_1,\ldots,j_{q-2}\}}}^{J} \mathbb{P}(U_{ij_1} > U_{ij}, \ldots, U_{ij_{q-1}} > U_{ij}, U_{ij} \geq U_{il}, l \in \mathcal{I}\setminus\{j_1, \ldots, j_{q-1}, j\}|(u_{ik})_{k=1}^{J}, \eta_{ij} = s)f(s)ds$$

which results in:

$$J\mathbb{P}(U_{i,(q-1)} > U_{ij} \geq U_{i,(q)}|x_i, z_j) \longrightarrow \frac{\exp(U(x_i, z_j))}{1+\int \exp(U(x_i, s))m(s)ds} \times \left(\frac{\int \exp(U(x_i, s))m(s)ds}{1+\int \exp(U(x_i, s))m(s)ds}\right)^{q-1}$$

We can thus derive the following result:

$$J\mathbb{P}(U_{ij} \geq U_{i,(q)}|x_i, z_j) \longrightarrow \frac{\exp(U(x_i, z_j))}{1 + \int \exp(U(x_i, s))m(s)ds} \times \sum_{k=1}^{q} \left(\frac{\int \exp(U(x_i, s))m(s)ds}{1 + \int \exp(U(x_i, s))m(s)ds} \right)^{k-1}$$
$$= \exp(U(x_i, z_j)) \times \left[1 - \left(\frac{\int \exp(U(x_i, s))m(s)ds}{1 + \int \exp(U(x_i, s))m(s)ds} \right)^{q} \right]$$

$$\mathbb{P}(U_{i0} \geq U_{i,(q)}|x_i, z_j) \longrightarrow \frac{1}{1 + \int \exp(U(x_i, s))m(s)ds} \times \sum_{k=1}^{q} \left(\frac{\int \exp(U(x_i, s))m(s)ds}{1 + \int \exp(U(x_i, s))m(s)ds} \right)^{k-1}$$
$$= \left[1 - \left(\frac{\int \exp(U(x_i, s))m(s)ds}{1 + \int \exp(U(x_i, s))m(s)ds} \right)^{q} \right]$$

This concludes the proof of Proposition 2.

1.A.3 Proof of Theorem 1

I start by proving part (i) and (ii) of Theorem 1. As in Menzel (2015), I first restrict the space of functions to which the solutions to the fixed point problem described in Equation 1.6 can belong. Namely, I show that we can restrict ourselves to a Banach space of continuous functions.

Assume that there exists a pair of functions Γ^*_w and Γ^*_m that solve the fixed point problem. By definition of Ψ_w and using that $\Gamma^*_m \geq 0$, we have for any $q \geq 1$:

$$\Gamma^*_w(x) = \Psi[\Gamma^*_m](x) = \int \exp\{U(x, s) + V(x, s)\} \left[1 - \left(\frac{\Gamma^*_m(s)}{1 + \Gamma^*_m(s)} \right)^{q} \right] m(s)ds$$
$$\leq \int \exp\{U(x, s) + V(x, s)\} m(s)ds$$
$$\leq \exp\{\bar{U} + \bar{V}\}$$

where $\bar{U}$ and $\bar{V}$ are the upper bounds of the functions U and V, respectively. Given that this bound holds also for $q = 1$, this implies that we can bound similarly $\Gamma^*_m(x) \leq \exp\{\bar{U} + \bar{V}\}$. I now establish continuity of the solutions Γ^*_w and Γ^*_m. By definition, I rewrite:

$$\Gamma^*_w(x) = \Psi_w[\Psi_m[\Gamma^*_w]](x)$$

$$= \int \exp\{U(x,s)+V(x,s)\}\left[1-\left(\frac{\int \frac{\exp\{U(t,s)+V(t,s)\}}{1+\Gamma_w^*(t)}w(t)dt}{1+\int \frac{\exp\{U(t,s)+V(t,s)\}}{1+\Gamma_w^*(t)}w(t)dt}\right)^q\right]m(s)ds$$

Similarly, I write:

$$\begin{aligned}\Gamma_m^*(z) &= \Psi_m[\Psi_w[\Gamma_m^*]](z)\\ &= \int \frac{\exp\{U(t,z)+V(t,z)\}}{1+\int \exp\{U(t,s)+V(t,s)\}\left[1-\left(\frac{\Gamma_m^*(s)}{1+\Gamma_m^*(s)}\right)^q\right]m(s)ds}w(t)dt\end{aligned}$$

Since U and V are continuous and all the integrals are nonnegative, $\Psi_w[\Psi_m[\Gamma_w^*]]$ and $\Psi_m[\Psi_w[\Gamma_m^*]]$ are also continuous which establishes continuity of the solutions Γ_w^* and Γ_m^*. Differentiability of Γ_w^* and Γ_m^* also follows from differentiability of U and V which is stated in Assumption 1. We can thus restrict the spaces in which Γ_w^* and Γ_m^* belong to a Banach space of nonnegative bounded continuous functions that I call $\mathcal{C}^*$.

Consider now two pairs of functions (Γ_w, Γ_m) and $(\tilde{\Gamma}_w, \tilde{\Gamma}_m)$ belonging to $\mathcal{C}^* \times \mathcal{C}^*$. I first rewrite:

$$\log \Psi_w[\log \Gamma_m](x) = \int \exp\{U(x,s)+V(x,s)\}\left[1-\left(\frac{\exp\{\log \Gamma_m^*(s)\}}{1+\exp\{\log \Gamma_m^*(s)\}}\right)^q\right]m(s)ds$$

Given that Ψ_w and Ψ_m are Gâteaux differentiable, I use the mean value inequality to establish that:

$$\begin{aligned}&\left|\left|\log \Psi_w[\Gamma_m](x) - \log \Psi_w[\tilde{\Gamma}_m](x)\right|\right|_\infty\\ &\qquad\leq \sup_{a\in[0,1]}\left|\left|d\log \Psi_w[a\log \Gamma_m + (1-a)\log \tilde{\Gamma}_m](x)\right|\right|_\infty \left|\left|\log \Gamma_m(x) - \log \tilde{\Gamma}_m(x)\right|\right|_\infty\end{aligned}$$

where we can write:

$$d\log \Psi_w[\log \Gamma_m](x) = -\frac{1}{\Psi_w[\log \Gamma_m](x)}\int \frac{\exp\{U(x,s)+V(x,s)\}}{1+\Gamma_m(s)}\left(\frac{\Gamma_m(s)}{1+\Gamma_m(s)}\right)^q m(s)ds$$

Rearranging this expression gives the following:

$$\begin{aligned}d\log \Psi_w[\log \Gamma_m](x) = &-\frac{1}{\Psi_w[\log \Gamma_m](x)}\int \exp\{U(x,s)+V(x,s)\}\left[1-\left(\frac{\Gamma_m(s)}{1+\Gamma_m(s)}\right)^q\right]\\ &\times \frac{q(\Gamma_m(s))^q}{(1+\Gamma_m(s))^{q+1}-(\Gamma_m(s))^q-(\Gamma_m^*(s))^{q+1}}m(s)ds\end{aligned}$$

Since Γ^*_m has to be positive, we can show that:

$$\begin{aligned}
\frac{q(\Gamma_m(s))^q}{(1+\Gamma_m(s))^{q+1}-(\Gamma_m(s))^q-(\Gamma_m(s))^{q+1}} &= \frac{q(\Gamma_m(s))^q}{\sum_{k=0}^{q+1}\frac{(q+1)!}{k!(q+1-k)!}(\Gamma_m(s))^k-(\Gamma_m(s))^q-(\Gamma_m(s))^{q+1}} \\
&= \frac{q(\Gamma_m(s))^q}{\sum_{k=0}^{q-1}\frac{(q+1)!}{k!(q+1-k)!}(\Gamma_m(s))^k+q(\Gamma_m(s))^q} \\
&\leq \frac{q(\exp\{\bar{U}+\bar{V}\})^q}{\sum_{k=0}^{q-1}\frac{(q+1)!}{k!(q+1-k)!}(\exp\{\bar{U}+\bar{V}\})^k+q(\exp\{\bar{U}+\bar{V}\})^q} := \lambda_q
\end{aligned}$$

This implies that:

$$\begin{aligned}
|d\log\Psi_w[\log\Gamma_m](x)| &\leq \frac{\lambda_q}{\Psi_w[\log\Gamma_m](x)}\int \exp\{U(x,s)+V(x,s)\}\left[1-\left(\frac{\Gamma^*_m(s)}{1+\Gamma^*_m(s)}\right)^q\right]m(s)ds \\
&= \lambda_q \leq 1
\end{aligned}$$

From this, I conclude that:

$$\sup_{a\in[0,1]}\left|\left|d\log\Psi_w[a\log\Gamma_m+(1-a)\log\tilde{\Gamma}_m](x)\right|\right|_\infty \leq \lambda_q$$

which implies that:

$$\left|\left|\log\Psi_w[\Gamma_m](x)-\log\Psi_w[\tilde{\Gamma}_m](x)\right|\right|_\infty \leq \lambda_q\left|\left|\log\Gamma_m(x)-\log\tilde{\Gamma}_m(x)\right|\right|_\infty$$

Given that this holds for any $q \geq 1$, I conclude that the mapping $(\log\Gamma_w, \log\Gamma_m) \mapsto (\log\Psi_w[\Gamma_m], \log\Psi_m[\Gamma_w])$ is a contraction which proves claim (i) of Theorem 1. The proof of part (ii) is a direct implication of the Banach fixed point theorem.

Before proving part (iii) of Theorem 1, intermediary steps are needed. In what follows, I first prove that the size of opportunity sets grow at a rate $\sqrt{n}$ for any $q \geq 1$. From this, I then show that the dependence between opportunity sets and taste shocks under the extremal matchings vanishes as n grows to infinity. I then use this result to show that we can approximate inclusive values arising from any stable match by inclusive value functions which have an approximate fixed point representation. I then finally prove that the solution to the finite sample fixed point problem converges to the unique solution of the population fixed point problem which concludes the proof of Theorem 1.(iii).

Rate of Size of Feasible Choice Sets

Define, for a given stable matching μ^*, the number of firms feasible to worker i and the number of workers feasible to firm j as:

$$J_{wi}^* = \sum_{j=1}^{n_m} \mathbb{1}\{V_{ji} \geq V_{j,(q)}(W_j(\mu^*))\} \quad \text{and} \quad J_{mj}^* = \sum_{i=1}^{n_w} \mathbb{1}\{U_{ji} \geq U_{i,(1)}(M_i(\mu^*))\}$$

Similarly, define the number of firm that worker i would accept and the number of workers that firm j would accept:

$$L_{wi}^* = \sum_{j=1}^{n_m} \mathbb{1}\{U_{ji} \geq U_{i,(1)}(M_i(\mu^*))\} \quad \text{and} \quad L_{mj}^* = \sum_{i=1}^{n_w} \mathbb{1}\{V_{ji} \geq V_{j,(q)}(W_j(\mu^*))\}$$

To characterize the limit of the conditional matching probabilities, we need to know at which rate these objects grow. Menzel (2015) showed that for $q = 1$, we can bound each of these by quantities that grow at a rate $\sqrt{n}$. I show that this extends to any $q > 1$ by proving the following:

Lemma 1 *Under Assumptions 1 and 2 and for any stable matching μ^*, we have for any finite $q \geq 1$:*

$$n^{1/2} \frac{\exp(-\bar{V} + \gamma_m)}{1 + \exp(\bar{U} + \bar{V} + \gamma_w)} \leq J_{wi}^* \leq n^{1/2} \exp(\bar{V} + \gamma_m)$$

$$n^{1/2} \frac{\exp(-\bar{U} + \gamma_w)}{1 + \exp(\bar{U} + \bar{V} + \gamma_m)} \leq J_{mj}^* \leq n^{1/2} \exp(\bar{U} + \gamma_w)$$

$$n^{1/2} \frac{\exp(-\bar{U} + \gamma_m)}{1 + \exp(\bar{U} + \bar{V} + \gamma_m)} \leq L_{wi}^* \leq n^{1/2} \exp(\bar{U} + \gamma_m)$$

$$n^{1/2} \frac{\exp(-\bar{V} + \gamma_w)}{1 + \exp(\bar{U} + \bar{V} + \gamma_w)} \leq L_{mj}^* \leq n^{1/2} \exp(\bar{V} + \gamma_w)$$

for each $i = 1, ..., n_w$ and $j = 1, ..., n_m$ with probability approaching 1 as $n \to \infty$.

PROOF: I rely on two important observations:
(a). We can bound, for any $q \geq 1$, $V_{j,(q)}(W_j(\mu^*))$ from above by $V_{j,(1)}(W_j(\mu^*))$ and similarly $U_{i,(q)}(M_i(\mu^*))$ by $U_{i,(1)}(M_i(\mu^*))$ for all $i = 1, ..., n_w$ and $j = 1, ..., n_m$.
(b). As in Menzel (2015), we can define exogenous sets $\bar{W}_j = \{i : U_{ij} \geq U_{i0}\}$ and $\bar{M}_i = \{j :$

$V_{ij} \geq V_{0j}\}$ such that $W_j(\mu^*) \subset \bar{W}_j$ and $M_i(\mu^*) \subset \bar{M}_i$ as well as $W_j^\circ = \{i : U_{ij} \geq U_{i,(1)}(\bar{M}_i)\}$ and $M_{i,(q)}^\circ = \{j : V_{ij} \geq V_{j,(q)}(\bar{W}_j)\}$ such that $W_j^\circ \subset W_j(\mu^*)$ and $M_{i,(q)}^\circ \subset M_i(\mu^*)$.

A first important result is that (a) implies that for any $q > 1$, $M_{i,(1)}^\circ \subset M_{i,(q)}^\circ \subset M_i(\mu^*)$. From this, I construct the following bounds on J_{wi}^*:

$$J_{wi}^\circ = \sum_{j=1}^{n_m} \mathbb{1}\{j \in M_{i,(1)}^\circ\} \leq \sum_{j=1}^{n_m} \mathbb{1}\{j \in M_i(\mu^*)\} \leq \sum_{j=1}^{n_m} \mathbb{1}\{j \in \bar{M}_i\} = \bar{J}_{wi}$$

from there, using Proposition 2, we can show that:

$$\mathbb{E}[\bar{J}_{wi}|x_i, z_1, ..., z_{n_m}] = \frac{1}{J}\sum_{j=1}^{n_m} \frac{\exp\{V(x_i, z_j)\}}{1 + \frac{1}{J}\exp\{V(x_i, z_j)\}} + o(1) \leq \frac{n_m}{J}\exp\{\bar{U}\} + o(1)$$

which implies under Assumption 2 that:

$$\mathbb{E}[\bar{J}_{wi}] \leq n^{1/2}\exp\{\bar{V} + \gamma_m\} + o(1)$$

Following the same steps as Menzel (2015) we can then show that the variance of $\bar{J}_{wi}$ converges to zero which implies that:

$$n^{-1/2}(\bar{J}_{wi} - \mathbb{E}[\bar{J}_{wi}]) \to 0$$

We have thus established that $J_{wi}^* \leq n^{1/2}\exp\{\bar{V} + \gamma_m\}$ with probability approaching 1 as $n \to \infty$. Following the same steps, we can show symmetrically that:

$$J_{mj}^* \leq n^{1/2}\exp\{\bar{U} + \gamma_w\}$$

$$L_{wi}^* \leq n^{1/2}\exp\{\bar{V} + \gamma_m\}$$

$$L_{mj}^* \leq n^{1/2}\exp\{\bar{U} + \gamma_w\}$$

with probability approaching 1 as $n \to \infty$. We now consider the lower bound J_{wi}°. We can again use Proposition 2 to show that:

$$\begin{aligned}\mathbb{E}[J_{wi}^\circ|(x_l)_{l\in\bar{W}_j}, z_1, ..., z_{n_m}] &= \frac{1}{J}\sum_{j=1}^{n_m} \frac{\exp\{V(x_i, z_j)\}}{1 + \frac{1}{J}\sum_{l\in\bar{W}_j}\exp\{V(x_l, z_j)\}} + o(1) \\ &\geq \frac{n_m}{J}\frac{\exp\{-\bar{V}\}}{1 + \frac{\bar{J}_{mj}}{J}\exp\{\bar{V}\}} + o(1)\end{aligned}$$

Using the higher bound for J^*_{mj} derived just above and Jensen's inequality, we can finally show that:

$$\mathbb{E}[J^\circ_{wi}] \geq n^{1/2} \frac{\exp\{-\bar{V} + \gamma_m\}}{1 + \exp\{\bar{V} + \bar{U} + \gamma_w\}} + o(1)$$

Following Menzel (2015) we can then also show that the variance of J°_{wi} converges to zero which implies that:

$$n^{-1/2}(J^\circ_{wi} - \mathbb{E}[J^\circ_{wi}]) \to 0$$

This establishes that $J^*_{wi} \geq n^{1/2} \frac{\exp\{-\bar{V}+\gamma_m\}}{1+\exp\{\bar{V}+\bar{U}+\gamma_w\}}$ with probability approaching 1 as $n \to \infty$. Following the same steps, we can show that symmetrically, we have:

$$J^*_{mj} \geq n^{1/2} \frac{\exp\{-\bar{U} + \gamma_w\}}{1 + \exp\{\bar{V} + \bar{U} + \gamma_m\}}$$

$$L^*_{wi} \geq n^{1/2} \frac{\exp\{-\bar{U} + \gamma_m\}}{1 + \exp\{\bar{V} + \bar{U} + \gamma_m\}}$$

$$L^*_{mj} \geq n^{1/2} \frac{\exp\{-\bar{V} + \gamma_w\}}{1 + \exp\{\bar{V} + \bar{U} + \gamma_w\}}$$

with probability approaching 1 as $n \to \infty$. This concludes the proof of Lemma 1.

Exogeneity of Feasible Choice Sets

We now need to show that as $n \to \infty$, the dependence between agents taste shocks and opportunity sets vanishes. Again, a proof exists for $q = 1$ in Menzel (2015) but I show that this result extends to $q > 1$.

For the rest of the proof, I define the following set of indicator functions $E^*_{ij} = \mathbb{1}\{i \in W_j(\mu^*)\}$ and $D^*_{ij} = \mathbb{1}\{j \in M_i(\mu^*)\}$ for all workers $i = 1, ..., n_w$ and firms $j = 1, ..., n_m$. The first result to establish is that the probability that changing one availability indicator affects another agents' opportunity set converges to zero as $n \to \infty$ for any $q \geq 1$. I first prove the following result:

Lemma 2 *Suppose Assumption 1 and 2 hold and suppose we change one availability indicator E^*_{ij} exogenously to $\tilde{E}_{ij} = 1 - E^*_{ij}$ and then iterate the deferred acceptance algorithm from this point until convergence. Denote the resulting availability indicators $\{\tilde{E}_{lk}, \tilde{D}_{lk} : l = 1, ..., n_w, k = 1, ..., n_m\}$. We have for any $q \geq 1$ and any worker l and firm k:*

(i). $\mathbb{P}(\tilde{D}_l \neq D_l^* | D_l^*, D_{ij}^* = 0) = \mathbb{P}(\tilde{E}_k \neq E_k^* | E_l^*, D_{ij}^* = 0) = 0$

(ii). There exist constants $\bar{a} < \infty$ *and* $0 < \lambda < 1$ *such that:*

$$\mathbb{P}(\tilde{D}_l \neq D_l^* | D_l^*, D_{ij}^* = 1) \leq n^{-1/2} \frac{\bar{a}}{1-\lambda}$$

$$\mathbb{P}(\tilde{E}_k \neq E_k^* | E_l^*, E_{ij}^* = 1) \leq n^{-1/2} \frac{\bar{a}}{1-\lambda}$$

The same result holds for an exogenous change of D_{ij} *to* $\tilde{D}_{ij} = 1 - D_{ij}$.

PROOF: Suppose we change E_{ji}^* exogenously to $\tilde{E}_{ji} = 1 - E_{ji}$ and that we iterate the deferred acceptance algorithm from this stage. This will only trigger a chain of rematches if this affects the indirect utility of either i or j. Suppose $D_{ij}^* = 0$ and that $E_{ij}^* = 0$ meaning that firm j is not feasible to worker i and vice versa. Suppose now that $\tilde{E}_{ji} = 1 - E_{ij}^* = 1$, meaning that suddenly worker i's preference for firm j increase such that worker i becomes feasible for firm j. This will not affect the indirect utility of firm j nor worker i given that firm j is not feasible to worker i. This change will thus not trigger a chain of rematches. A similar argument can be used in the case where E_{ij}^* changes from 1 to $\tilde{E}_{ji} = 1 - E_{ij}^* = 0$. This establishes part (i) of Lemma 2 and does not depend on the value of q.

Now suppose that $D_{ij}^* = 1$ such that if $\tilde{E}_{ji} = 1 - E_{ij}^* = 1$, now firm j and worker i will want to rematch together or if $\tilde{E}_{ji} = 1 - E_{ij}^* = 0$ firm j and worker i will break their current match. This will trigger a chain of rematches than can potentially cycle back to worker i or firm j's opportunity set. I start by showing that, for any $q > 1$ at each step s of these subsequent rematches, there is at most one indicator in the vector $D_l^{(s)}$ corresponding to a firm k with $E_{lk}^{(s)} = 1$ that will change. The idea of the proof is the following: suppose that a given worker l matched to firm k in step $(s-1)$ becomes unavailable to firm k in step s. This firm will then replace this worker by the next q^{th} ranked feasible applicant, which will only change the availability indicator of this firm to this newly hired worker. On the other hand, if a given worker becomes available to a firm while this firm prefers this worker to one of its top q matched employees, then it will replace the q^{th} ranked worker by this new employee, making this firm unavailable to the kicked out employee. In both cases, this will only change at most one availability indicator among the workers who are willing to match with this firm. Note that at each of these steps, there is a chance that the chain is terminated if the next q^{th} ranked feasible worker is the outside option. A similar argument can be used symmetrically from the workers perspective by replacing q by 1.

The rest of the proof now consists in bounding the probability that the chain is terminated by either (a) firm k or worker l preferring the outside option to any other option in their opportunity set or (b) a change in availability indicators of worker k D_k. I define μ^s the state of the match in iteration s of the deferred acceptance algorithm following an exogenous change of E_{ji} to $\tilde{E}_{ji} = 1 - E_{ji}$. The first step bounds the probability that the chain is terminated by the outside option at stage s.

I start from the following observation: given that $\mathbb{P}(V_{lk} > V_{k,(q)}(W_k(\mu^s))|x_l, z_k) \geq \mathbb{P}(V_{lk} > V_{k,(1)}(W_k(\mu^s))|x_l, z_k)$ and that $W^{\circ}_{k,(1)} \subset W^*_k \subset \bar{W}_k$, we have from Proposition 2 and Lemma 1 that for any firm k and worker l:

$$
\begin{aligned}
\mathbb{P}(V_{lk} > V_{k,(q)}(W_k(\mu^s))|x_l, z_k) &\geq \mathbb{P}(V_{lk} > V_{k,(1)}(\bar{W}_k)|x_l, z_k) \\
&= n^{-1/2} \frac{\exp(V(z_k, x_l))}{1 + \frac{1}{J}\sum_{i \in \bar{W}_k} \exp(V(z_k, x_i))} + o(1) \\
&\geq n^{-1/2} \frac{\exp(V(z_k, x_l))}{1 + \exp(\bar{U} + \bar{V} + \gamma_w)} + o(1)
\end{aligned}
$$

This implies that, conditional on D^*_i and as n approaches infinity:

$$
\mathbb{P}(V_{0k} > V_{k,(q)}|D^*_i, x_i, z_k) \geq \frac{1}{1 + \exp(\bar{U} + \bar{V} + \gamma_w)} =: p_s
$$

Following now the same steps as Menzel (2015), we have, by Bayes law that:

$$
\mathbb{P}(V_{0k} > V_{k,(q)}|D^*_l, \tilde{D}^{(s)}_{lk} = 1, x_l, z_k) \geq \frac{\underline{L} p_s}{\bar{L}(1 - p_s) + \underline{L} p_s}
$$

where $\bar{L}$ and $\underline{L}$ are respectively the upper and lower bounds on L^*_{mj} taken from Lemma 1. From there, we finally get that:

$$
1 - \mathbb{P}(V_{0k} > V_{k,(q)}|D^*_l, \tilde{D}^{(s)}_{lk} = 1, x_l, z_k) \leq \frac{\bar{L}\exp(\bar{U} + \bar{V} + \gamma_w)}{\bar{L}\exp(\bar{U} + \bar{V} + \gamma_w) + \underline{L}} =: \lambda < 1
$$

This essentially means that the probability that the chain is not terminated at stage s is bounded away from 1.

Now we bound the probability that the chain leads to a change in D_l at stage s. We can thus bound the following probability using Proposition 2 and Lemma 1:

$$\begin{aligned}
&\mathbb{P}(V_{lk} > V_{k,(q)}(W_k(\mu^s))|x_l, z_k) \\
&\quad \leq \mathbb{P}(V_{lk} > V_{k,(q)}(W^\circ_{k,(1)})|x_l, z_k) \\
&\quad = n^{-1/2}\exp(V(z_k, x_l)) \left[1 - \left(\frac{\frac{1}{J}\sum_{i\in W^\circ_{k,(1)}} \exp(V(z_k, x_i))}{1 + \frac{1}{J}\sum_{i\in W^\circ_{k,(1)}} \exp(V(z_k, x_i))}\right)^q\right] + o(1) \\
&\quad \leq n^{-1/2}\exp(V(z_k, x_l)) \left[1 - \left(\frac{\frac{J^\circ_{mk}}{J}\exp(-\bar{V})}{1 + \frac{J^\circ_{mk}}{J}\exp(-\bar{V})}\right)^q\right] + o(1) \\
&\quad \leq n^{-1/2}\exp(V(z_k, x_l)) \left[1 - \left(\frac{\exp(-\bar{V} - \bar{U} + \gamma_w)}{1 + \exp(-\bar{V} - \bar{U} + \gamma_w) + \exp(\bar{U} + \bar{V} + \gamma_m)}\right)^q\right] + o(1) \\
&\quad \leq n^{-1/2}\exp(\bar{V}) + o(1)
\end{aligned}$$

This implies that for n sufficiently large, we have:

$$\begin{aligned}
&\mathbb{P}(\tilde{D}^{(s)}_l \neq D^*_l | D^*_l, \tilde{D}^{(s)}_{lk} = 1, x_l, z_k) \\
&\quad \leq \frac{n^{-1/2}\exp(\bar{V})\bar{L}}{n^{-1/2}\exp(\bar{V})\bar{L} + \underline{L}} \leq n^{-1/2}\exp(\bar{V})\frac{\bar{L}}{\underline{L}} = n^{-1/2}\bar{a}
\end{aligned}$$

Using the law of total probability, we can thus bound as $n \to \infty$ the conditional probability that $\tilde{D}_l \neq D^*_l$:

$$\mathbb{P}(\tilde{D}_l \neq D^*_l | D^*_l) \leq \sum_{s=1}^{\infty} \lambda^s n^{-1/2}\bar{a} \leq \frac{n^{-1/2}\bar{a}}{1-\lambda}$$

which proves part (b) of Lemma 2.

From there, I state the main result that the dependence between taste shocks and agents' opportunity sets vanishes as $n \to \infty$ for any $q \geq 1$. I first define the joint distribution of $\eta_i = (\eta_{i1}, ..., \eta_{in_m})'$, $\epsilon_j = (\epsilon_{1j}, ..., \epsilon_{n_wj})'$ and the availability indicators D^W_i, E^W_j, D^M_i, E^M_j corresponding to the worker-optimal and the firm-optimal stable matches. Note that I consider these two specific matches since the worker-optimal and firm-optimal stable matches are defined with probability 1 conditional on the realization of the taste shocks η_i and ϵ_j. Indeed, the distribution of availability indicators arising from an arbitrary stable match D^*_i would not be well defined. I also define: $D^W_{i,-j} = (D^W_{i1}, ..., D^W_{i(j-1)}, D^W_{i(j+1)}, ..., D^W_{in_m})$ and $E_{-i,j} = (E^W_{1j}, ..., E^W_{(i-1)j}, E^W_{(i+1)j}, ..., E^W_{n_wj})$ with analogous notations for the firm optimal match. I then define the conditional c.d.f.s:

$$G^W_{\eta|D}(\eta|\boldsymbol{d}) = \mathbb{P}(\eta_i \leq \eta | D^W_i = \boldsymbol{d}), \quad \boldsymbol{d} \in \{0,1\}^{n_m}$$

$$G^W_{\eta,\epsilon|D,E}(\eta,\epsilon|\boldsymbol{d},\boldsymbol{e}) = \mathbb{P}(\eta_i \leq \eta, \epsilon_j \leq \epsilon | D^W_{i,-j} = \boldsymbol{d}, E^W_{-i,j} = \boldsymbol{e}), \quad \boldsymbol{d} \in \{0,1\}^{n_m-1}, \boldsymbol{e} \in \{0,1\}^{n_w-1}$$

with analogous definitions for the firm-optimal stable match and associated p.d.f.s $g^W_{\eta|D}$ and $g^W_{\eta,\epsilon|D,E}$. The main result is the following:

Lemma 3 *Under Assumption 1 and 2, we have:*

(i). $g^W_{\eta|D}$ and $g^M_{\eta|D}$ satisfy:

$$\lim_n \Big| \frac{g^W_{\eta|D}(\eta|D^W_i)}{g_\eta(\eta)} - 1 \Big| = \lim_n \Big| \frac{g^M_{\eta|D}(\eta|D^M_i)}{g_\eta(\eta)} - 1 \Big| = 1$$

(ii). $g^W_{\eta,\epsilon|D,E}$ and $g^M_{\eta,\epsilon|D,E}$ satisfy:

$$\lim_n \Big| \frac{g^W_{\eta|D}(\eta,\epsilon|D^W_{i,-j}, E^W_{-i,j})}{g_{\eta,\epsilon}(\eta,\epsilon)} - 1 \Big| = \lim_n \Big| \frac{g^M_{\eta|D}(\eta,\epsilon|D^M_{i,-j}, E^M_{-i,j})}{g_{\eta,\epsilon}(\eta,\epsilon)} - 1 \Big| = 1$$

The same results holds for the firm side of the market.

PROOF: Let $g^W_{\eta,D}$ be the joint p.d.f. of taste shocks and availability indicators under the worker optimal stable match. We can rewrite, by definition of a conditional density:

$$\frac{g^W_{\eta|D}(\eta|D^W_i)}{g_\eta(\eta)} = \frac{g^W_{\eta,D}(\eta, D^W_i)}{g_\eta(\eta)P(D^W_i)} = \frac{P(D^W_i|\eta_i = \eta)g_\eta(\eta)}{g_\eta(\eta)P(D^W_i)} = \frac{P(D^W_i|\eta_i = \eta)}{P(D^W_i)}$$

I then follow similar steps as in Menzel (2015) to show that:

$$\Big| \frac{P(D^W_i|\eta_i = \eta)}{P(D^W_i)} - 1 \Big| \leq \sup_{\eta_1,\eta_2} \Big| \frac{P(D^W_i|\eta_i = \eta_1)}{P(D^W_i|\eta_i = \eta_2)} - 1 \Big|$$

such that I only need to bound the probability that shifting η_i from η_1 to η_2 changes worker i's opportunity set. This insight does not depend on q. We know from Lemma 2 that changing an availability indicator will trigger a chain of rematches that could change worker

i's opportunity set with probability less than $\frac{n^{-1/2}\bar{a}}{1-\lambda}$ as n approaches infinity. Here, we can show that shifting agent i's taste shocks would trigger at most two chains of rematches. Indeed, if the shift in taste shocks makes agent i prefers firm l with $D_{il} = 1$ instead of her current employer firm j, this changes both E_{ij} from 1 to 0 and E_{il} from 0 to 1. Thus, this would trigger two chains of rematches where both firm j and the worker which was displaced from firm l by worker i would need to find a new match. We can thus conclude that:

$$\frac{P(D_i^W|\eta_i = \eta_1)}{P(D_i^W|\eta_i = \eta_2)} - 1 \leq 2\frac{n^{-1/2}\bar{a}}{1-\lambda}$$

as $n \to \infty$ which can be shown to hold also in absolute value. As the right hand side converges to 0 as $n \to \infty$, this proves the first part of claim (i). Now consider the symmetrical case where we would shift firm j's taste shocks. Following a similar argument, we can see that this would create at most $2q$ chains of rematches. Indeed, assuming that such a shift in firm j's taste shocks would make it want to replace all of its q employees, this implies that the q workers which were let go along with the (potentially) q firms which lost one of their employees would need to find a new match. This implies that:

$$\frac{P(E_j^W|\epsilon_j = \epsilon_1)}{P(E_j^W|\epsilon_j = \epsilon_2)} - 1 \leq qn^{-1/2}\frac{2\bar{a}}{1-\lambda}$$

as $n \to \infty$ which can be shown to hold also in absolute value. As the right hand side converges to 0 as $n \to \infty$, this proves the first part of claim (i).

For part (ii), note that the argument can be extended in a similar way. If you change both firm j and worker i's taste shocks this can trigger at most $2(q+1)$ chains of rematches such that we can bound the probability of a shift in opportunity sets by $(q+1)n^{-1/2}\frac{2\bar{a}}{1-\lambda}$ which can be made arbitrarily close to 0 as n approaches infinity.

Bounds for Inclusive Values

Since I have established exogeneity of opportunity sets under the firm-optimal and worker-optimal stable matches, the rest of the analysis focuses on characterizing the limit of inclusive values that arise under these extremal matchings. As in Menzel (2015), I show that both converge to a unique limit, implying that inclusive values arising from any stable matching also converge towards this limit.

I define $I^W_{wi} = I_{wi}(\mu^W)$ and $I^W_{mj} = I_{mj}(\mu^W)$ the inclusive values that arise from the worker-optimal stable match. Similarly, I define I^M_{wi} and I^M_{mj} as the inclusive values that arise from the firm-optimal stable match such that for any stable match μ^*, we have $I^W_{wi} \geq I_{wi}(\mu^*) \geq I^M_{wi}$ and $I^W_{mj} \leq I_{wi}(\mu^*) \leq I^M_{mj}$. I state the following result:

Lemma 4 *Under Assumption 1 and 2:*
(i). For all $i = 1, ..., n_w$ and $j = 1, ..., n_m$:

$$I^M_{wi} \geq \hat{\Gamma}^M_{wn}(x_i) + o_p(1) \quad \textit{and} \quad I^M_{mj} \leq \hat{\Gamma}^M_{mn}(z_j) + o_p(1)$$

where the analogous result holds for the worker-optimal stable match with the side of inequalities reversed.
(ii). If the weight functions $\omega(x, z) \geq 0$ are bounded and form a Glivenko-Cantelli class in x, then

$$\sup_{x \in \mathcal{X}} \frac{1}{n} \sum_{j=1}^{n_m} \omega(x, z_j)(I^M_{mj} - \hat{\Gamma}^M_m(z_j)) \leq o_p(1)$$

and

$$\inf_{z \in \mathcal{Z}} \frac{1}{n} \sum_{i=1}^{n_w} \omega(x_i, z)(I^M_{wi} - \hat{\Gamma}^M_w(x_i)) \geq o_p(1)$$

The analogous conclusion holds for the worker-optimal stable match where the sign of the inequalities is reversed and if $\omega(x, z) \geq 0$ are bounded and form a Glivenko-Cantelli class in z.

Proof: I first show that we can bound conditional choice probabilities given an opportunity set arising from a stable match using the extremal matchings. I first define the conditional probability that worker i chooses firm j given the realization of opportunity set M^M arising from the firm-optimal stable match:

$$\Lambda^M_w(x, z, M^M) = \mathbb{P}(U_{ij} \geq U_{i,(1)}(M^M_i) | M^M_i = M^M, x_i = x, z_j = z)$$

Similarly, I define the equivalent object from firm j's perspective:

$$\Lambda^M_m(x, z, W^M) = \mathbb{P}(V_{ij} \geq V_{j,(q)}(W^M_j) | W^M_j = W^M, x_i = x, z_j = z)$$

I also define the conditional choice probabilities under exogenous opportunity sets as:

$$\Lambda_w(x, z, M) = \mathbb{P}(U_{ij} \geq U_{i,(1)}(M)|x_i = x, z_j = z)$$

$$\Lambda_m(x, z, W) = \mathbb{P}(V_{ij} \geq V_{i,(q)}(W)|x_i = x, z_j = z)$$

As there are several stable matches such that $M_i^* = M_i^M$ and $W_j^* = W_j^M$ we can show that:

$$J\Lambda_w^M(x, z, M_i^M) \leq J\Lambda_w(x, z, M_i^M) + o_p(1)$$

$$J\Lambda_m^M(x, z, W_j^M) \geq J\Lambda_m(x, z, W_j^M) + o_p(1)$$

Similarly, we have:

$$J\Lambda_w^W(x, z, M_i^W) \geq J\Lambda_w(x, z, M_i^W) + o_p(1)$$

$$J\Lambda_m^W(x, z, W_j^W) \leq J\Lambda_m(x, z, W_j^W) + o_p(1)$$

Using Proposition 2, we can then show that for $i = 1, ..., n_w$, $l_1 = 1, ..., n_m$ and $l_2 \neq l_1$:

$$E[J(D_{il_1}^M - \Lambda_m^M(x_i, z_{l_1}, I_{ml_1}^M))|I_{ml_1}^M, x_i, z_{l_1}] \to 0$$

and

$$E[J^2(D_{il_1}^M - \Lambda_m^M(x_i, z_{l_1}, I_{ml_1}^M))(D_{il_1}^M - \Lambda_m^M(x_i, z_{l_2}, I_{ml_2}^M))|I_{ml_1}^M, I_{ml_2}^M, x_i, z_{l_1}, z_{l_2}] \to 0$$

Therefore, since under Assumption 1, we know that $\exp(U(x_i, z_j))$ is bounded, we can thus conclude that:

$$\text{Var}\left(\frac{1}{n}\sum_{k=1}^{n_m} \exp\{U(x_i, z_k)\}J(D_{ik}^M - \Lambda_m^M(x_i, z_k, I_{mk}^M))\right) \to 0$$

which implies that:

$$\frac{1}{n}\sum_{k=1}^{n_m} \exp\{U(x_i, z_k)\}J(D_{ik}^M - \Lambda_m^M(x_i, z_k, I_{mk}^M)) = o_p(1)$$

Given that from Proposition 2:

$$J\Lambda_m^M(x,z,W_j^M) \geq \exp\{V(x,z)\}\left[1-\left(\frac{I_{mj}^M}{1+I_{mj}^M}\right)^q\right]+o_p(1)$$

This implies that:

$$\frac{1}{n}\sum_{k=1}^{n_m}\exp\{U(x_i,z_k)\}\left(JD_{ik}^M-\exp\{V(x_i,z_k)\}\left[1-\left(\frac{I_{mk}^M}{1+I_{mk}^M}\right)^q\right]\right)\geq o_p(1)$$

which proves the first claim of part (i) of Lemma 4. From firm j's perspective, I show using the same arguments that:

$$\frac{1}{n}\sum_{l=1}^{n_w}\exp\{V(x_l,z_j)\}J(E_{lj}^M-\Lambda_w^M(x_l,z_j,I_{wl}^M))=o_p(1)$$

which implies, using Proposition 2:

$$\frac{1}{n}\sum_{l=1}^{n_w}\exp\{V(x_l,z_j)\}\left(JE_{lj}^M-\frac{\exp\{U(x_l,z_j)\}}{1+I_{wl}^M}\right)\geq o_p(1)$$

This prove part (i) of Lemma 4. Note that analogous arguments can be used to bound inclusive values arising from the worker-optimal stable match.

Part (ii) follows from part (i) of the Lemma and the boundedness condition on ω which implies pointwise convergence. The Glivenko-Cantelli condition on ω then implies uniform convergence. This concludes the proof of Lemma 4.

The next step consists in establishing uniform convergence with respect to $\Gamma_w \in \mathcal{T}_w$ and $\Gamma_m \in \mathcal{T}_m$ of the fixed point mappings $\hat{\Psi}_w$ and $\hat{\Psi}_m$ to their population counterparts. I define:

$$\hat{\Psi}_w[\Gamma_m](x)=\frac{1}{n}\sum_{j=1}^{n_m}\psi_w(z_j,x;\Gamma_m)$$

where ψ_w is defined as:

$$\psi_w(z_j,x;\Gamma_m)=\exp\{U(x,z_j)+V(x,z_j)\}\left[1-\left(\frac{\Gamma_m(z_j)}{1+\Gamma_m(z_j)}\right)^q\right]$$

Similarly, I define:

$$\hat{\Psi}_m[\Gamma_w](z) = \frac{1}{n}\sum_{i=1}^{n_w} \psi_m(z, x_i; \Gamma_w)$$

where ψ_m is defined as:

$$\psi_m(z, x_i; \Gamma_w) = \frac{\exp\{U(x_i, z) + V(x_i, z)\}}{1 + \Gamma_w(x_i)}$$

I define the class of functions $\mathcal{F}_w : \{\psi_w(., x; \Gamma_m) : x \in \mathcal{X}, \Gamma_m \in \mathcal{T}_m\}$ and $\mathcal{F}_m : \{\psi_m(., x; \Gamma_w) : x \in \mathcal{X}, \Gamma_w \in \mathcal{T}_w\}$.

Lemma 5 *Under Assumption 1:*
(i). The classes of functions $\mathcal{F}_w$ and $\mathcal{F}_w$ are Glivenko-Cantelli.
(ii). As $n \to \infty$:

$$(\hat{\Psi}_w[\Gamma_m](x), \hat{\Psi}_m[\Gamma_w](z)) \to (\Psi_w[\Gamma_m](x), \Psi_m[\Gamma_w](z))$$

uniformly in $\Gamma_w \in \mathcal{T}_w$, $\Gamma_m \in \mathcal{T}_m$ and $(x, z) \in \mathcal{X} \times \mathcal{Z}$.

PROOF: Under Assumption 1, $\exp\{U(x, z) + V(x, z)\}$ is Lipschitz in x and z such that this class of functions is Glivenko-Cantelli. Γ_m and Γ_w are bounded and have bounded $p \geq 1$ derivatives which makes the class of functions $\mathcal{F}_w \cup \mathcal{F}_m$ Glivenko-Cantelli. Finally, note that the transformation $\psi_m(g, h) = \frac{g}{1+h}$ is bounded and continuous since h and g are bounded and continuous and $h \geq 0$. Similarly, the transformation $\psi_w(g, h) = g\left[1 - \left(\frac{h}{1+h}\right)^q\right]$ is also bounded and continuous for any $q \geq 1$. Theorem 3 in van der Vaart and Wellner (2000) implies claim (i) of Lemma 5. Part (ii) of Lemma 5 is a direct implication of part (i).

Proof of Theorem 3.1 (iii)

I finally turn to the proof of part (iii) of Theorem 1. I first apply Lemma 4 to show that for any $q \geq 1$:

$$\begin{aligned}
\hat{\Gamma}_w^M(x) &= \frac{1}{n}\sum_{j=1}^{n_m} \exp\{U(x, z_j) + V(x, z_j)\}\left[1 - \left(\frac{I_{mj}^M}{1 + I_{mj}^M}\right)^q\right] \\
&\geq \frac{1}{n}\sum_{j=1}^{n_m} \exp\{U(x, z_j) + V(x, z_j)\}\left[1 - \left(\frac{\hat{\Gamma}_m^M(z_j)}{1 + \hat{\Gamma}_m^M(z_j)}\right)^q\right] + o_p(1)
\end{aligned}$$

Similarly, I show that:

$$\hat{\Gamma}_m^M(z) \leq \frac{1}{n}\sum_{i=1}^{n_w} \frac{\exp\{U(x_i, z) + V(x_i, z)\}}{1 + \hat{\Gamma}_w^M(x_i)} + o_p(1)$$

Analogous bounds can be formed for the inclusive value functions of the worker-optimal stable match. We thus have that:

$$\hat{\Gamma}_w^M \geq \hat{\Psi}_w^M[\hat{\Gamma}_m^M] + o_p(1) \quad \text{and} \quad \hat{\Gamma}_m^M \leq \hat{\Psi}_m^M[\hat{\Gamma}_w^M] + o_p(1)$$

$$\hat{\Gamma}_w^W \leq \hat{\Psi}_w^W[\hat{\Gamma}_m^W] + o_p(1) \quad \text{and} \quad \hat{\Gamma}_m^W \geq \hat{\Psi}_m^W[\hat{\Gamma}_w^W] + o_p(1)$$

Given that $\hat{\Psi}_w[\Gamma_m]$ and $\hat{\Psi}_m[\Gamma_w]$ are nonincreasing and Lipschitz continuous in Γ_m and Γ_w, we have:

$$\hat{\Gamma}_w^M \geq \hat{\Psi}_w^M[\hat{\Gamma}_m^M] + o_p(1) \geq \hat{\Psi}_w^M[\hat{\Psi}_m^M[\hat{\Gamma}_w^M]] + o_p(1)$$

Thus for any pair of functions (Γ_w^*, Γ_m^*) solving the fixed point problem:

$$\Gamma_w^* = \hat{\Psi}_w[\Gamma_m^*] + o_p(1) \quad \text{and} \quad \Gamma_m^* = \hat{\Psi}_m[\Gamma_w^*] + o_p(1)$$

we thus have:

$$\hat{\Gamma}_w^M \geq \Gamma_w^* + o_p(1) \quad \text{and} \quad \hat{\Gamma}_m^M \leq \Gamma_m^* + o_p(1)$$

However, we know that the mapping $\hat{\Psi}$ is a contraction in logs, which means that it has a unique fixed point (Γ_w^*, Γ_m^*). We also know, by definition, that:

$$\hat{\Gamma}_w^M \leq \hat{\Gamma}_w^W \quad \text{and} \quad \hat{\Gamma}_m^M \geq \hat{\Gamma}_m^W$$

which implies that:

$$\Gamma_w^* + o_p(1) \geq \hat{\Gamma}_w^W \geq \hat{\Gamma}_w^M \geq \Gamma_w^* + o_p(1)$$

$$\Gamma_m^* + o_p(1) \leq \hat{\Gamma}_m^W \leq \hat{\Gamma}_m^M \leq \Gamma_m^* + o_p(1)$$

which in turn implies that:

$$\hat{\Gamma}_w^M = \Gamma_w^* + o_p(1) \quad \text{and} \quad \hat{\Gamma}_m^M = \Gamma_m^* + o_p(1)$$

$$\hat{\Gamma}_w^W = \Gamma_w^* + o_p(1) \quad \text{and} \quad \hat{\Gamma}_m^W = \Gamma_m^* + o_p(1)$$

Combining this with Lemma 3, this gives us for all $i = 1, ..., n_w$ and all $j = 1, ..., n_m$:

$$I_{wi}^M = \Gamma_w^* + o_p(1) \quad \text{and} \quad I_{mj}^M = \Gamma_m^* + o_p(1)$$

$$I_{wi}^W = \Gamma_w^* + o_p(1) \quad \text{and} \quad I_{mj}^M = \Gamma_m^* + o_p(1)$$

Note that given that inclusive value functions that would arise under any stable match μ^* defined as I_{wi}^* and I_{mj}^* are such that $I_{wi}^M \leq I_{wi}^* \leq I_{wi}^W$ and $I_{mj}^M \geq I_{mj}^* \geq I_{mj}^W$ the equality written above holds also for any I_{wi}^* and I_{mj}^*.

I have shown that inclusive values can be approximated by the solution of the finite sample fixed point problem. Lemma 5 finally implies that the solution of the finite sample fixed point problem converges towards the solution of its population equivalent. This proves Theorem 1.(iii).

1.A.4 Proof of Proposition 3

Assume that firm j is matched with a group of $k = 2$ workers and that we want to characterize the limit of the following CCP:

$$\mathbb{P}(\mu_m(j) = \{i, l\} \cup \{0\}^{q-2} | x_i, x_l, z_j)$$

We can rewrite it as follows:

$$\begin{aligned}
&\mathbb{P}(U_{ij} \geq U_{i,(1)}(M_i(\mu^*)), U_{lj} \geq U_{l,(1)}(M_l(\mu^*)), V_{ij} > V_{lj} > V_{0j} \geq V_{j,(3)}(W_j(\mu^*)) | x_i, x_l, z_j) \\
&\quad + \mathbb{P}(U_{ij} \geq U_{i,(1)}(M_i(\mu^*)), U_{lj} \geq U_{l,(1)}(M_l(\mu^*)), V_{lj} > V_{ij} > V_{0j} \geq V_{j,(3)}(W_j(\mu^*)) | x_i, x_l, z_j) \\
=&\mathbb{P}(U_{ij} \geq U_{i,(1)}(M_i(\mu^*)) | x_i, z_j) \times \mathbb{P}(U_{lj} \geq U_{i,(1)}(M_i(\mu^*)) | x_l, z_j) \\
&\quad \times \left[\mathbb{P}(V_{ij} > V_{lj} > V_{0j} \geq V_{j,(3)}(W_j(\mu^*)) | x_i, x_l, z_j) + \mathbb{P}(V_{lj} > V_{ij} > V_{0j} \geq V_{j,(3)}(W_j(\mu^*)) | x_i, x_l, z_j)\right]
\end{aligned}$$

We can then decompose the rank ordered CCPs as follows:

$$\begin{aligned}
&\mathbb{P}(V_{ij} > V_{lj} > V_{0j} \geq V_{j,(3)}(W_j(\mu^*)) | x_i, x_l, z_j) \\
=&\mathbb{P}(V_{ij} \geq V_{j,(1)}(W_j(\mu^*)) | x_i, x_l, z_j) \times \mathbb{P}(V_{lj} \geq V_{j,(1)}(W_j(\mu^*) \setminus i) | x_i, x_l, z_j)
\end{aligned}$$

$$\times \mathbb{P}(V_{0j} \geq V_{j,(1)}(W_j(\mu^*) \setminus \{i, l\})|x_i, x_l, z_j)$$

In the limit, removing one arbitrary alternative from opportunity sets does not affect inclusive values:

$$\begin{aligned} n^{-1/2} \sum_{i \in W_j(\mu^*) \setminus \{l\}} \exp\{V(x_i, z_j)\} &= n^{-1/2} \sum_{i \in W_j(\mu^*)} \exp\{V(x_i, z_j)\} - n^{-1/2} \exp\{V(x_l, z_j)\} \\ &= n^{-1/2} \sum_{i \in W_j(\mu^*)} \exp\{V(x_i, z_j)\} + o_p(1) = I^*_{mj} + o_p(1) \end{aligned}$$

We can thus conclude that:

$$\begin{aligned} &\mathbb{P}(V_{ij} > V_{lj} > V_{0j} \geq V_{j,(3)}(W_j(\mu^*))|x_i, x_l, z_j) \\ =&\mathbb{P}(V_{ij} \geq V_{j,(1)}(W_j(\mu^*))|x_i, x_l, z_j) \times \mathbb{P}(V_{lj} \geq V_{j,(1)}(W_j(\mu^*))|x_i, x_l, z_j) \\ &\times \mathbb{P}(V_{0j} \geq V_{j,(1)}(W_j(\mu^*))|x_i, x_l, z_j) + o(1) \end{aligned}$$

Which, using Proposition 2 and Theorem 1, implies that:

$$n^2 \mathbb{P}(\mu_m(j) = \{i, l\} \cup \{0\}^{q-2}|x_i, x_l, z_j) \longrightarrow \frac{2 \exp\{U(x_i, z_j) + U(x_l, z_j) + V(x_i, z_j) + V(x_l, z_j)\}}{(1 + \Gamma^*_w(x_i))(1 + \Gamma^*_w(x_l))(1 + \Gamma^*_w(z_j))^3}$$

where Γ^*_w and Γ^*_m are the solutions of the fixed point problem described in Equation 1.6. To extend to proof to any k, a similar argument applies, except that the number of rank ordered CCPs becomes $k!$. This proves part (i) of Proposition 3.

A similar argument can be used to prove part (ii).

1.A.5 Proof of Proposition 6

From Theorem 1 and Proposition 2, we know that for any $q \geq 1$ and for a given finite w:

$$n^{-1/2} \mathbb{P}(U_{ij} \geq U_{i,(1)}(M_i(\mu^*))|x_i = x, z_j = z, w_i = w) \to \frac{\exp\left(U(x, z)\right)}{1 + \Gamma^*_w(x, w)}$$

and

$$J \mathbb{P}(V_{ij} \geq V_{j,(1)}(W_j(\mu^*))|x_i = x, z_j = z, w_i = w) \to \frac{\exp\left(V(x, z) + g(w)\right)}{1 + \Gamma_m(z)}$$

where Γ_m^* and Γ_w^* solve the following fixed point problem:

$$\Gamma_w^* = \Psi_w[\Gamma_m^*] \quad \text{and} \quad \Gamma_m^* = \Psi_m[\Gamma_w^*] \tag{1.7}$$

where

$$\Psi_w[\Gamma_m](x,w) = \int \exp(U(x,s) + V(x,s) + g(w) + \gamma_m) \times \left[1 - \left(\frac{\Gamma_m(s)}{1+\Gamma_m(s)}\right)^q\right] m(s)ds$$

$$\Psi_m[\Gamma_w](x) = \int\int \frac{\exp(U(s,z) + V(s,z) + g(t) + \gamma_w)}{1+\Gamma_w(s,t)} w_x(s)w_w(t)ds$$

However, as w goes to infinity, we have that:[14]

$$\lim_{w\to\infty} \mathbb{P}(V_{ij} \geq V_{j,(q)}(W_j(\mu^*))|x_i = x, z_j = z, w_i = w) = 1$$

This implies that:

$$\lim_{w_i\to\infty} n^{-1/2} I_{wi} = \int \exp\{U(x,s) + \gamma_m\} m(s)ds$$

which in turn implies that:

$$\lim_{w\to\infty} n\mathbb{P}(U_{ij} \geq U_{i,(1)}(M_i(\mu^*))|x_i = x, z_j = z, w_i = w) = \frac{\exp\left(U(x,z)\right)}{1+\int \exp\{U(x,s)+\gamma_m\}m(s)ds}$$

Similarly,

$$\lim_{w\to\infty} n\mathbb{P}(U_{i0} \geq U_{i,(1)}(M_i(\mu^*))|x_i = x, z_j = z, w_i = w) = \frac{1}{1+\int \exp\{U(x,s)+\gamma_m\}m(s)ds}$$

Taking the log of these ratios separately identifies U from the joint surplus. Given that the joint surplus is identified for finite w, we can then recover $V + g$. V can be separately identified by evaluating $V + g$ at $\bar{w}$.

[14]The probability that an option is in workers' opportunity sets only goes to 1 when making the shifter w go to infinity. In He et al. (2021) and Agarwal and Somaini (2022) this is not the case as cutoffs are fixed and finite since the number of "colleges" or "products" is fixed. As both the number of firms and workers grow to infinity in our case, the cutoffs grow to infinity as the size of the market grows.

Chapter 2

Teacher Compensation and Structural Inequality: Evidence from Centralized Teacher School Choice in Peru

with Matteo Bobba, Gianmarco Leon-Ciliotta, Christopher Neilson and Marco Nieddu

Abstract. This paper studies how increasing teacher compensation at hard-to-staff schools can reduce inequality in access to qualified teachers. Leveraging an unconditional change in the structure of teacher compensation in Perú, we first show causal evidence that increasing salaries at less desirable locations attracts teachers who score 0.45 standard deviations higher in standardized competency tests, leading to an average increase in student test scores of 0.33-0.38 standard deviations. We then estimate a model of teacher preferences over local amenities, school characteristics, and wages using geocoded job postings and rich application data from the nationwide centralized teacher assignment system. A policy that sets compensation at each job posting taking into account teacher preferences is more cost-effective than the actual policy in terms of reducing structural inequality in access to learning opportunities, and it possibly enhances the efficiency of the education system.

2.1 Introduction

Children born in remote and rural communities face significant disadvantages in achieving comparable levels of academic achievement as their peers born in urban areas (World Bank, 2018). Part of these wide inter-regional disparities reflect structural and historically persistent differences across geographic areas. Current policies can contribute to further widening the gap in the formation of human capital if the pre-existing inequality is not compensated for (Glewwe and Muralidharan, 2016). Key policies that reduce inequalities in access to key, high quality, educational inputs, i.e., teachers, are at the forefront of policy and academic discussions, yet mechanisms designed to ensure that high performing teachers accept positions in disadvantaged areas are still relatively under explored.[1]

In this paper, we study how inequality in the access to learning opportunities is amplified or reduced by policies that shape the geographic distribution of teachers. We shed light on this question in the context of Perú, a developing country with a heterogeneous geography and a population that is characterized by different languages, cultures, and ethnicities. In this diverse setting, rigidities that affect wage setting in the public sector (e.g., collective bargaining agreements) leads teachers sort on non-pecuniary aspects of employment (Rosen, 1986) thereby reinforcing the stark inequality in school inputs across rural and urban communities.

We study the impact of a recent policy reform that significantly increased compensation at hard-to-staff schools in rural areas. We use a combination of a regression discontinuity design and an empirical model of teacher school choice to characterize the effects of the policy and the underlying mechanisms through which it affects teacher sorting. Higher compensation at rural schools increases the supply of qualified teachers and improve student learning outcomes, and the gains are larger for under-achieving students. Despite the positive effects of increased compensation, we show that the current policy is both inefficient –since it fails to account account for teachers' heterogenous preferences over job postings– and not large enough to effectively undo the inequality of initial conditions that hard-to-staff schools and their communities face.

We then turn to investigate whether it is possible to redesign the compensation policy to achieve a more equitable allocation of teachers at a lower cost. To do so, we use the

[1] There is ample evidence that teachers matter for student outcomes in e.g., the US (Chetty et al., 2014a; Jackson, 2018), Ecuador (Araujo et al., 2016), Pakistan (Bau and Das, 2020) and Uganda (Buhl-Wiggers et al., 2017).

estimates of teacher preferences and information on school vacancies to characterize a menu of counterfactual wage schedules that can further enhance teachers sorting toward disadvantaged locations without modifying the current assignment mechanism. We finally use our framework to assess the relative cost-effectiveness of additional policy instruments that may complement wage incentives, such as investing in rural school/community infrastructure or increasing the local supply of teachers in disadvantaged areas.

We begin our empirical analysis by presenting descriptive evidence on the structural divide in school inputs and academic outcomes between rural and urban areas. Administrative data on school infrastructure, teacher qualifications, and student achievement show large gaps between rural and urban schools. Job amenities (or the lack thereof) is one of the main reason why rural schools have limited capacity to attract competent teachers. Using detailed data of teachers' preferences over job postings, we provide direct supporting evidence for this hypothesis and show that job applications are highly skewed towards vacancies in urban areas. Conversely, the school system is hard-pressed to staff many small rural public schools scattered throughout the poorest parts of the country. The scarcity of teachers applying for jobs at rural schools could be due to several factors, including insufficient compensation (Jackson et al., 2014).

Against this backdrop, the government implemented a policy that increased compensation for teaching positions in rural public schools, and its allocation was based on a coarse set of school and community attributes. We exploit discrete jumps in teacher compensation at specific thresholds of the local population to show causal evidence that, while higher wages significantly increased the demand for vacancies in rural locations, this did not translate into a higher share of vacancies being filled. Importantly though, teachers who chose positions that offer higher wages have scored significantly higher (0.45σ) on the national competency test, compared to those who chose lower-paying positions in otherwise similar teaching positions. We show that the increase in teacher quality in high-wage schools does not come at the expense of a reduction of qualified teachers just below the eligibility threshold for the rural wage bonus (or more broadly from other schools that don't offer the wage bonus, see Sections 2.4.5 and 2.5.4). Importantly, these effects are observed only for teachers who were assigned through a centralized mechanism that follows transparent and strict priority rules (contract teachers). As found in other settings (Duflo et al., 2015; Estrada, 2019), the local institutions determining how teachers are evaluated and assigned are an important necessary condition for the effectiveness of human resource policies in the public sector.

The increase in teacher compensation also spurs better student academic achievement in math and language (0.3σ and 0.35σ, respectively). This is only true for schools that had an available teaching position in the recruitment drives we analyze, even though incumbent teachers were also paid more. This evidence confirms that that higher wages do not prompt an effort response from incumbent teachers, mirroring recent findings that show that unconditional wage increases do not affect student outcomes in a setting where most teachers are public servants with permanent contracts (de Ree et al., 2018). In our context, instead, a large proportion of school vacancies targeted by the wage reform are filled by contract teachers, which is a common feature of teacher labor markets in Latin America and Africa. This feature creates significant flexibility in the labor market thus allowing wage incentives to play an important role in attracting higher quality teachers and consequently improving student outcomes. Higher wages have a more pronounced effect on reducing the proportion of students who score in the the first two deciles of the test score distribution, while the effects are smaller in magnitude and relatively uniform for better-performing students. This evidence suggests that policies that incentivize teachers sorting toward disadvantaged areas can be effective from an equity and efficiency point of view.

Our regression-discontinuity estimates provide credible causal evidence on the local effects of the policy, which may or may not hold more generally (e.g., in the presence of other wage bonuses and/or equilibrium sorting effects). To evaluate the policy away from the eligibility cutoffs, we estimate an empirical model of teacher school choice and construct counterfactual assignments in the absence of the policy. We estimate the model parameters taking advantage of teachers' revealed preferences observed in the contract-teacher assignment mechanism. The system follows a serial dictatorship algorithm where job applicants are ranked by their competency scores and sequentially assigned to their preferred school among those that still have an open vacancy. Together with detailed information on every school vacancy, teacher characteristics, and final assignments, this setting is ideal for estimating a flexible model of heterogeneous teacher preferences over wages and job attributes (Agarwal and Somaini, 2020). The model is able to replicate the main features of the data, including the spatial sorting of teachers, the local effects around the wage discontinuity, and broader trends along the support of the variables that characterize rural areas (e.g., locality population and proximity to the provincial capital). The estimated preference parameters quantify key trade-offs between wages and local amenities, school characteristics, teacher-school match effects, or moving costs. Importantly, teachers belonging to ethnic minorities who predominantly reside in rural areas are more willing to work at schools in communities from their own ethnolinguistic group, thus requiring a lower compensation to staff these positions.

The model of teacher school choice provides a rich perspective on the effects of the recent reform of the wage schedule on teacher's spatial distribution. By comparing simulated assignment outcomes with and without the increase in compensation, we can characterize the policy effects on teacher sorting away from the discontinuities generated by the eligibility cutoffs of the reform. We show that, while most of the impact of the policy on teacher quality happens close to the population threshold, its effect on the share of filled vacancies seems spatially concentrated in less desirable locations that are farther away from the cutoffs. The evidence drawn from the estimated model also indicates that wage bonuses generate, on net, a positive reallocation effect across the entire country, which is explained by the inflow of applicants who are matched to a school vacancy due to the wage incentives.

The changes in predicted teachers' utility with and without the increase in compensation are markedly heterogenous within geographic areas that pay the same wage, indicating large scope for improvement in the targeting design of the current policy. We thus turn to the evaluation of alternative compensation schemes using a matching-with-contracts framework (Hatfield and Milgrom, 2005). In this framework, we maintain the allocation mechanism that is currently in place for contract teachers (deferred acceptance, DA) and allow schools to increase the wages they offer sequentially until they fill their vacancies, either unconditionally or conditionally on the quality of the assigned teacher. We characterize the school-optimal stable allocation under a generalized DA algorithm and show that a policy that sets compensation at each job posting using the information generated by the matching platform is more efficient in terms of reducing structural inequality in access to learning opportunities. In comparison, a rigid system that ignores teacher preferences will indirectly reinforce such inequalities.

While flexibly incorporating information on teacher preferences, the counterfactual policy achieves the same objectives of the current system of wage bonuses at a much lower cost for the government. We also find that filling every school with at least one teacher would require a lower budget (in terms of the total wage bill) than the current policy. However, shifting the supply of highly qualified teachers towards hard-to-staff schools is significantly more costly. Given the existing stock of prospective teachers and school vacancies, it would take almost seven times the current budget to assign a teacher in every school with the median competency level of urban areas in the status quo. This result can be explained by the fact that such policy objective would require many unassigned and high-quality applicants to accept a teaching position within the system. Investing in local infrastructures in our setting would entail achieving our policy objectives at total costs that are 20-30 percent lower. Place-based incentives aimed at enhancing the pool of teachers in locations where the supply is relatively scarce would entail saving 40 percent of the total cost of the policy that assigns a teacher in every school. This last result highlights the predominant roles of moving costs and of the ethnolinguistic match effects in explaining teacher preferences over job postings in our setting.[2]

Given that the returns to having a high quality teacher are more pronounced at the lower end of the distribution of learning outcomes, our findings suggest that there is large scope for efficiency gains with respect to the current policy by reallocating more qualified personnel towards remote rural locations. Putting together the regression-discontinuity estimates and the policy counterfactual simulations, we estimate that the share of under-performing students in the most disadvantaged locations would decrease from 80 percent to at least 50 percent, at the same total cost for the government.

This paper contributes to a growing literature that uses equilibrium models to study the implications of compensation policies on the spatial distribution of teachers (Boyd et al., 2013; Tincani, 2021a; Biasi et al., 2021a). Our results expand the literature evaluating teacher wage setting policies in developed and developing countries. There is large body of work studying the effectiveness of pay-for-performance schemes.[3], while relatively fewer studies consider policy effects of unconditional wage increases on teacher turnover (Clotfelter et al., 2008) and student outcomes (de Ree et al., 2018; Pugatch and Schroeder, 2018; Cabrera and Webbink, 2020). Our work contributes to this literature by providing well identified reduced form estimates of the effects of unconditional wage increases on the distribution of quality teachers across regions and their effects on educational outcomes. Further, our structural estimates allow us to go beyond the policy evaluation and provide counterfactual estimates

[2] Ajzenman et al. (2021) show evidence that teacher applications to hard-to-staff schools can also be influenced by information interventions or behavioral nudges.

[3] See for example Muralidharan and Sundararaman (2011); Barrera-Osorio and Raju (2017); Gilligan et al. (2022); Leaver et al. (2021a); Brown and Andrabi (2020).

on the effects of alternative compensations policies.

More broadly, we contribute to the recent literature studying different personnel and organizational policies in the public sector (Finan et al., 2017). For instance, Dal Bo et al. (2013) show that increased compensation for public sector positions in Mexico led to a larger pool of applicants, and a higher quality of hired employees. In Uganda, Deserranno (2019a) finds that higher financial incentives attract more applicants and increase the probability of filling vacancies while crowding out pro-socially motivated health workers. We complement this literature by incorporating a empirical market design approach (Agarwal and Budish, 2021), which leverages matching platforms to study the design of compensation schemes for public school teachers. This approach, which was pioneered by Agarwal (2015,0), provide important insights into the potential for improving the equity and efficiency of public service provision through its effects in the reallocation of high-quality public employees.

2.2 Data

In this paper, we combine several administrative data sources from the Ministry of Education of Perú over the period 2015-2018. While the resulting dataset spans the universe of public-sector teachers and schools, we restrict our analysis to primary schools for two reasons. First, secondary schools are much less prevalent in rural areas. Public schools serve 74% of the primary school enrollment countrywide. In rural communities (i.e., those with less than 2,000 inhabitants), public schools are generally the only option.[4] To the extent that the geographic distribution of schools is key to understanding disparities in access to competent teachers, we need to focus on primary schools that are well represented throughout the country. Second, in primary schools, all students in a classroom are taught by a single teacher, instead of having one teacher for each subject. This setup allows us to more precisely match students and their teachers, and estimate the effect of the newly assigned teachers on student achievement in the empirical analysis in Section 2.4.

Our first data source is the *centralized teacher job application and assignment system*. This dataset includes information on all job vacancies posted at every public school in the country during the first two rounds of the national recruitment of public sector teachers (2015 and 2017), the scores in the standardized evaluations for every applicant, and detailed

[4]There are more than 6,000 public primary schools in rural areas catering to 98 percent of school-aged children in 2015.

information on all the steps of the job application process that we discuss in Section 2.3.2. Figure 2.A.1 shows some relevant individual-level correlates of teacher performance in the standardized test. During the first (second) national recruitment drive, 64,000 (72,000) applicants competed for 18,000 (25,500) vacancies in primary schools. Table 2.A.1 reports basic descriptive statistics on applicants across types of contracts in the public sector. About 8% of the applicants report no prior teaching experience (neither in the public sector nor in the private sector). More than one-fourth of the applicants in our sample report speaking *Quechua* or *Aymara* as their main language, thus likely belonging to the ethnic groups that are concentrated in the Andean highlands, while an additional 2% belong speak one of the many other languages spread in the Amazon forests.[5]

Our second administrative data source is the *teacher occupation and payroll system (NEXUS)*. This is a longitudinal dataset collected and maintained by the Ministry of Education, which contains the complete records of all teachers employed in the public sector. In particular, the dataset includes individual identifiers for all teachers, the school in which they work (but not the specific grade), and the type of contract/position they hold (permanent or contract, number of hours, etc.). This information is collected at the start, middle, and end of each school year, allowing us to precisely trace both the school of origin (if any) and the school of destination (if any) for each applicant to the national recruitment drives. About two-third of the applicants had previously been employed as public sector teachers.

We obtain data on school and locality characteristics from the national *school census*. These data include information on the number of students, school infrastructure (libraries, computers, classrooms, sports facilities), and staff (teaching and administrative) at each school. Additionally, the dataset includes information on local amenities, e.g., access to basic services (electricity, sewage, water source) and infrastructure (community phone, internet, bank, police, public library). This information is reported annually by school principals. Table 2.A.2 reports basic descriptive statistics for some key school characteristics for urban and rural areas, respectively.

Our fourth data source is the administrative records on *student academic outcomes*. The *Evaluación Censal de Estudiantes* (ECE) is a national standardized test that covers curricular knowledge of math and language (Spanish). The test is administered by the Ministry of

[5]The information on the main language spoken by the applicants is only available for the 2015 recruitment drive.

Education at the end of every school year at selected grades at both public and private schools with an enrollment of more than five pupils. We have access to individual test scores from 2014–16 and 2018 for fourth grade students in public primary schools (widespread floods in the country led the government to cancel the 2017 exams).

Finally, in collaboration with the Ministry of Education, we administered an online survey among the applicants to the permanent teaching positions during the 2015 centralized job application process. The response rate is slightly below 20% (5,553 applicants), and observable teacher characteristics of respondents are not different from those that did not respond. Among several questions on teachers' application decisions, we asked applicants to rank the their preferred school's characteristics. As shown in Panel B of Table 2.A.3, 44% of teachers say "being close to home" is one of the key characteristics guiding their preference ranking. Other often cited attributes of the teaching job are prestige, safety and "cultural reasons". While "prestige" is admittedly a somewhat vague concept, "cultural reasons" mainly refers to ethnolinguistic similarities between teachers and the communities where the schools are located. Interestingly, distance and prestige are disproportionately appreciated by teachers who scored the highest grades (top quartile) in the centralized test, compared to the average teacher. These survey results partly motivate the empirical model that we propose and estimate in Section 2.5.

2.3 Context and Institutions

2.3.1 Inequality of Education Inputs

Perú is a country that spans a vast and varied geography, which includes mountainous areas in the Andes, the Amazon forest, and coastal regions. It is composed of culturally and linguistically diverse people, who have lived under extractive systems of governance as a Spanish colony. The legacy of colonial institutions and policies is one of the root causes of current structural inequalities. These previous policies were often targeted to the highlands and jungle regions, where most of the natural resources are located. Currently, those areas show high poverty rates and a large concentration of indigenous people.

Over the last decade, the government has undertaken several efforts to improve educational outcomes in poor, rural areas, such as implementing a large-scale conditional cash transfer program, investing in school infrastructure projects, and improving access to drinking water and sewage (Bertoni et al., 2020). However, large differences still exist in the access to educational inputs such as school infrastructure. In Table 2.A.2 we document some differences between schools in urban and rural areas across a broad set of indicators of schools, teachers, students, and community characteristics. Schools in rural areas predominantly hold (90%) mixed classes, with a single teacher serving students of several grades at the same time. About one-third of the rural schools lack access to basic services such as running water or electricity.

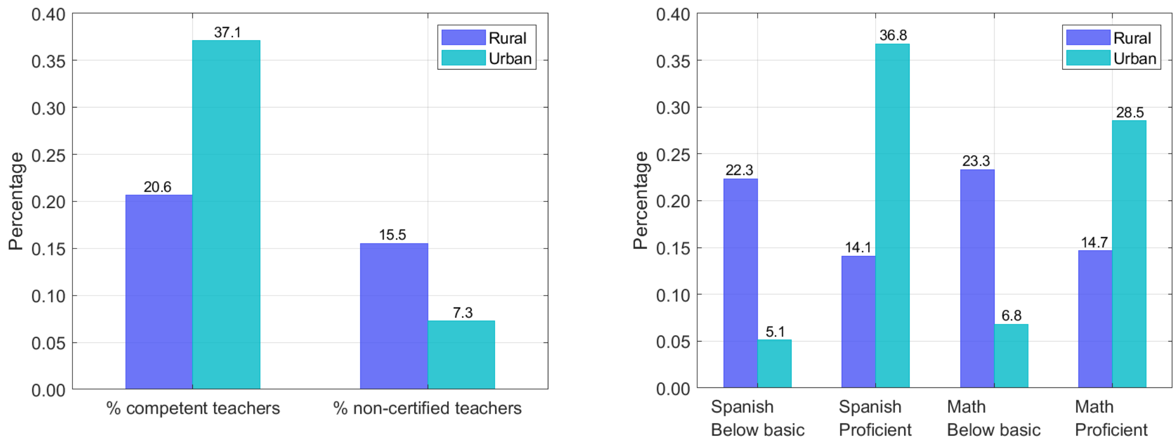
Figure 2.1 documents the stark differences in teacher quality and student achievement between urban and rural primary schools. Panel A shows that teachers at rural public schools are half as likely to pass the requirements set by the government for permanent teachers ("competent teachers"), and are twice as likely to lack teaching credentials (non-certified teachers).[6] Panel B of Figure 2.1 displays students' academic performance on the national standardized evaluation in two subjects – Spanish and math. Approximately one in four students enrolled in rural schools are classified as performing below the basic curricular requirements in either of the two subjects, whereas the corresponding shares in urban schools are only around 5%.

Figure 2.A.2 shows the geographic distribution of competent teachers across provinces alongside the corresponding distribution of student test scores. Competent teachers are heavily concentrated in the richer, coastal cities, while they are nearly absent in the highlands and the inner amazonian regions (Panel A). The spatial variation in students' achievement outcomes, shown in Panel B of Figure 2.A.2, is almost a mirror image of the spatial distribution of competent teachers across the country.

We document inequality in schooling inputs and outputs across Perú. While local amenities and school infrastructures likely reflect structural differences between urban and rural areas, the unequal spatial distribution of teacher quality suggests a margin where policy can play an important role. To better understand the reasons behind the current allocation of

[6] Competent teachers are defined in Figure 2.1 as those who attain a score of at least 60% in the curricular and pedagogical knowledge module of the standardized test used to both screen and recruit teachers (see Section 2.3.2). Subject competency test have shown to correlate with teacher value added and other dimensions of teacher quality in several contexts (Bold et al., 2017; Estrada, 2019; Gallegos et al., 2019; Araujo et al., 2020). For the Peruvian case, Bertoni et al. (2021) document strong correlations between various measures of teaching effectiveness and the score in the curricular and pedagogical knowledge module of the evaluation test.

Figure 2.1: Teachers and Students in Urban Vs. Rural Areas



a) Teachers **b) Students**

NOTES: These figures show different summary statistics about teachers and students in urban and rural areas. Panel A shows, separately for rural and urban schools, the average share of teachers classified as competent based on the curricular and pedagogical knowledge of their subjects of specialization and the average share of teachers who lack teaching certifications. Panel B shows how academic performance in the Spanish and math modules of the national standardized evaluation differs between students of urban and rural schools. Table 2.A.2 in the Appendix presents a broader set of indicators for school and community-level characteristics across urban and rural areas.

teachers across different geographic areas, we now describe the institutions that govern the labor market for public school teachers.

2.3.2 Contracts, Wages, and Sorting of Public School Teachers

Public school teachers in Perú are hired under two distinct types of contracts. Permanent teachers (*docentes nombrados*) are civil servants with stable employment conditions (i.e. indefinite contracts). Alternatively, teachers can be hired by the central administration to work at a specific school for an academic year as contract teachers (*docentes contratados*). This contract has the option of being renewed for up to one more year, conditional on being approved by the school's administration. Short-term contracts are routinely used in most education systems around the world and are often designed as entry-level positions in the teaching career.[7] In our setting, about one out of five primary school instructors in urban areas is hired as a contract teacher, while these contracts are more widespread in rural areas, where they reach almost half of the labor force in the most remote schools.

[7]Research in India and Kenya shows that locally hired teachers on annual contracts have better performance, and their students score higher in standardized test scores (Muralidharan and Sundararaman, 2011; Duflo et al., 2015), although in Kenya these gains tend to vanish when the contracts are administered by the government, rather than by a non-government organization (Bold et al., 2018).

The compensation of public-school teachers in Perú depends on (i) the type of contract (permanent or contract teacher), (ii) seniority, and (iii) specific location or school characteristics. In 2016, the base monthly wage for primary-school teachers under a short-term contract was S/ 1,396 (US$ 402), while that for permanent teachers was S/ 1,550 (US$ 447), although more experienced permanent teachers can earn up to S/ 4,043 (US$ 1330). Additional wage bonuses are given to all teachers (irrespective of the contract) working in specific types of schools, such as multi-grade or single-teacher schools, or schools located in disadvantaged communities.[8] According to the national household survey (ENAHO 2016), the earnings of primary school teachers are ranked second to last among the liberal professions in Perú, followed only by translators and interpreters. Nationally representative survey data on teachers (ENDO 2014) document that the average monthly wage for teachers working in primary schools in the private sector is approximately S/ 950. Only private school teachers in the top ten percent of the distribution earn more than the base wage of a teacher in the public sector.

Permanent and contract teachers in Perú were recruited in a decentralized fashion until 2015. As in most countries, regional and local level officials often had significant discretion in teacher hiring and allocation decisions (Bertoni et al., 2019; Estrada, 2019). In an effort to make the process more transparent and meritocratic, the Ministry of Education established a nation-wide recruitment process in which school-level job postings and teacher job applications are processed on a single, centrally-managed platform. The first national recruitment drive took place in 2015, followed by another round in 2017. Teachers recruited through the 2015 and 2017 drives started teaching in the 2016 and 2018 academic years (March-December), respectively.

The recruitment process is structured in two phases.

Permanent teacher recruitment. Every vacancy for permanent teachers across all education levels are posted in a centralized platform. The opening of each of these positions depend on previous retirements and transfers and the ability of local governments to secure

[8]Figure 2.A.3 shows the different wage bonuses, which vary between 4% (bilingual school) and 36% (extremely rural locations, as defined in Section 2.3.3) of the monthly base wage. Schools can satisfy multiple criteria (e.g. multi-grade and bilingual), in which case the bonuses are cumulative. Accredited bilingual teachers are eligible for an additional bonus of S/ 100. There are also some compensation adjustments throughout the year, such as a holiday bonus, which usually represents less than 5% of the total monthly wage. The wage bonuses for multi-grade and single-teacher schools were cancelled in 2017 and reinstated in 2020.

permanent funding for the position. Applicants are required to have a teaching accreditation (i.e. a teacher degree) and to have taken the standardized competency evaluation.[9] Those who correctly answer at least 60 percent of the questions in each of the three parts of the test are eligible for a permanent position, and can in turn submit a ranked-order list of school preferences of up to five available positions within a given school district.[10] In our data, about 10% of the applicants are eligible for a permanent teaching position. Once preferences are submitted, teachers move on to a decentralized stage of evaluation in which each school interviews a short-list of the highest scoring teachers who express a preference for that vacancy. In this second evaluation, teachers are given another score based on their performance in a typical class that they have to teach and an in person interview with the principal and other school stakeholders. Additionally points can be also assigned based on their CV. Permanent positions are finally allocated based on an overall score that comprise the competency test and the decentralized evaluation.

Short-term/contract teacher recruitment. The goal of this stage is to fill as many of the remaining positions with a certified teacher. About half of the applicants who cleared the bar to be eligible for permanent positions eventually participated in this round of the assignment mechanism. These are mostly teachers who were not selected for their preferred positions and thus opted for a temporary position. Importantly, these teachers are not systematically different to those who are assigned to a permanent position (see Table 2.A.1). Unlike the assignment of permanent teachers, short-term teaching positions are allocated through a serial dictatorship algorithm. In this mechanism, school preferences are taken to be a strict ranking of teachers' competency scores. Applicants sequentially (starting by the highest ranked) choose from the list of open vacancies in a given school district. Once a vacancy is filled, it is eliminated from the list of the available options in that district, and the next lower-ranked teacher is allowed to pick her preferred option. This iterative process continues until all vacancies are filled, or until the lowest-ranked teacher in each school district is allowed to choose among the remaining vacancies. After the first round

[9]The test is divided into three modules, which carry different weights in the total score: logical reasoning (25 percent), reading comprehension (25 percent), and curricular and pedagogical knowledge (50 percent).

[10]In the 2015 screening and recruitment process, candidates were allowed to rank a maximum of five schools. This constraint was removed from 2017 onwards, and applicants were free to submit an unlimited number of options. In total there are 218 school districts. There is substantial within-district variation in the rural status of the school vacancies. For the average school district, 71 percent of vacancies are in rural locations. In 33 school districts all available vacancies are in urban locations (15 percent).

of the matching process, unassigned applicants are given another chance to choose among the remaining open vacancies from other districts. Positions that are not filled through the serial dictatorship mechanism are eventually filled through a decentralized secondary market, where non-certified teachers are also included.[11]

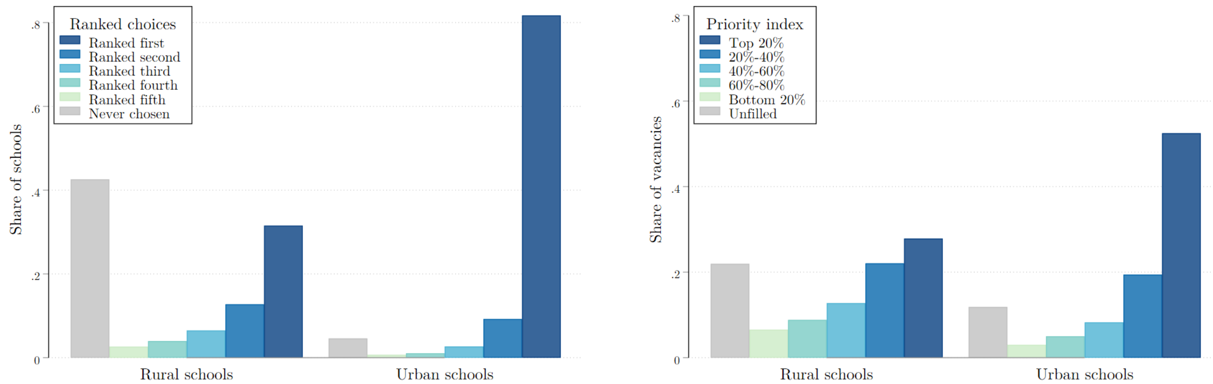
Figure 2.2 shows data from the applications to primary-school vacancies, showing teachers' preferences for different types of schools. While 80% of schools in urban areas are ranked first by at least one applicant to a permanent position, vacancies posted in rural areas receive significantly fewer applications – nearly half of rural schools are never even ranked in applicants' preference lists. As a result, more than two-thirds of job vacancies for permanent positions remain unfilled in rural schools, while three-fourths of vacancies are filled in urban schools through the centralized assignment mechanism. Panel B considers the sample of contract teacher positions by plotting the quintiles of the priority indices for the positions that are filled in the serial dictatorship algorithm. Short-term teaching vacancies in urban areas are in higher demand, as more than half of these postings get filled by teachers ranked in the top 20% of the pool of applicants in their respective school districts whereas the distribution of the assigned teachers at short-term vacancies in rural areas is clearly more skewed toward less qualified personnel (as measured by the competency score). Overall, the centralized assignment process fills almost 90 percent of short-term vacancies in urban areas and slightly less than 80 percent of short-term vacancies in rural areas.[12]

We conclude that the spatial inequality in access to qualified teachers displayed in Panel A of Figure 2.1 can be (at least in part) explained by teachers' preferences and choices over locations. Teachers in poor rural areas face numerous challenges: scarcity of basic school inputs, lack of services and public goods, few local amenities, and (for some) being far from friends and family. To the extent that wage-setting protocols do not compensate for the lack of these amenities, these jobs will be less attractive. Indeed, the data on job postings and teacher rank-order applications show that applications are skewed toward positions in urban areas, and the system is hard-pressed to staff the roughly 14,000 positions in rural public schools in the poorest parts of the country. As a consequence, many of these vacancies are

[11] Over 53,000 applicants for short-term teaching positions (88%) were not assigned within the first two (centralized) rounds of the 2015 assignment mechanism. More than three-quarters of them re-applied in the 2017 assignment mechanism.

[12] While on average there are seven applicants per vacancy within the centralized application platform, there are more than two vacancies per applicant in indigenous communities in the forest inlands, which explain the reason why these vacancies are more likely to remain unfilled (50% vs. 21% in the overall sample).

Figure 2.2: Teacher Choices over Job Postings



a) Permanent Teachers **b) Contract Teachers**

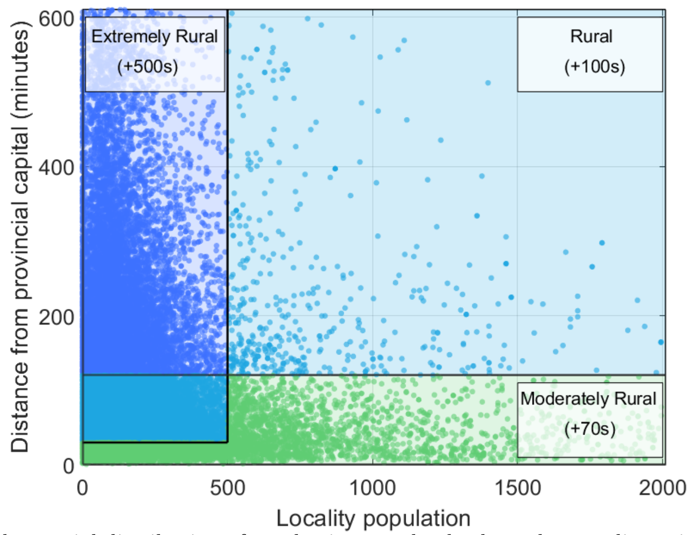
NOTES: This figure depicts the demand for teaching positions in rural and urban schools. Panel A plots the relative share of schools by the highest preference received, so that "ranked first" means that at least one teacher from among all applicants ranked it as number one, "ranked second" means that no teachers ranked the school as number one, but at least one teacher ranked it as number two, and so forth. Similarly, the grey bar indicates the relative share of schools that were not mentioned in any of the permanent teacher rankings. Panel B plots the priority order (grouped in quintiles of the teacher competency score) in which a short-term position is filled, together with the share of vacancies that remained unfilled (not filled by a certified teacher). The numbers are obtained by pooling the data from the two recruitment drives from 2015 and 2017.

eventually filled using short-term contracts by teachers who, on average, have competency scores that are 0.5 standard deviations lower than those assigned to urban schools, while the remaining portion are filled by non-certified teachers through the decentralized secondary market.

2.3.3 Policy Changes to Compensation in Rural Locations

The government recently implemented a reform to the wage bonus that significantly increased teacher compensation at positions in select rural schools. The new policy established three distinct categories of rurality according to the school locality's population and its proximity to the provincial capital (see Figure 2.3). The population of the locality is measured by population counts in the latest available census (2007). Travel time from the locality to the provincial capital is used as a proxy for how remote a community is, and it is computed based on the school's GPS coordinates, the types roads available at the time of the measurement, and the most frequent modes of transport. *Extremely Rural* schools are those in localities with less than 500 inhabitants, and for which it takes more than 120 minutes to reach the province capital. The second category of schools, labeled as *Rural*, is reserved for either: (a) schools in localities with less than 500 inhabitants and are located between 30 and 120

Figure 2.3: The Distribution of Rural Schools and the Wage Bonuses



NOTES. This figure shows the spatial distribution of rural primary schools along the two dimensions that determine assignment of the rural wage bonus. *Extremely Rural* schools are the dark blue dots, *Rural* are light blue and *Moderately Rural* schools are green.

minutes from the province capital, or (b) schools in localities with 500–2,000 inhabitants that are farther than 120 minutes from the province capital. The third category of *Moderately Rural* schools are either: (c) in localities with 500–2,000 people that are within 120 minutes of the province capital, or (d) in localities with less than 500 inhabitants which are within 30 minutes of the province capital. All other schools are classified as Urban, and are therefore not entitled to the wage bonus.

Rurality bonuses were first introduced in January 2014, and only permanent teachers were eligible to receive them. In August 2015, the wage bonuses were extended to contract teachers. Importantly, these changes were only announced briefly before they were actually implemented (in August, i.e. in the middle of the school year) and thus right before the first centralized recruitment drive (October 2015), which marks the start of our study period. The bonus for *Extremely Rural* schools is fairly generous: for contract teachers, it ranges between 25 and 36 percent of the base wage, depending on the school year considered (contract teacher wages increased from S/ 1,396 in 2016 to S/ 2,000 in 2017); for permanent teachers, it ranges between 25 and 32 percent of the base wage.

Figure 2.3 displays a scatter plot of the distribution of the 25,000 rural primary schools in Perú over the population (x-axis) and the proximity to the provincial capital of the communities where the schools are located (y-axis). There is a large mass of schools around both the time cutoffs (30 minutes and 120 minutes from the provincial capital) and the population cutoff (500 inhabitants) for the rural wage bonuses. As the localities become more remote, schools are more likely to be located in communities that are small and predominantly fall into the *Extremely Rural* category. Likewise, for localities with populations above 1,000 inhabitants, there are more communities that are closer to the provincial capitals (*Moderately Rural*).

2.4 Causal Effects of the Increase in Compensation

2.4.1 Regression Discontinuity Design

Offering higher wages for positions at rural locations could potentially lead to better student outcomes through two main mechanisms. On the one hand, at the extensive margin, higher wages could attract more and higher-quality teachers. On the other hand, higher levels of compensation may also motivate incumbent and newly hired teachers to exert higher levels of effort. Our empirical analysis identifies the causal effects of unconditional wage increases on teacher application behavior, teacher selection, and student outcomes. We do this by exploiting the classification rules of the rural wage bonus, and compare (i) the characteristics of teachers who choose/are assigned to a position at a high vs low paying school, and (ii) student test scores between schools that offer high vs low compensation to their teachers. Additionally, to discern whether changes at the extensive or at the intensive margin of teacher quality can explain the effect of the wage reform on students' academic achievement we can compare student outcomes in schools with and without open vacancies in the national recruitment drives.[13]

The introduction of the rural wage bonus may generate incentives for school principals and administrators to manipulate the information used to determine bonus eligibility. The population threshold is based on census data, and as such, it is difficult to manipulate, whereas the time-to-travel measure is gathered by inspectors from the Ministry of Education, who physically go to the schools and take the GPS coordinates of the school's location. The procedure was originally done in 2014 and then repeated in 2017 to account for possible

[13]Table 2.A.7 shows that there is no effect of the wage bonus on the probability that a school has an open position for permanent or contract teachers. Figure 2.A.4 displays scatter plots similar to the one reported in Figure 2.3 for schools with and without vacancies in the national recruitment drives of 2015 and 2017, respectively.

changes in the transportation network. By the time the information was to be updated, the previous measurement had become public information, and hence some schools located just below the 120-minute threshold may have gained eligibility to the S/ 500 wage bonus by slightly manipulating the GPS measurement. The data shows that there is a significantly larger mass of schools that falls just above the time-to-travel threshold for the assignment process that took place in 2017, while there are no significant jumps in the density of schools at the population threshold for either of the years of interest (see Figures 2.A.6, 2.A.7, and 2.A.8).[14]

We thus rely on the population-based assignment rule as the only source of exogenous variation in teacher wages for this part of the analysis. Table 2.A.5 shows the estimated wage increases explained by crossing the population threshold. Contract teachers in localities with slightly less than 500 inhabitants earn on average over S/ 250 more than those in localities that are just above the cutoff. This represents an increase in the monthly wage of about 13 percent. The corresponding average increase in wages for newly recruited permanent teachers (i.e., no experience) due to the rural bonus reform is S/ 225, or 11 percent of their monthly wage.[15]

Given continuity of potential outcomes around the population cutoff, the following specification identifies the effect of a higher wage bonus:[16]

$$y_{jt} = \gamma_0 + \gamma_1 \mathbf{1}(pop_{jt} < pop_c) + g(pop_{jt}) + \delta_t + u_{jt}, \tag{2.1}$$

where y_{jt} is an outcome variable for school j at time t, $g(\cdot)$ is a flexible polynomial in the population of the locality of the school at both sides of the population cutoff, δ_t denotes time indicators for the specific year of the recruitment drive (included only for teachers' outcomes),

[14]While the locality population is a good predictor for the eligibility to the rural wage bonus in both years, time-to-travel in 2015—which we observe to be less prone to manipulation—does not help predict the policy eligibility status in 2017 and therefore doesn't provide useful variation for estimating the effects of the wage bonus in 2017 (see Figure 2.A.9). An alternative strategy would be to limit the sample to observations that are above the 120-minute time-to-travel cutoff, however, this implies conditioning on a partially manipulated variable. This sample restriction would also exclude a large portion of schools, and in particular, some located in the lower-right quadrant of Figure 2.3, thereby missing relevant variation in wages in the data.

[15]These effects are unconditional weighted averages –pooled across school years– of the different wage increases induced by crossing the population cutoff from above for different values of the time-to-travel variable.

[16]Table 2.A.6 shows that pre-determined school and locality-level covariates are smooth around the population threshold, with point estimates that are very small and not statistically different from zero in all but five cases for 2015, and in all cases for 2017 (29 covariates considered).

and u_{jt} is an error term clustered at the school×year level for teachers' outcomes and clustered at the school level for students' outcomes (that we observe in only one year, see Section 2.4.4). The parameter of interest is γ_1, which represents the average outcome difference between schools, teachers, or students in localities that are just above or below the population cutoff, and therefore that are marginally eligible to receive (or not) an unconditional increase of about 15% in teacher wages. We estimate γ_1 non-parametrically using the robust estimator proposed by Calonico et al. (2014) through bias-corrected local linear regressions that are defined within the mean squared error optimal bandwidths.

We exclude from the estimation sample all urban and rural schools in localities within 30 minutes of the province capital since for them, crossing the population cutoff does not lead to an increase in the bonus. We further restrict the sample to schools with non-missing observations for the different outcome categories considered in our analysis. We present all the results pooling the data from the two recruitment drives from 2015 and 2017. The results split by year are shown in Tables 2.A.8 and 2.A.9 and are broadly consistent with the patterns described in the main text.[17]

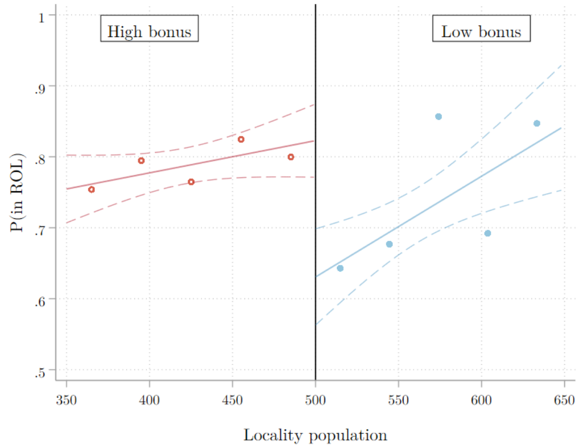
2.4.2 Teacher Choices over Job Postings

We start by showing how teachers' application behavior is affected by higher wages, providing direct evidence on the effects of wage increases on teachers' labor supply decisions. We document graphical evidence of the threshold crossing effects separately by job applications for permanent and short-term teaching positions. Panel A of Figure 2.4 documents clear evidence that applicants for permanent teaching positions are more likely to include in their applications schools in localities with a population just below the cutoff (eligible for a higher wage bonus), as opposed to options just above (not eligible for a higher wage bonus). Away from the cutoff, the observed positive correlation between teachers' choices over job postings and the population of the community is consistent with the notion that the population captures some valuable amenities in the locality.

Panel B considers the choices over job posting for contract teachers. As in Section 2.3.2, we infer teachers' preferences over positions from choices observed in the serial dictatorship. To do this, we normalize the ranking in which a position is chosen within a school district,

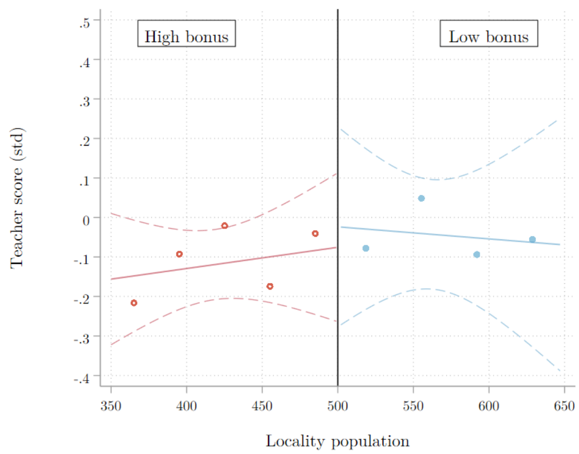
[17] The main estimates reported in this section are robust to alternative specifications and estimation choices. The results of these specification checks are reported in Figures 2.A.10 and 2.A.11.

Figure 2.4: Teacher Choices over Job Postings

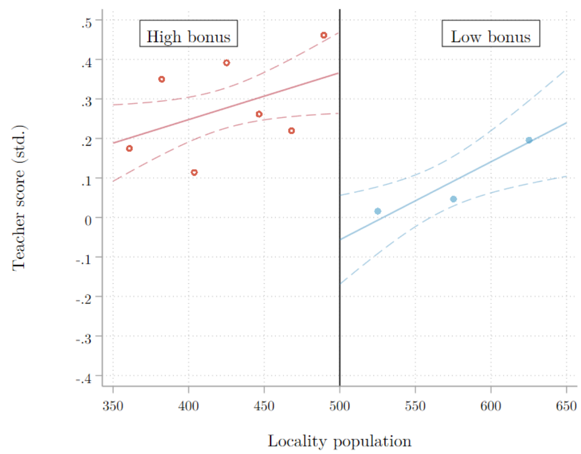


a) Stated Preferences, Permanent Teachers

b) Revealed Preferences, Contract Teachers



c) Competency Score, Permanent Teachers



d) Competency Score, Contract Teachers

NOTES. This figure shows how applicants' preferences and quality vary based on the difference between the 500-inhabitants cutoff and the population of the community where the school is located. Panels A and C focus on the assignment process of permanent teachers. In Panel A the outcome variable is a dummy equal to one if a school was mentioned in at least one application, while in Panel C the outcome variable is the standardized (total) score obtained in the centralized test by the newly-assigned permanent teacher. Panels B and D are analogous to A and C for the assignment process of contract teachers. Panel B uses as outcome variable the priority in which a vacancy was chosen in the serial dictatorship mechanism (normalized so that it takes value from zero to one), while Panel D uses the standardized score obtained in the centralized test by the newly-assigned contract teacher. Each marker indicates the average of the outcome variable within each bin, defined following the IMSE-optimal evenly spaced method by Calonico et al. (2015). Solid lines represent the predictions from linear regressions estimated separately for observations to the left and to the right of the cutoff.

so that the index takes the value of zero if the position is filled last and one if the position is filled first. Short-term positions that are just below the population cutoff get filled at higher priority order when compared to those above the cutoff, which again indicates that the wage bonus increases the demand for these positions.

Table 2.1 reports the corresponding regression-discontinuity (RD) estimates from the empirical specification in equation (2.1) using data at the school/vacancy level. In Column (1) the dependent variable is either an indicator that takes the value of 1 if a school was mentioned in at least one application for a permanent teaching position (Panel A) or the normalized priority index at which a short-term position is filled (Panel B). In the neighborhood of the population discontinuity defined by the MSE-optimal bandwidth (RD sample), the average school is mentioned in 76% of permanent teacher rankings. This proportion increases by 19 percentage points for schools that offer higher wages. Similarly, the average short-term position in localities with a population slightly above 500 inhabitants is filled by a teacher ranked in the 37^{th} percentile $(1 - 0.63)$ of the score distribution of applicants, while schools that offer a wage bonus manage to fill the position with an applicant in the 24^{th} percentile $(1 - 0.63 - 0.13)$.

The priority index of contract teachers reported in Column (1) and the competency scores for both permanent and contract teachers reported in Column (3) of Table 2.1 are defined for the subset of the open vacancies that got filled in the centralized stages of the matching process. To deal with this potential endogenous selection into the sample, we report RD bounds below the point estimates using the approach outlined in Gerard et al. (2020). The bounds are in general quite tight, thereby suggesting that the censorship in the density of the observations due to the fact that some vacancies remain unfilled is inconsequential for the RD estimates.

The evidence presented in this subsection show that vacancies at schools that receive a higher wage bonus become more desirable: They are requested more often by applicants for permanent positions and are filled faster by contract teachers. The increased competition for vacant positions can lead to an increase in the quality of applicants who select into these higher-paying jobs and/or an increase in the quantity of teachers matched to those rural vacancies. In turn, we explore these potential margins of response to the wage reform in the next subsection.

2.4.3 Teacher Sorting Patterns

A first-order objective of the centralized assignment system is to fill as many position as posible. If the vacancies go unfilled, schools either recruit teachers without credentials or increase the workload for the existing teachers at the school, presumably reducing their

Table 2.1: Teacher Choices and Sorting

Panel A: Sample of Permanent Teachers			
	(1)	(2)	(3)
	Stated Preferences	Vacancy filled	Competency score
High Bonus	0.188	0.026	-0.037
	(0.069)	(0.074)	(0.155)
Bounds			[-.302; .205]
Mean dep. var. (Low Bonus)	0.755	0.371	-0.080
Bandwidth	166.986	169.330	259.213
Schools	835	847	830
Observations	1009	1725	1167
Panel B: Sample of Contract Teachers			
	(1)	(2)	(3)
	Revealed Preferences	Vacancy filled	Competency score
High Bonus	0.130	0.051	0.483
	(0.036)	(0.048)	(0.124)
Bounds	[.116; .138]		[.391; .5]
Mean dep. var. (Low Bonus)	0.634	0.900	0.063
Bandwidth	150.781	159.432	152.768
Schools	836	935	851
Observations	1917	2199	1955

Notes. This table reports the effect of crossing the population threshold on different outcomes. Panel A uses the sample of permanent teachers. In Column (1) the outcome variable is a dummy equal to one if a school was mentioned in at least one application, while in Column (2) is an indicator for whether the vacancy was filled by a certified teacher in the assignment process for permanent teachers. The regression displayed in the last column uses the standardized competency score obtained by the teachers in the centralized test as outcome variable. Panel B focuses on the selection process of contract teachers. Column (1) shows the effects on the rank in which a vacancy was chosen in the deferred acceptance mechanism (normalized so that it takes values from zero to one), while Columns (2) and (3) are analogous to those from Panel A. Cells report the bias-corrected regression-discontinuity estimates obtained using the robust estimator proposed in Calonico et al. (2014). Regressions are defined within a mean-square error optimal bandwidth (BW), reported at the bottom part of the table. In Column (1) of Panel A and in Column (3) the sample is restricted to vacancies that were actually filled by a certified teacher. In those cases, the table also reports the RD bounds estimated using the procedure developed in Gerard et al. (2020). The table also reports the mean of the dependent variable computed within the interval $(-BW, 0]$ (Low Bonus). Standard errors are clustered at the school×year level.

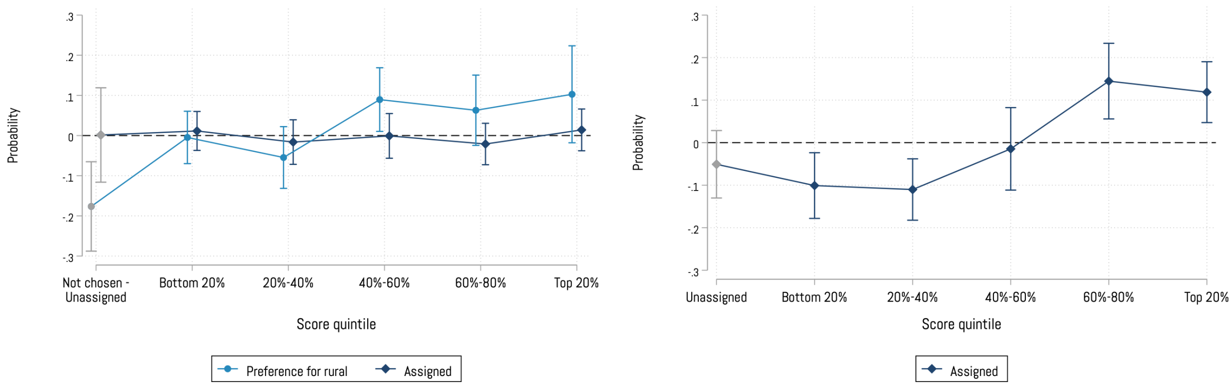
effectiveness. Column (2) of Panel A in Table 2.1 presents regression-discontinuity estimates for the probability that a vacancy is filled in the selection process of permanent teachers, while Panel B shows analogous estimates for contract teachers. Permanent teacher positions that offer higher wages are not more likely to be filled, compared to those offering lower wages. For contract teachers, instead, we find a positive but not statistically significant effect of higher wages on the probability that a vacancy is filled. This evidence can be reconciled with the "local" nature of the estimates shown here. While it may be the case that higher wages induce some teachers to accept a position in a more disadvantaged location, this margin of response to the wage bonus may be active elsewhere in the spatial distributions of schools shown in Figure 2.3. Indeed, 90% of the rural vacancies are filled in the low bonus areas of the RD sample. In Section 2.6.1, we address this issue directly by simulating the global sorting patterns triggered by the system of wage bonuses currently in place using the estimated model of teachers' preferences that we discuss in Section 2.5.

We next investigate whether the observed boost in competition for high-paying positions in extremely rural locations leads to an increase in teacher quality, as measured by the competency score used to define priorities in the assignment algorithm. The two-sided nature of the assignment process for permanent teachers may possibly explain the small and insignificant effects of a higher wage bonus on the quantity and quality of realized matches in *Extremely Rural* schools, as reported in Figure 2.4 (Panel C) and in Column (3) of Table 2.1 (Panel A). Figure 2.5 shows estimates reflecting the preferences and final assignments of permanent teachers (Panel A) and contract teachers (Panel B) for the different quintiles of the test score distribution. Schools offering higher bonuses are more likely to be included in the ranked-order lists of more competent teachers (light blue line in Panel A). However, this change in demand triggered by the wage incentives does not translate into a disproportional assignment of higher quality teachers in these schools (dark blue line in Panel A). The decentralized stage of the assignment mechanism may have potentially undone the positive sorting toward disadvantaged locations induced by higher wages.

Both the graphical evidence displayed in Figure 2.4 (Panel D) and the RD estimates in Column (3) of Panel B of Table 2.1 show that contract teachers who select into schools that offer a higher wage bonus have higher competency scores, on average, than those who choose a position in another rural school. The magnitude of the effect is 0.48 standard deviations of the distribution of the competency score, a very large effect, which points towards quantitatively important sorting implications within the assignment system. The magnitude of the effect is consistent with the fact that a larger proportion of teachers in the top two quintiles of the test score distribution disproportionally sort into higher paying positions (see Panel B of Figure 2.5). To put this magnitude in perspective, the average gap in teachers' competency between *Extremely Rural* schools and other rural schools is approximately 0.3 standard deviations, whereas the average gap between rural and urban schools is about 0.5 standard deviations.

In sum, a higher wage bonus targeted at disadvantaged locations shifted applications toward schools offering both permanent and short-term positions. This change in teachers'

Figure 2.5: Wage Bonuses and the Selection of Competent Teachers



a. Permanent Teachers **b. Contract Teachers**

Notes. The figure displays the effect of crossing the population threshold on different measures of the demand for teaching positions and the resulting quality of the recruited teachers. Circles in panel A indicate the point estimates from a set of regression of the form of Equation (2.1) where the dependent variable is either a dummy equal to one if a school was not mentioned in any application for a permanent teaching position or a set of binary indicators for whether the school was mentioned by at least a teacher whose score falls into the quintile of the distribution of the competency score reported on the x-axis. Similarly, diamonds in Panel A and B are the point estimates from a set of regressions where the dependent variable is either a dummy equal to one if a teaching position remained unfilled, or was filled by a non-certified teacher, or a set of binary indicators for whether the vacancy is filled by a teacher whose score falls into the quintile of the distribution of the competency score reported on the x-axis. Markers and vertical lines indicate the robust bias-corrected regression-discontinuity estimates and confidence interval (at the 90% level) obtained using the robust estimator proposed in Calonico et al. (2014).

labor supply does not seem to significantly affect the probability of creating new matches. While for permanent teachers this result is arguably due to the design of the assignment mechanism, for contract teachers it can be explained by the fact that there is little scope for a substantial increase in the share of filled vacancies at the margin. Increased compensations in rural schools leads, instead, to a large inflow of more competent teachers for short-term positions.[18]

2.4.4 Student Achievement

To the extent that contract teachers account for nearly half of the teaching positions in the RD sample (where each school has three teachers, on average), the increased quality of new teachers documented in the previous subsection may generate substantial improvements in student learning outcomes. We document this effect by implementing the same empirical strategy we used to identify the causal effect of compensation on teachers' outcomes. Hence,

[18]This evidence is consistent with recent findings reported in Agarwal (2017), which document that the primary effect of financial incentives were to increase the quality, not numbers, of medical residents in rural America.

we compare student test scores in schools in localities that have less than 500 inhabitants with those with a slightly larger population. In Table 2.2 we report separate results for standardized test scores in Spanish (Panel A) and math (Panel B) administered to fourth graders three years after the policy change. We focus on test scores collected at the end of the 2018 academic year, since this increases the likelihood that any given cohort of students in fourth grade has been exposed to teachers recruited through the centralized system after the introduction of the rural wage bonuses.[19]

Recall that wage bonuses apply to both incumbent and newly recruited teachers in an eligible school. Higher wages may therefore also affect the behavior of the teachers who started working in the school before the introduction of the centralized recruitment drive or the bonuses. To separate this effort margin from the selection effects of the wage bonuses, we compare schools offering higher vs lower bonuses among those that did not have an open teaching vacancy to fill in the 2015 or 2017 recruitment drives. Column (1) shows the RD estimate of the cumulative learning gains for this subsample. The point estimates are very small and statistically insignificant, suggesting that there is no effort response to higher wages for incumbent teachers. In Column (2), we focus instead on the subsample of schools with an open vacancy in 2015 and/or 2017, for either permanent or contract teacher positions. Students in these bonus-eligible schools performed much better in Spanish and math, with effect sizes of 0.3-0.35 standard deviations.

The evidence in Columns (1) and (2) of Table 2.2 suggests that the recruitment effect of the wage bonus documented in Section 2.4.3 is the main driver of the observed increase in student test scores. Consistently with the fact that higher wages do not affect the selection of permanent teachers, in Column (3) we document that in schools with open vacancies only for permanent teachers the effect of higher wages on student performance is very small and statistically insignificant.[20]

Finally, in Column (4) of Table 2.2 we consider the subsample of schools with an open vacancy for short-term teaching positions in the 2015 and/or 2017 centralized recruitment

[19] As mentioned in Section 2.2, the data available does not allow us to precisely match teachers to classes within a school, and hence we are unable to isolate the precise effect of having a better teacher (due to higher wages) in the classroom.

[20] As most of the permanent positions that remain unfilled in the assignment process are later posted as vacancies for a contract teacher (see Section 2.3.2), the sample that we use in Column (3) of Table 2.2 excludes schools that, besides having had a vacancy for a permanent position, also had an opening for a short-term position.

Table 2.2: Wage Bonus and Student Achievement

Panel A: Dependent Variable is Spanish Test (z-score)				
	No vacancy	Vacancy		
	(1)	(2)	(3)	(4)
		Any vacancy	Permanent teacher	Contract teacher
High Bonus	0.001	0.327	-0.012	0.330
	(0.160)	(0.130)	(0.195)	(0.139)
Mean dep. var. (Low Bonus)	-0.471	-0.469	-0.383	-0.491
Bandwidth	123.388	104.702	173.915	113.922
Schools	368	662	286	615
Observations	3916	9386	3355	8893

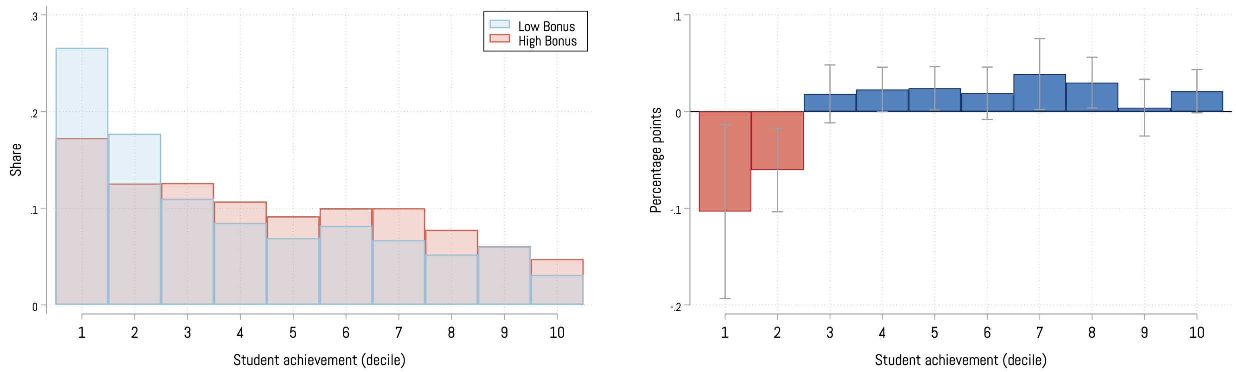
Panel B: Dependent Variable is Math Test (z-score)				
	No vacancy	Vacancy		
	(1)	(2)	(3)	(4)
		Any vacancy	Permanent teacher	Contract teacher
High Bonus	0.028	0.378	-0.016	0.483
	(0.177)	(0.144)	(0.249)	(0.160)
Mean dep. var. (Low Bonus)	-0.436	-0.408	-0.296	-0.417
Bandwidth	125.996	108.117	162.799	101.104
Schools	379	691	274	561
Observations	4014	9698	3196	8146

Notes. This table reports the effect of crossing the population threshold on student achievement in Math and Spanish. In all columns, the outcome variable is the standardized 2018 test scores in Spanish (Panel A) and Math (Panel B) for students in fourth grade. The sample in Columns (1) and (2) is split based on whether the school had an open vacancy (of any type) in the 2015 and/or 2017 centralized recruitment drives. In Column (3) and (4), the sample is further restricted to schools that had vacancies for permanent or contract teachers, respectively. Each cell reports the bias-corrected regression-discontinuity estimates obtained using the robust estimator proposed in Calonico et al. (2014). Regressions are defined within a mean-square error optimal bandwidth (BW), reported at the bottom of the table. The table also reports the mean of the dependent variable computed within the interval $(0, +BW)$ (Low Bonus). Standard errors are clustered at the school level.

drives. Consistently with the substantial increase in the competency level of newly recruited contract teachers, students in schools that receive higher wages perform much better in the Spanish and math achievement tests relative to students in schools that had contract teacher vacancies but were not eligible for the wage bonus. The effect sizes on student performance are very similar to the effect of higher wages on teacher competency scores, as shown in Panel B of Table 2.1. The magnitudes of the standardized effects reported in Column (4) Table 2.2 imply an increase of 7% in Spanish scores and of 11% in Math scores, relative to the local averages (in levels) at the right-hand side of the population cutoff (low bonus).

We further explore the relative effects of the recruitment of a more competent teacher along the test score distribution. Panel A in Figure 2.6 displays the relative shares computed

Figure 2.6: Wage Bonus and Composition Effects on Student Achievement



a. Shares at Population Cutoff **b. RD Estimates**

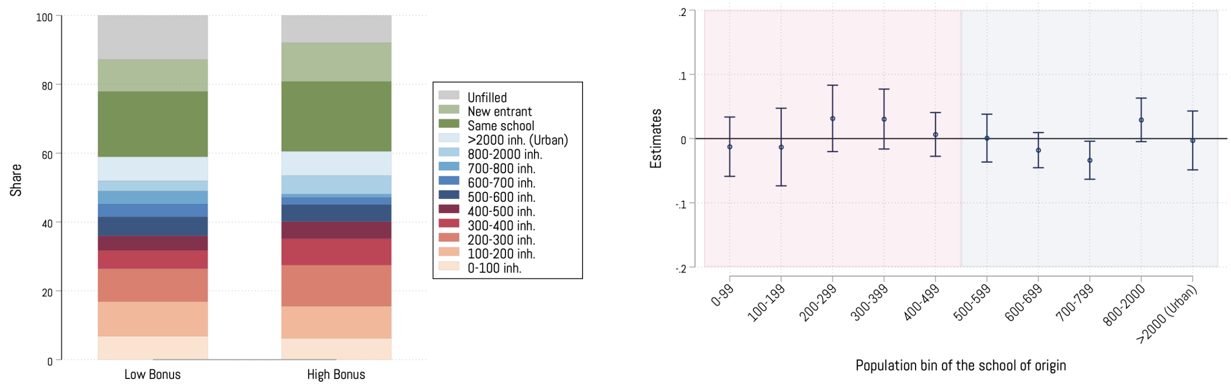
Notes. Panel A reports the relative shares of students by decile of the distribution of the average score in Spanish and math, separately for schools located to the right (Low Bonus) and left (High Bonus) of the population cutoff. Bars and vertical lines depicted in Panel B indicates the corresponding bias-corrected regression-discontinuity estimates of crossing the population threshold and the associated confidence intervals at the 90% level (Calonico et al., 2014). The sample includes schools with an open position for contract teachers.

at both sides of the population threshold by the deciles of the average score in Spanish and math. Higher wages have a more pronounced effect on reducing the proportion of students who score in the the first two deciles of the test score distribution, while the effects are smaller in magnitude and relatively uniform for better-performing students. In Panel B of Figure 2.6 we confirm these asymmetric match effects between the newly-assigned teachers and students using the deciles of the average score as dependent variables in separate RD regressions.

2.4.5 Additional Evidence

One potential concern with the identification of our main estimates is that the observed threshold-crossing effect could possibly violate SUTVA, whereby high-quality teachers who end up choosing a school in a locality with slightly less than 500 inhabitants would have otherwise chosen a school in a somewhat more populated locality. While a priori this may be an issue, we argue that it is not warranted in our setting. First, it is important to remark that differently sized localities are not necessarily geographically close to one another. In fact the median geodesic distances between the three closest below-cutoff schools and the schools just above the cutoff for the sample of contract teachers are approximately 10km, 20km, and 30Km, respectively (see Figure 2.A.5 for the full distribution).

Figure 2.7: Wage Bonus and the Origin of Newly Recruited Teachers



a. Shares at Population Cutoff **b. RD Estimates**

Notes. Panel A displays the relative shares (computed at the population cutoff) of the contract teachers who are assigned through the assignment mechanism based on the location of the previous schools recorded in the teacher occupation and payroll system (*NEXUS*), separately for schools located to the right (Low Bonus) and left (High Bonus) of the population cutoff. Panel B reports the effect of crossing the population threshold on the probability that the vacancy is filled by a teacher whose previous location falls into the population bin indicated in the x-axis. The sample includes all contract teacher vacancies assigned to a certified teacher in the 2015 and 2017 processes. Bars report the bias-corrected regression-discontinuity estimates along with confidence intervals at the 90% level obtained using the robust estimator proposed in Calonico et al. (2014).

Second, Panel A in Figure 2.7 shows the relative shares of the assigned applicants at both sides of the population threshold by the size of the localities of the schools where they were previously teaching (if any). The share of teachers who were working in a school located just above the 500-inhabitant population cutoff is small (4-5%) and it does not seem to vary between schools that are eligible for a high bonus and a low bonus. The estimates displayed in Panel B of Figure 2.7 confirms this visual pattern, showing fairly precise zero sorting effects for teachers who were previously working in schools located around the population cutoff (i.e. the 400-500 and the 500-600 population bins).

Third, in the next Section we estimate teacher preferences over wages and job attributes in order to properly construct counterfactual assignments in the absence of the wage bonus policy. We show in Figure 2.A.13 simulation-based evidence that is inconsistent with potential SUTVA violations around the population cutoff that determines eligibility to the higher wage bonus (see Section 2.5.4).

Increased competition for vacant positions can lead to an increase in the quality of applicants who select into higher-paying teaching jobs either by selecting a larger pool of prospective teachers into the public sector or by reallocating existing competent teachers from urban or other rural schools toward *Extremely Rural* locations. To show that the selection margin

Table 2.3: Wage Bonus and the Selection of New Entrant Teachers

	All	Age		Private sector experience	
	(1)	(2)	(3)	(4)	(5)
		<30	≥30	Yes	No
High Bonus	0.047	0.042	0.004	0.051	-0.003
	(0.038)	(0.017)	(0.031)	(0.028)	(0.025)
Mean (Low Bonus)	0.156	0.033	0.127	0.093	0.066
Bandwidth	139.031	155.634	159.715	135.421	168.192
Schools	801	911	935	776	993
Observations	1917	2151	2199	1869	2311

NOTES. This table reports the effect of crossing the population threshold on the selection of new entrant teachers. In Column (1), the outcome variable is a binary indicator for whether the vacancy is filled by a teacher who was not previously teaching in any public school (new entrant teacher). In Columns (2) to (5), the outcome variable is the interaction between the new entrant indicator and a set of additional characteristics of the assigned teacher. Cells report the bias-corrected regression-discontinuity estimates obtained using the robust estimator proposed in Calonico et al. (2014). Regressions are defined within a mean-square error optimal bandwidth (BW), reported at the bottom part of the table. The table also reports the mean of the dependent variable computed within the interval $(-BW, 0]$ (left-hand-side of the cutoff). Standard errors are clustered at the school×year level.

is active in our context, and hence that the results are not entirely driven by teachers sorting within the public education system, we focus on the subset of applicants who were not previously teaching in any public school. As mentioned above, these new entrants in the system represent a non trivial share of the assigned applicants who earn a position as contract teacher in our data. In Table 2.3 we show RD estimates on the pure selection effect of a higher wage bonus for the new entrants in the public sector, which is relatively large and positive (but noisy). This effect can be explained by a large and more precisely estimated inflow of recent graduates (column 2) as well as by applicants who had prior teaching experience in private schools (column 4).

Overall, the evidence suggests that the effect of the wage bonus on teachers' sorting patterns is not merely a zero-sum game. We reconsider this issue in Section 2.6.1 in the context of the estimated model of teacher preferences (see Figures 2.12-2.13). These findings provide further support for the notion that positive net inflows from the outside option partly explain the overall reallocation patterns induced by the current system of wage bonuses.

We conclude this Section by ruling out alternative mechanisms through which the wage bonus may affect student outcomes. For example, wage bonuses could affect student achievement by changing the size and composition of the teaching staff. However, Table 2.A.10 shows that the wage reform has small and statistically insignificant effects on the number of teachers, the relative share of permanent and contract teachers, and student-to-teacher ratios. Alternatively, teachers may be more likely to stay in their jobs for longer periods in the presence of the wage bonus, although Table 2.A.11 shows that wage bonuses do not affect retention rates during the study period.[21]

Taken together with the results shown in Table 2.2, this evidence strongly suggests that the inflow of more competent teachers mostly explains the large improvements in learning outcomes for the students enrolled in higher-bonus schools. While there may be an effort margin due to the wage incentives for the newly recruited teachers, the evidence reported in Table 2.A.12 documents little if no composition effects along teachers' observable characteristics. This seems to suggest that selection based on unobserved traits such as intrinsic or extrinsic motivation is unlikely to operate in this setting.

2.5 An Empirical Model of Teacher Preferences

2.5.1 Utility and Preferences

Following the discrete choice literature, we specify an empirical model of teachers' preferences that flexibly capture substitution patterns between school or local amenities and the compensation offered at every specific job postings throughout the country. We model the (indirect) utility that teacher i gets from being matched with school j as:

$$v_{ij} = \alpha_i w_j + \boldsymbol{\beta}_i' \boldsymbol{z}_j + \boldsymbol{\delta}' \boldsymbol{d}_{ij} + \boldsymbol{\lambda}' \boldsymbol{m}_{ij} + \epsilon_{ij}, \tag{2.2}$$

where w_j is the wage posted at school j in thousands of Peruvian Soles and $\boldsymbol{z}_j$ is a vector of locality and schools' characteristics that generate variation in teachers' utility across job postings. The vector $\boldsymbol{z}_j$ contains a poverty index, an infrastructure score at the locality level capturing the overall level of amenities associated to a given area, a polynomial in the population of the locality of the school and the time-to-travel (in hours) between the locality of the school and the province's capital.[22] It also includes a set of indicator variables for

[21] The effect on retention rates between academic years is partly mechanical, since these are temporary positions with a duration of one or two years.

[22] The poverty index is an asset-based measure of poverty at the individual level (poverty score) computed by the Ministry of Economy and Finance that we aggregate at the locality level. The infrastructure score collapses a set of indicators measuring infrastructure quality at the locality level through a multiple correspondence analysis (see Panel D of Table 2.A.2).

whether a given school belongs to specific regimes that determine eligibility for other wage bonuses such as multi-grade, single-teacher, bilingual, and/or to the specific geographic areas (see Figure 2.A.3).

We account for the fact that individual-specific factors may affect the extent to which teachers' labor supply vary with respect to wages and other school or locality characteristics. For example, men may be more sensitive to wages than women due to gender norms and/or gender differences in outside options. Similarly, teachers at an early stage of their professional life may be more or less sensitive to wages and other local amenities due to life cycle considerations or career concerns. We flexibly capture such patterns through the vectors α_i and $\boldsymbol{\beta}_i$, which are defined as:

$$\begin{aligned}\boldsymbol{\beta}_i &= \boldsymbol{\gamma}_0 + \Gamma_1 \boldsymbol{x_i},\\ \alpha_i &= \alpha_0 + \boldsymbol{\alpha_1'} \boldsymbol{x}_i + \sigma \nu_i,\end{aligned}$$

where $\boldsymbol{x}_i$ is a vector of indicator variables for teacher characteristics, such as gender, experience, residential location, and competency and Γ_1 is a matrix of coefficients, which is conformable with $\boldsymbol{x_i}$ and $\boldsymbol{z}_j$. We also include ν_i, a log-normally distributed random coefficient capturing unobserved preference heterogeneity for wages which would not be accounted for by $\boldsymbol{x}_i$. The presence of heterogenous preferences in our model generates flexible substitution patterns between wages and other school and locality characteristics that are key to interpreting the role of the wage schedule as well as school and locality amenities in the counterfactual analysis that we present in Section 2.6.

In addition to the fairly rich structure of preferences for the different school-level factors specified above, the discrete choice model described by equation (2.2) features two different sources of match-specific preference heterogeneity. Moving costs and other costs associated to switching jobs are captured by $\boldsymbol{d}_{ij}$, a vector of linear splines in the geodesic distance between the location of school j and teacher i, as measured by the location of the school where this teacher was working in the previous academic year. For novice teachers we use the location of the university/institute from which they recently graduated. For the remaining non-novice teachers with no prior experience in the public sector (new entrants) we use the locality of residence in 2013. Alternatively, $\boldsymbol{d}_{ij}$ may also reflect the fact that applicants may not be aware of all the available positions across the entire country and/or of their specific attributes—especially those far away from their location (see Panel B of Table 2.A.4). In this case, the parameter vector $\boldsymbol{\delta}$ should be interpreted as a combination of moving/switching costs as well as the probability that a given job posting lies within teacher i's consideration set.

The vector $\boldsymbol{m}_{ij}$ contains ethnolinguistic match effects, indicating whether teacher i's indigenous native language (if any) and school j's secondary language of instruction (if any) coincide. These capture language barriers that teachers might face when working in a school from a different ethnolinguistic group and, more broadly, any specific taste for living in a community with shared cultural traits. In settings with rich ethnolinguistic diversity, such as in Peru, these type of match effects may be particular relevant to characterize the current population of applicants (see Section 2.2). To avoid sparseness in the data, beyond the two most prominent ethnolinguistic groups (*Quechua* and *Aymara*) we consider the two most popular and well-defined indigenous groups of the Amazonian regions, the *Ashaninka* and *Awajun*, and lump together all the remaining minorities into one residual category.

All residual unobserved tastes of teacher i for school j are captured in the ϵ_{ij} term that is assumed to be distributed *iid* across i and j through a Gumbel distribution with normalized scale and location. Finally, we include all private schools that are not part of the centralized assignment mechanism or any other labor market opportunity not observed in the data as being part of the outside option.

We specify the utility of the outside option as:

$$v_{i0} = \boldsymbol{\eta}_0 + \boldsymbol{\eta}_1' \boldsymbol{q}_i + \epsilon_{i0}, \tag{2.3}$$

where $\boldsymbol{q}_i$ is a rich set of characteristics for teacher i. These characteristics include gender, experience in both the public and the private sector, ethnicity, the competency score, the population of the place of residence, and the time-to-travel between the provincial capital and the place of residence.

2.5.2 Identification and Estimation

We observe data on teachers' choices over job postings from two sources. The first data source is the rank-ordered lists of applications for permanent positions. The second source of information is the realized match for short-term positions given teachers' competency scores and choice of school district. We choose to estimate the discrete choice model presented in the previous subsection using exclusively the second source of information for several reasons. First, the vast majority of applicants are not eligible for a long-term position and among the 10% of teachers that do qualify, half either reject all offers or do not get any and eventually participate in the assignment mechanism for short-term positions (see Table 2.A.1). Second, and perhaps more importantly given our previous finding that wage bonuses do not affect the sorting outcomes of permanent teachers, studying the behavior of this sub-population becomes less relevant for the purpose of the optimal targeting of wage bonuses aimed at reducing inequalities in the allocation of public-sector teachers in Perú. Finally, the design of the assignment mechanism for permanent positions gives rise to incentives for teachers not to report their preferences truthfully in the submitted rank-order lists. Our survey elicits preferences over job postings that are unconditional on the institutional constraints of the application system. Almost one third of the surveyed teachers do not apply to their most preferred school, which clearly indicates the presence of strategic considerations in our setting (see Table 2.A.4). Learning about teachers' preferences from the available data on the rank-ordered lists would require more involved methods and additional data that go beyond the scope of this paper.[23]

None of these issues arise when focusing on contract teachers. Recall from Section 2.3.2 that within each administrative unit (school district), contract teachers are ranked based on their competency score and are sequentially assigned to their preferred school among the options that still have open vacancies. This procedure is iterated until all vacancies are filled and/or all teachers are assigned. Given the structure of the assignment mechanism, we assume that the realized matching equilibrium is stable, meaning that teachers would not be accepted by a school that they strictly prefer with respect to their current match. The assignment mechanism, indeed, directly implies that the match is stable within each school district. Overall stability might be compromised if teachers do not correctly predict in which school district their preferred feasible school is located. However, the presence of an aftermarket that assigns the remaining unfilled vacancies mitigates these concerns.

[23]Beyond a model of supply and demand, the complex nature of the assignment process would require taking into account that teachers might have biased beliefs regarding their admission chances (Kapor et al., 2020), which we don't observe in our data. It would also be important to carefully model the dynamic incentives between permanent positions and short-term positions that necessarily arise due to the sequential nature of the assignment mechanism. This extension is outside of the scope of the current project and is left for future research.

Stability implies that the observed match between schools and teachers can be interpreted as the outcome of a discrete choice model with individual-specific choice sets that depend only on teachers' competency scores (Fack et al., 2019). Under the distributional assumptions stated in Section 2.5.1, we can thus write the following log-likelihood function for the n teachers who apply to short-term positions through the centralized application system:

$$L(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \log \left\{ \int_{0}^{\infty} \left(\frac{\exp \tilde{v}_{i\mu(i)}}{\sum_{k \in \Omega(s_i) \cup \{0\}} \exp \tilde{v}_{ik}} \right) dF(\nu_i) \right\}, \tag{2.4}$$

where $\mu(i)$ is the school assignment of each teacher i, $\Omega(s_i)$ is the feasible choice set, which depends on teacher i's competency score s_i, and $\tilde{v}_{ij}$ is the deterministic component of the indirect utility function in (2.2). The term inside the brackets of equation (2.4) is the conditional probability that teacher i chooses school j from her feasible choice set, which is also a function of the cumulative distribution function of the log normal distribution, $F(\cdot)$. We compute the integral in (2.4) numerically using a Gaussian-Hermite quadrature (Judd, 1998).

In this model, preference parameters $\boldsymbol{\theta}$ are identified if (i) the observable characteristics $(w_j, \boldsymbol{z_j}, \boldsymbol{d_{ij}}, \boldsymbol{m_{ij}}, \boldsymbol{x_i}, \boldsymbol{q_i})$ are independent of both taste shifters ϵ_{ij} and the random coefficient ν_i and (ii) the feasible choice sets $\Omega(s_i)$ are independent from the taste shifters ϵ_{ij} conditional on observables. The first assumption implies that the set of observables has to be rich enough such that residual preference heterogeneity can be modeled as an exogenous shock. This might be problematic if, for instance, we believe that we are omitting a set of relevant variables that would be correlated with wages.[24] However, given that wages are set exogenously via deterministic rules and that we are controlling flexibly for all relevant wage determinants, we are confident that this assumption is reasonable in our setting. The second condition may not hold if there is a possibility that the decision by teacher i to accept or reject a given job posting may trigger a chain of acceptance or rejections by other teachers that may feed back into teacher i's set of feasible schools (Menzel, 2015). Preference cycles of this sort are ruled out in our setting, since schools rank applicants according to the same criterion (i.e. the competency score). Another potential concern that may arise in this setting is that some

[24]In the context of the centralized matching between residents and hospitals in the US, Agarwal (2015) employs a control function approach to deal with the potential endogeneity between salaries and unobserved program characteristics. The approach relies on the availability of an instrument that is excludable from the preferences of the residents.

schools of a specific type $\boldsymbol{z_j}$ may be unreachable to low scoring teachers. To mitigate these concerns and restore full support, as a proxy for teacher quality in the model, we include in the $\boldsymbol{x_i}$ vector a discrete measure of curricular and pedagogical knowledge, instead of the total competency score that determines priorities in the system (see Section 2.3.1).

2.5.3 Estimation Results

Panel A of Table 2.4 reports selected preference estimates for relevant school and locality characteristics such as wages, poverty, infrastructure, and indicators for whether a school is multigrade or single teacher. The full set of estimated parameters of the model described in Equation (2.2) is presented in Table 2.A.13. The estimated preferences for wages (α_i) are heterogeneous along both observed and unobserved dimensions. For example, male applicants are much more responsive to compensations than females. Applicants living in urban areas and more competent teachers are also more sensitive to changes in wages, which is consistent with the fact that living in cities is more expensive and that ability and/or effort are likely to determine wage sensitivity. We do not find any significant heterogeneity with respect to teaching experience in the public-sector, suggesting that many different channels may be at play that are potentially cancelling each other out. For instance, career concerns for novice teachers may push down the wage coefficient while at the same time life-cycle considerations are consistent with a positive correlation between experience and the sensitivity to the wage posted in a given location.

The large and significant standard deviation of the random coefficient ν_i displayed in Panel A of Table 2.4 indicates the presence of substantial unobserved taste heterogeneity with respect to wages that is not explained by the observed teacher characteristics included in the model. Figure 2.8 displays the wage elasticities implied by the estimates of the model. These estimates combine both observed and unobserved sources of preference heterogeneity with respect to the wages posted at each vacancy, and they range from close to 0 to around 6, with a global average of 2.19. Several interesting patterns emerge from these distributions. For instance, increasing wages seems to be a more prominent "pull" factor for attracting teachers in rural schools than in urban schools. This result highlights the trade-off between amenities, which are more scarce in rural areas, and wages, implying that wages enter more prominently into teachers' compensating differentials.

Preference estimates for other job characteristics ($\boldsymbol{\beta}_i$) are also displayed in Panel A of Ta-

Table 2.4: Model Estimates – Selected Parameters

Panel A: Wage (α) and School/Locality Characteristics ($\boldsymbol{\beta}$)										
	Wage		Poverty Score		Infrastructure		Multigrade		Single Teacher	
	0.815	(0.120)	-0.201	(0.035)	-0.054	(0.054)	-0.237	(0.119)	-0.786	(0.192)
× Male	0.611	(0.157)	0.115	(0.032)	-0.060	(0.048)	0.019	(0.099)	0.519	(0.137)
× Experience $\geq$ 4 yrs	0.070	(0.053)	0.097	(0.036)	0.132	(0.052)	-0.284	(0.118)	0.020	(0.181)
× Urban	0.115	(0.061)	-0.060	(0.044)	0.036	(0.068)	0.009	(0.170)	-0.125	(0.242)
× Competent	0.170	(0.067)	-0.065	(0.047)	0.198	(0.076)	-0.782	(0.185)	-0.752	(0.351)
Std. Deviation (σ)	0.560	(0.053)								

Panel B: Teacher-School Match Effects					
	Ethnolinguistic Match ($\boldsymbol{\lambda}$)		Switching/moving Costs ($\boldsymbol{\delta}$)		
Quechua × Quechua	1.488	(0.158)	Distance < 20km	-0.187	(0.003)
Aymara × Aymara	1.375	(0.537)	20km < Distance < 100km	-0.033	(0.001)
Ashaninka × Ashaninka	2.243	(0.558)	100km < Distance < 200km	-0.018	(0.001)
Awajun × Awajun	2.086	(1.020)	200km < Distance < 300km	-0.017	(0.002)
Other × Other	0.995	(0.113)	Distance > 300km	-0.002	(0.000)

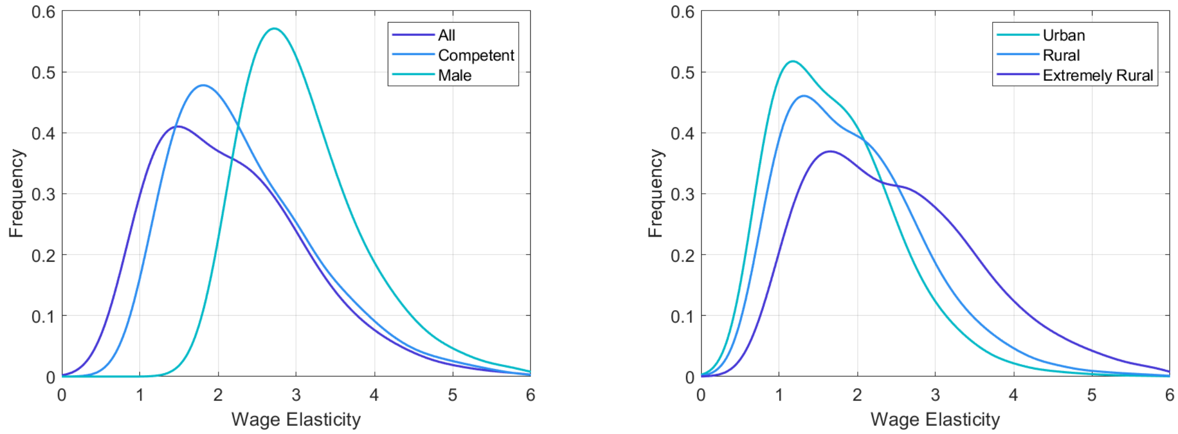
NOTES. This table displays selected estimates and standard errors (in parentheses) of the parameters of the model described in Equation (2.2). Panel A shows the estimated coefficients associated to a selected set of schools/locality characteristics while Panel B shows estimated preferences for geographical proximity as well as the interaction between schools' language of instruction and teachers' own native language. The data used contains choices of the pool of 59,949 applicants (note that 500 applicants are left out due to missing data) that participated in the allocation of short-term contracts for public primary schools in 2015. Estimates are obtained by maximizing the likelihood described in Equation (2.4) where the integral is computed numerically in an inner loop via a Gaussian-Hermite quadrature. Table 2.A.13 displays the full set of the estimated coefficients.

ble 2.4. The estimates show that on average teachers have a strong distaste for localities with high levels of poverty, for schools that are multigrade, or those with a single teacher. These patterns are more evident among competent and more experienced teachers, which suggests that complementary policies aiming at broadly improving school and locality infrastructures may be effective at reducing spatial inequalities in the allocation of public-sector teachers.

Panel B of Table 2.4 displays the ethnolinguistic match effects and the effect of the geodesic distance between teachers and schools. The magnitudes of the estimated parameters show that both play a very important role in teachers' choices over schools. Figure 2.9 documents heterogeneity across applicants in terms of the implied wages needed to compensate teachers from moving farther away from where they live (Panel A) as well as their willingness to pay for being assigned to a school offering a bilingual education that corresponds to the own ethnolinguistic group (Panel B). Moving costs are estimated to be substantial in our context. It would take on average 2.75 times the current base wage to make teachers willing to move 50km. away from where they currently live.[25] Similarly, the average teacher who

[25]As mentioned in Section 2.5.1, the estimated preference parameters for distance may also capture applicants' limited awareness about job postings that are located farther away from their current locations. Overall, the high sensitivity to distance found here is consistent with recent evidence that draws directly from the rank-ordered lists of permanent teachers in Perú (Bertoni et al., 2019).

Figure 2.8: Wage Elasticities



a) Teacher Characteristics **b) School Characteristics**

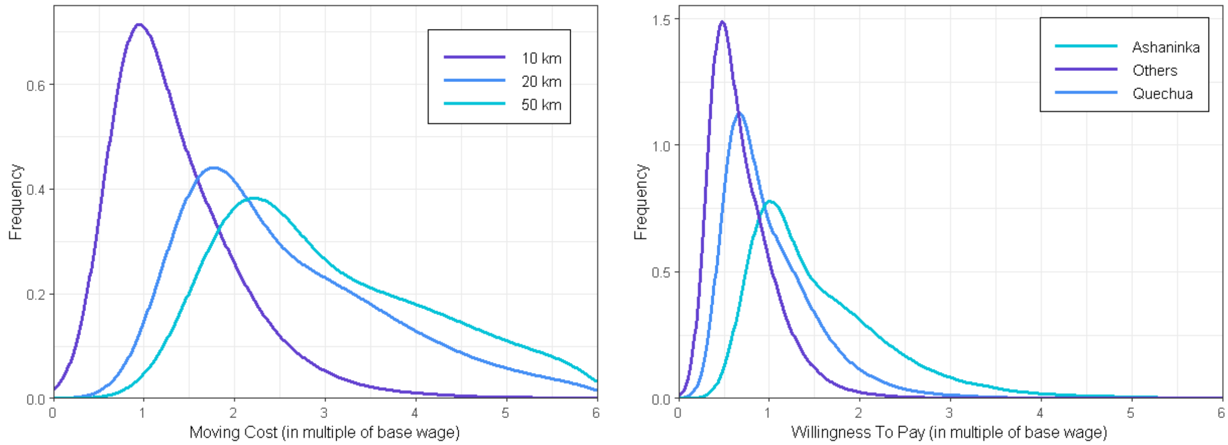
NOTES. This figure depicts the distribution of the wage elasticities which are computed using the estimates from Table 2.4. These elasticities give the % change in the conditional probability that teacher i chooses school j, which we denote P_{ij}, resulting from a 1% increase in the wage proposed in school j: $\frac{\partial P_{ij}}{\partial w_j} \frac{w_j}{P_{ij}} = \alpha_i w_j (1 - P_{ij})$. Panel A plots the distribution of this elasticity for different groups of teachers (all, competent, and male), while Panel B displays heterogeneity of this distribution with respect to the rurality of schools' locality.

speaks a native language would be willing to pay up to the amount of the base wage in order to teach in a school from her own ethnolinguistic group, with higher willingness-to-pay for the minority groups such as Ashaninka or Awajun. To the extent that these minorities are mostly located in rural areas with school vacancies that are in excess demand for bilingual teachers, place-based policies aimed at leveraging these strong match-specific effects (both ethnic and geographic) might be a promising alternative to wage incentives as a way to enhance the local supply of teachers.

2.5.4 Model Validation

In this subsection we assess the validity of our model by evaluating how well its estimated parameters predict some key moments in the data. In particular, it is important to test the empirical plausibility of the estimated wage elasticities from Figure 2.8, given that the counterfactual analysis in Section 2.6 will mainly rely on those preference parameters. To do so, we verify the consistency between the sorting patterns predicted by the model and the estimated effects at the 500-inhabitant population threshold for eligibility of the rural bonus discussed in Section 2.4. The predicted size of the effects in teacher sorting outcomes can be used for model validation since its magnitude would be entirely explained by the

Figure 2.9: Match Effects



a) Moving Costs **b) Willingness to Pay**

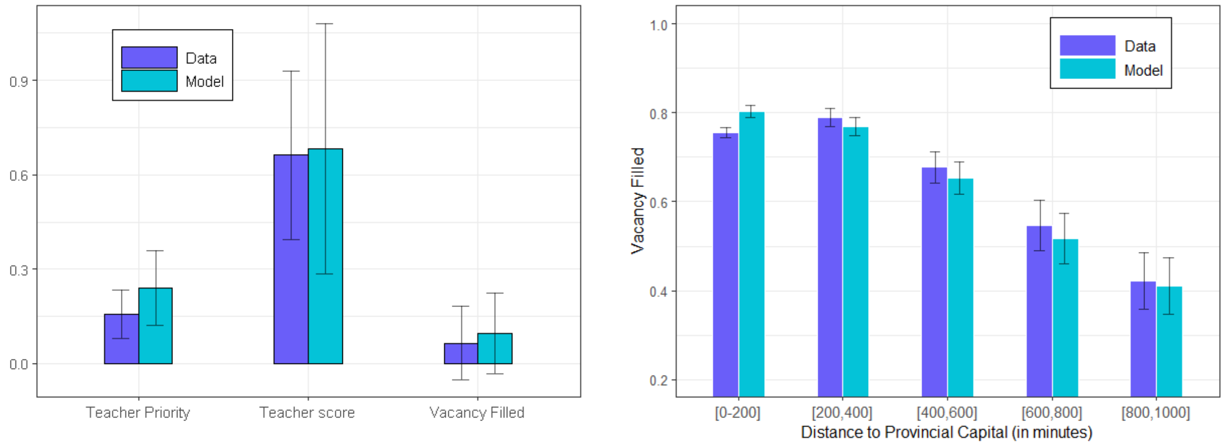
NOTES. Panel A plots the estimated distributions of the cost incurred by teachers when moving away from their previous location by 10km, 20km and 50km, respectively. These figures are computed using the estimates of the distance spline coefficients and the random coefficient on wages displayed in Table 2.4. Panel B plots the estimated distributions of the willingness to pay (in multiple of the base wage) for indigenous teachers to get assigned to (bilingual) schools with secondary language of instruction that is the same as their own language.

estimated wage elasticity. We thus simulate teachers' choices using the estimated preference parameters, replicate the RD analysis on simulated data, and compare the resulting estimates with those obtained with the actual data. In addition, we assess the overall fit of the model in terms of the global sorting patterns by the degree of remoteness of the localities where schools are situated.

Figure 2.10 shows the corresponding estimates of this exercise along with the associated 95% confidence intervals. The evidence reported in Panel A documents that the estimated model seems to predict very well the different sorting patterns as induced by the wage bonuses that we observe in the data. This validation exercise alleviates concerns about the potential correlation between wages and unobserved school characteristics (see Section 2.5.2). This is even more reassuring given that the rural bonus policy explains only a small portion (less than 10%) of the total variation in wages across job postings that is used to identify the wage coefficient in the choice model. The evidence shown in Panel B further confirms that our model precisely replicates the negative gradient between the proximity of the locality to the provincial capital and the share of filled vacancies. We provide additional measures of model fit in Figure 2.A.12.

We finally use the estimated model to provide supporting evidence for the RD analysis.

Figure 2.10: Comparing RD Estimates, Observed Sorting and Simulated Data



a) Threshold Crossing Effects (pop.=500) **b) Share of Filled Vacancies by Remoteness**

Notes. Panel A in this figure shows the estimated RD jump in vacancy filled, teacher score and the teacher priority index at the 500 locality population threshold both in the actual data and in the simulated data. The simulated assignment is generated by running the serial dictatorship algorithm using predicted utilities computed from the estimates of Table 2.4 as well as a randomly drawn set of taste shocks ϵ_{ij}. Panel B compares the share of vacancies filled in the actual data and in the simulated data depending on how far the schools posting the vacancies are located from the provincial capital.

More precisely, we use the model to evaluate whether the concerns about a possible violation of SUTVA, i.e., the possibility that high-quality teachers who sort into bonus-eligible schools would have chosen schools just above the population thresholds in the absence of the bonus, are warranted in our setting. We do this by simulating a counterfactual assignment with no rural wage bonuses and compare the resulting sorting patterns with the status quo scenario (i.e., with bonuses). Figure 2.A.13 shows RD charts based on these simulations. We find no systematic differences in teacher competency scores at the cutoff under the no-rural-bonus regime (Panel A), as expected. The introduction of bonuses at the 500-inhabitants threshold (Panel B) generates a discrete jump in teacher quality, which is comparable to the results in Panel D of Figure 2.4. More importantly, the intercepts and the slopes of the interpolating lines above the cutoff, that is, for schools in the low-bonus regime are virtually identical under the counterfactual and the status-quo regimes. This evidence is fully consistent with SUTVA.

2.6 Counterfactual Analysis

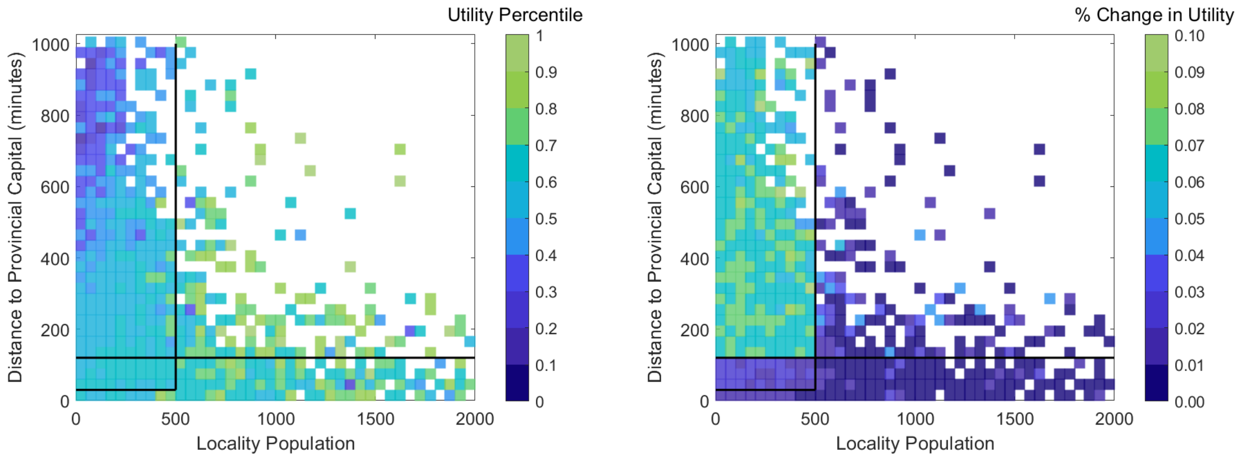
2.6.1 Evaluation of the Actual Wage Policy

Public-sector teachers who work in schools with a specific set of locality and school characteristics receive additional compensations that vary between 4% and 36% of the base wage (see Figure 2.A.3). The rural wage bonuses studied in Section 2.4 are part of this larger incentive scheme. We use the estimated preference parameters from the model in the previous Section to evaluate the effects of the overall system of wage bonuses currently in place for the universe of public sector teachers in Perú. Unlike the estimates discussed in Section 2.4, the structure of the model allows us to evaluate the policy effects away from the RD threshold, thus gaining a broader perspective on the equilibrium effects of wage bonuses on teacher sorting.

In order to generate our counterfactual of interest, we first run the serial dictatorship algorithm in which teachers are assigned to short-term positions using their estimated preferences but in the absence of any wage bonuses (including the rurality bonuses).[26] Panel A in Figure 2.11 plots the percentile of desirability, as measured by local averages in the median utility predicted by the model without any wage bonus in each school. The model estimates imply that we would need to offer the average teacher a wage that is 3.5 times higher than the base wage in order to make her indifferent between a school located in the first and in the last percentiles of desirability. The desirability index monotonically decreases with the distance to the provincial capital whereas it is only weakly correlated with the population of the locality. Schools located close to the cutoffs for eligibility to the rural bonus are not the least desirable, suggesting that some (if not most) of the effect of the wage bonus may actually show up more prominently in localities that are away from these cutoffs. This is confirmed by Panel B, which displays the cell-averages of the percentage changes in predicted utility between the status-quo (which include all the wage bonuses) and the "no bonus" counterfactual from Panel A. Changes in utility are heterogenous within the *Extremely rural* category, indicating large differences in the initial conditions of the schools that receive the same S/

[26]The school-specific and locality-specific determinants of the other wage bonuses are highly correlated with both dimensions of rurality (distance to provincial capital and population). Figures 2.A.14-2.A.15 separately show the impacts of the other wage bonuses (vis-a-vis the no-bonus scenario) and those of the rural bonus (vis-a-vis the other-bonus scenario) along the support of the univariate distributions of population and the time-to-travel to the provincial capital. The results suggest that the bulk of the policy effects on sorting outcomes are almost entirely driven by the rurality bonus.

Figure 2.11: Fitted Teacher Utility



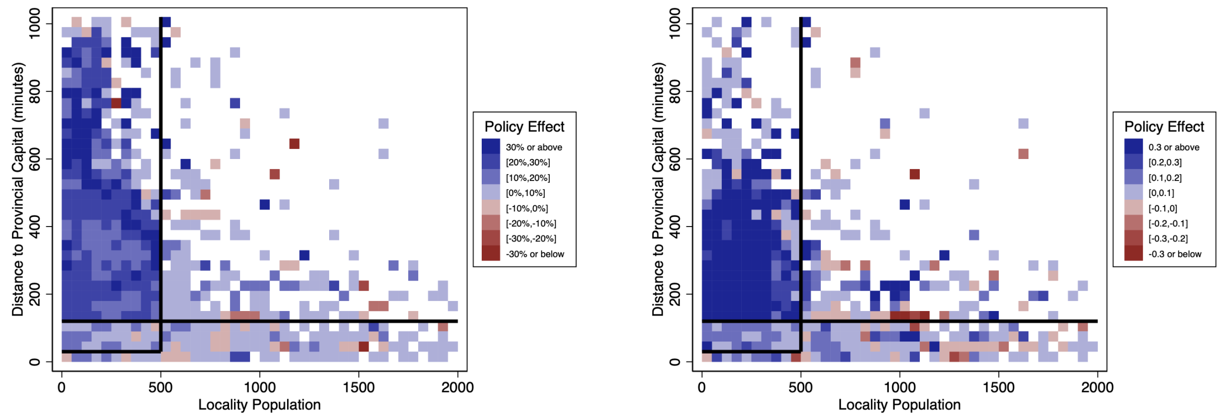
a) No Bonus **b) Wage Bonus**

NOTES. Panel A plots the average percentiles of the median predicted utility associated with each vacancy from the estimates reported in Table 2.A.13 for a fine grid in the population and distance to provincial capital space (each cell is 50×30). Panel B reports the average percentage changes in the median utility between the status quo and the counterfactual scenario with no wage bonuses.

500 rural bonus.

We next use the estimated preferences to simulate the allocation of teachers into schools under the counterfactual scenario where we remove all wage bonuses and compare it to the allocation obtained under the status quo with the system of wage bonuses actually in place. While the first two columns of Table 2.5 document some aggregate patterns related to each of these two assignments, in Figure 2.12 we display the spatially disaggregated differences between the actual wage bonus policy and the no bonus scenario. Each cell in the figure is defined by discrete values of population and time-to-travel. Most of the positive effects of the wage-incentive policy manifest in schools in localities with less than 500 inhabitants and that are farther than 120 minutes away from provincial capitals, which is due to the targeting and the magnitudes of the rural bonus. Consistently with the evidence reported in Section 2.4, the effects are not symmetric for the two sorting outcomes. Panel A shows that wage bonuses achieve a higher proportion of filled vacancies. These effects are relatively small and they do not vary systematically across the population threshold associated with the eligibility of the rural bonus, as shown by the vertical line in the figure. Indeed, the effects of the wage bonus appear more pronounced in very remote schools (i.e., in the upper left corner of Figure 2.12). Panel B instead shows larger effect sizes, with most of the effect on teacher quality that is concentrated in schools just below the population cutoff and near the time-to-travel cutoff

Figure 2.12: Policy Effects on Teacher Sorting



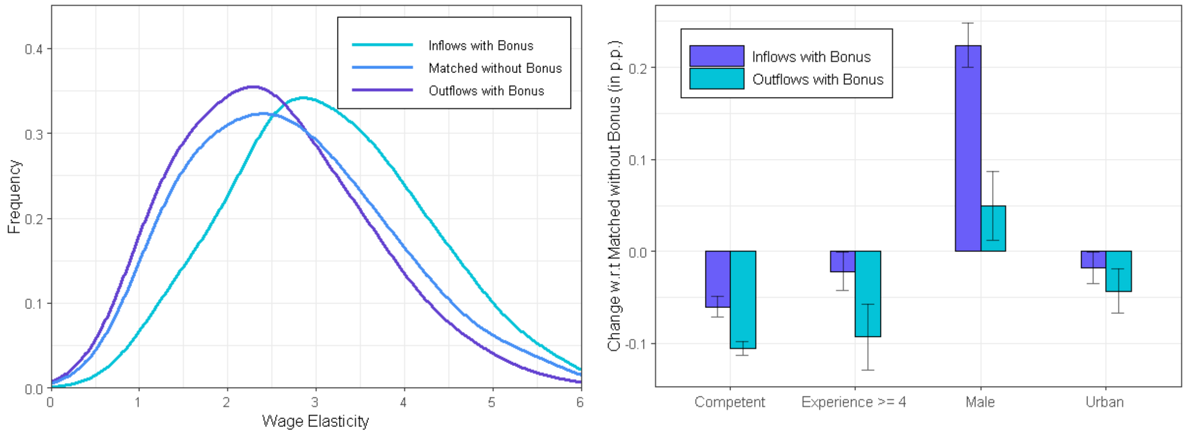
a) Share of Filled Vacancies **b) Teacher Competency Score**

NOTES. This Figure uses simulated assignment data computed by running the serial dictatorship algorithm with predicted utilities using the estimates from Table 2.4 as well as a randomly drawn set of taste shocks ϵ_{ij}. For each outcome variable, we compute kernel-weighted averages in the population and distance to provincial capital between the assignment simulated under the actual policy and a counterfactual scenario with no wage bonuses.

where the data is more concentrated (see Figure 2.3)

Importantly, our results further show that the effects of the wage bonus policy are positive across most of the support of the population and time-to-travel variables, alleviating the concern that the compensation policy generates a zero-sum reallocation. This is explained by the inflow (outflow) of teachers from (to) the outside option. We document these composition effects in Figure 2.13. Panel A compares the empirical density of the wage elasticity for assigned teachers under the no-bonus scenario with the corresponding distributions for those who choose a position under the actual system of wage bonuses and who would have otherwise (i.e. without bonuses) chosen the outside option, and for those who choose the outside option with the wage bonuses and would have otherwise been matched to a school vacancy in the absence of bonuses. As expected, the distribution of the wage elasticity for applicants who are drawn into short-term teaching jobs first-order stochastically dominates the distribution of the applicants who are displaced and/or pushed toward the outside option due to the wage incentives (p-value<0.001). Panel B displays the average percentage changes in selected characteristics between the two sub-populations of teachers who enter and exit as a result of the wage incentives with respect to the average levels of those who are matched without wage bonuses. The *Inflows with Bonus* are disproportionally more likely to be male, which is consistent with the higher wage elasticity of this sub-group of applicants shown in Figure 2.8.

Figure 2.13: The Effect of the Wage Bonus on the Selection of Teachers



a) Distributions of Wage Elasticity **b) % Change in Teacher Characteristics**

NOTES. Panel A of this Figure plots the empirical PDFs of the wage elasticity for the assigned teachers in the counterfactual scenario without any wage bonuses, along with (i) the *Inflows with Bonus* which are pulled out from the outside option thanks to the wage bonus policy and (ii) the *Outflows with Bonus* which are crowded out to the outside option because of the wage bonus policy. Panel B plots the percentage change in the average characteristics of the individuals belonging to these two groups with respect the assigned teachers under the no-bonus scenario.

They are also less competent (based on the discrete measure of curricular and pedagogical knowledge used in the model) and less experienced, when compared to the pool of existing teachers. Instead, when compared to the *Outflows with Bonus*, they are (slightly) more competent and more experienced.

2.6.2 Alternative Wage Policies

The evidence in Section 2.4.4 shows that policies that incentivize teachers sorting toward disadvantaged areas can increase efficiency along with equity given that competent teachers are more effective on low achieving students. In the last part of our analysis, we investigate whether we can achieve a more equitable allocation of teachers by redesigning the wage bonus policy. We focus on two independent policy goals that target either the extensive or the intensive margin of teachers' sorting outcomes, as discussed in Section 2.6.1. Objective (i) is having at least one filled vacancy in each school. Objective (ii) is to recruit at least one high-quality (i.e., above the median teacher in urban areas) teacher in each school.[27] Policy objective (i) and (ii) are equivalent, the only difference being that the set of appli-

[27]This threshold implies that objective (ii) mimics the size of the estimated effects of the wage bonus policy on teacher competency scores reported in Section 2.4.3—i.e. 0.45 standard deviations above the overall sample mean.

cants considered is not the same. The aim of this exercise is to determine what would be the cheapest wage bonus policy that achieves either objective (i) or (ii) *under the actual assignment mechanism in place* for contract teachers.

We consider a counterfactual economy where schools are allowed to propose different wages to teachers (Kelso and Crawford, 1982; Hatfield and Milgrom, 2005). We restrict the pool of available applicants to the set of high quality teachers under policy objective (ii).[28] To be consistent with the institutional framework, we impose that schools have to pay the same wage to all the teachers they hire. We then use the estimated preference parameters from Section 2.5 in order to infer how teachers rank each school-wage allocation, and we embed the policy objectives defined above into schools' preferences over each possible allocation through the following two assumptions:

(A1) For a fixed wage, we assume that schools have the same preferences as in the actual assignment mechanism. Teachers are individually ranked by test scores and the most preferred group of q teachers is the one composed of the q best scoring teachers.[29] Schools cannot leave a vacancy empty if a teacher would be willing to fill it at that wage.

(A2) We assume that schools have lexicographic preferences when ranking two allocations with different wages. Keeping all slots empty is dominated by any other allocation at any given wage. Otherwise, schools will always prefer the allocation with the lowest wage.

These preferences satisfy the substitute condition. The proof is provided in Appendix 2.B.1. We thus leverage the seminal result in Hatfield and Milgrom (2005) that shows that a stable set of contracts always exists in the proposed mechanism. There exists an allocation such that there is no school-teacher pair that would prefer to break their match and rematch together under any proposed wage. These stable contracts form a lattice where the largest and smallest elements are the school-optimal stable allocation and the teacher-optimal stable allocation, respectively. We can either use the school-proposing generalized DA algorithm or the teacher-proposing generalized DA algorithm to reach one or the other allocation.

We now state our main result. The proof is provided in Appendix 2.B.2:

[28]This restriction implicitly requires that there are enough high quality teachers to fill at least one vacancy per school.

[29]Hence, preference over groups of teachers are responsive (Roth and Sotomayor, 1992).

Proposition 1 *Under assumptions (A1)-(A2):*

(i) *The wage schedule and allocation resulting from the school-proposing generalized DA algorithm reaches each of the policy objectives (i) and (ii) at the lowest cost conditional on stability.*

(ii) *The allocation reached by the algorithm is implementable under the actual assignment mechanism by fixing wages to the accepted wage in each school.*

Hence, we can use the school proposing generalized DA algorithm to derive the cost-efficient wage bonus policy that would achieve either policy objective (i) or (ii) under the assignment mechanism currently in place in Peru.[30]

Table 2.5 presents summary statistics for counterfactual compensation policies, for both matching outcomes as well as the implied cost of the wage bonuses. As a benchmark, the first two columns replicate the exercise performed in the previous subsection on the evaluation of the actual wage bonus scheme. The actual policy has been effective at increasing the share of schools with filled vacancies as well as the overall quality of the recruited teachers (Panel A). While most of the benefits accrue to schools in *Extremely Rural* locations, other rural schools also benefit on average, while urban schools do not suffer any losses in matching outcomes. This evidence is consistent with the spatially disaggregated patterns depicted in Figure 2.12, whereby positive net inflows from the outside option partly explain the overall reallocation effect induced by the actual system of wage bonuses.

We explore a battery of counterfactual assignments computed under the matching-with-contracts algorithm described above (*Optimal Policy* in Table 2.5). The third and seventh columns show that by flexibly incorporating information on teacher preferences, the counterfactual policy derived under Proposition 1 achieves the same objectives of the actual system of wage bonuses at a much lower cost—23% of the total cost for objective (i) (one-filled vacancy per school) and 12% for objective (ii) (one high-quality teacher per school). Attracting competent teachers is significantly more costly than merely filling vacancies. While the counterfactual policy shown in the fourth column would fill at least one vacancy in every school at

[30]One could potentially reach the same objectives at a lower total cost by making a subset of schools deviate from the optimal stable allocation and increase wages. Such deviations may be optimal from the point of view of the overall system, illustrating a classic trade-off between stability and (aggregate) efficiency generated by the presence of externalities in two-sided matching markets. Under policy objective (ii), the allocation and wages derived are cost-efficient when considering the set of high quality teachers only. Any additional cost incurred by hiring low quality teachers is not taken into account.

Table 2.5: Counterfactual Policy Evaluation

Policy Objective			One Filled Vacancy per School				One High-Quality Teacher per School			
Wage Schedule	No Bonus	Actual Policy	Optimal Policy	Optimal Policy	Optimal Policy	Optimal Policy	Optimal Policy	Optimal Policy	Optimal Policy	Optimal Policy
Additional Features			Actual Allocation		Equal Amenities	Targeted Supply Increase	Actual Allocation		Equal Amenities	Targeted Supply Increase
Panel A: Matching Outcomes										
% Filled Schools	70.84	81.85	81.85	100	100	100	72.95	100	100	100
in Extremely Rural	55.43	75.99	75.99	-	-	-	59.42	-	-	-
in Rural	78.88	85.03	85.03	-	-	-	79.94	-	-	-
in Moderately Rural	84.73	88.36	88.36	-	-	-	85.11	-	-	-
in Urban	88.47	87.97	87.97	-	-	-	88.47	-	-	-
% Schools w. HQ Teacher	31.47	38.15	34.12	37.11	41.25	36.57	38.15	100	100	100
in Extremely Rural	16.39	28.36	21.06	23.90	29.40	22.14	28.36	-	-	-
in Rural	30.85	35.49	32.29	34.35	44.53	35.33	35.49	-	-	-
in Moderately Rural	40.46	44.85	41.22	43.51	48.09	43.89	44.85	-	-	-
in Urban	58.23	57.31	58.72	63.15	58.86	63.22	57.31	-	-	-
% Unassigned Teachers	90.45	88.85	88.85	87.51	87.44	87.30	90.09	84.84	84.86	84.82
% Unassig. HQ Teachers	85.28	82.28	84.32	83.24	81.80	83.46	82.28	59.37	59.50	59.37
Panel B: Wage Bonus (in millions of Soles)										
Total Cost	0	2.35M	0.55M	1.85M	1.26M	1.12M	0.29M	16.46M	13.64M	16.07M
Share of Total Cost										
in Extremely Rural	-	0.794	0.848	0.787	0.781	0.743	0.867	0.619	0.588	0.619
Dist.$\in [0, 400min]$	-	0.683	0.592	0.274	0.226	0.349	0.798	0.445	0.419	0.448
Dist.$\in [400, 800min]$	-	0.216	0.265	0.244	0.228	0.266	0.167	0.238	0.239	0.238
Dist.$> 800min$	-	0.102	0.142	0.481	0.546	0.385	0.035	0.317	0.342	0.315
in Rural	-	0.159	0.113	0.120	0.102	0.143	0.099	0.218	0.213	0.217
in Moderately Rural	-	0.039	0.022	0.031	0.023	0.040	0.026	0.058	0.063	0.058
in Urban	-	0.008	0.016	0.062	0.094	0.074	0.009	0.105	0.137	0.106

NOTES. This table displays the outcomes of different allocations that would result from counterfactual wage bonus policies under the assignment mechanism currently in place in Peru. For each counterfactual scenario, Panel A, describes the matching outcome by showing, by rurality category, the share of schools with at least one filled vacancy, the share of schools with at least one high-quality teacher, and the share of teachers in the outside option. Panel B displays the distribution of the counterfactual wage bonuses. *No Bonus* depicts the counterfactual scenario without all the bonuses currently in place. *Actual Policy* details the actual allocation and wage bonus policy. The remaining columns *Optimal Policy* describe the stable allocations and associated wage schedules resulting from the procedure described in Proposition 1 for both policy objectives.

a lower cost than the actual policy, it would take a total cost that is almost seven times the budget of the actual policy to fill every school with a teacher with the median competency level of urban areas in the status quo (eighth column). This can be explained by the fact that such objective would entail attracting approximately 4,000 high-quality applicants from the outside option. The share of unassigned high-quality teachers goes from 82% in the second column two to 59% in the eighth column.

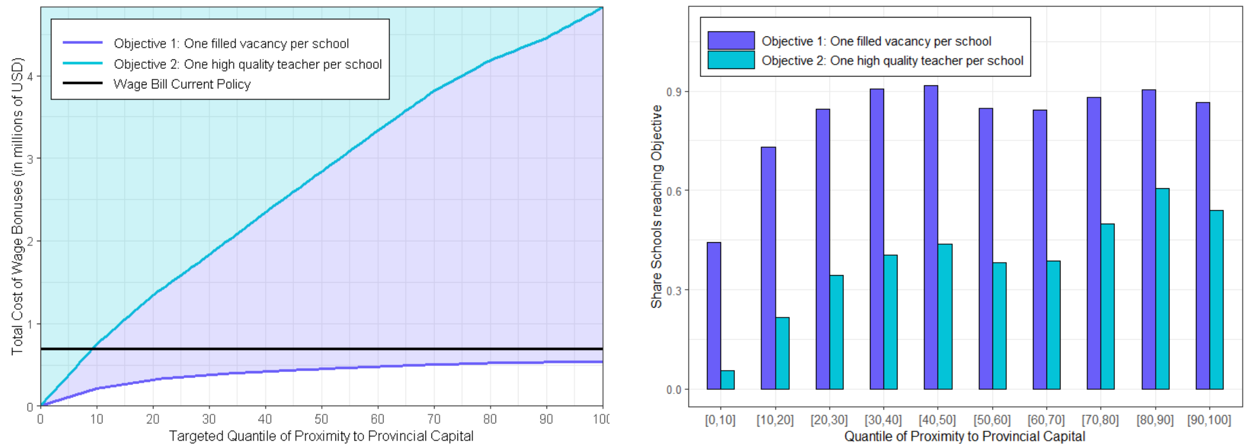
Panel B of Table 2.5 further characterizes the spatial distribution of the wage bonuses. While the actual wage bonus policy is heavily skewed toward schools in the *Extremely Rural* category, the most remote localities (Distance>800 minutes), which according to Figure 2.11 are also the least desirable for teachers, receive only 10% of the bonuses. The counterfactual policy, instead, targets those very remote localities more aggressively with almost half of the bonuses for achieving objective (i) and one-third of the bonuses for objective (ii). Urban localities receive almost no wage bonuses under the actual policy, but they are assigned a fair share of those (10% of the total cost) under the counterfactual policy when it comes to attracting high-quality teachers. This result may be explained by the fact that some urban localities may lack infrastructures and amenities that competent teachers value (see Table 2.4), which is reinforced by the upward pressure on wages due to competition among schools for relatively scarce high-quality teachers.

We next use our framework to assess the relative cost-effectiveness of additional policy instruments that may complement wage incentives in reducing spatial inequalities in the allocation of public-sector teachers. On the demand side, we remove all structural inequalities by considering a scenario where all the locality and school characteristics that potentially explain teachers' preferences are equalized across the country. Investing in local infrastructures in our setting would entail saving between 20% and 30% of the total cost in order to achieve the two policy objectives. An alternative policy consists in training prospective teachers to increase the pool of local applicants in the most disadvantaged locations. The counterfactual simulations shown in Table 2.5 mimic this supply-side intervention by "cloning" the four teachers who are most closely located to each of the 500 schools that propose the highest wages under the optimal policy. This gives a total of 2,000 new teachers –i.e., a 3% increase with respect to the overall number of applicants. Place-based incentives aimed at enhancing the local supply of teachers would entail saving 40% of the total cost that is needed to achieve objective (i). This result highlights the predominant roles of moving costs and of the ethnolinguistic match effects in explaining teachers' preferences over job postings in our setting (see Figure 2.9).[31]

It is also possible to selectively target the counterfactual bonus policy by allowing only a pre-specified subset of schools to increase wages. For example, one could be interested in knowing what would be the cost-efficient way of filling at least one vacancy or recruiting one high-quality teacher in every school belonging to a given quantile of the distribution of the proximity to the provincial capital. Figure 2.14 plots the results of this exercise. The cost-effective frontiers are concave, suggesting that achieving our policy objectives is

[31] The supply-side policy does not specifically target the overall quality of the pool of matched teachers. Hence, it is not surprising that there are no cost-advantages for achieving objective (ii) under this counterfactual.

Figure 2.14: Cost-Effective Frontiers



a) Optimal Policy **b) Actual Policy**

Notes. Panel A plots the total cost of the optimal policy targeting groups of schools which location's belong to different deciles of the remoteness distribution. Panel B plots the share of filled schools with at least one filled vacancy and the share of schools with at least one high quality teacher by decile of the remoteness distribution.

more expensive when targeting the most remote locations. This is consistent with the results reported in Panel B, which document that the actual policy is falling short on both objectives within those areas.

Panel A of Figure 2.14 confirms the findings reported in Panel B of Table 2.5, namely that attracting high-quality teachers is more challenging than filling vacancies, and hence we see that the associated cost of the policy grows large very rapidly. However, by targeting the schools in the bottom decile of the proximity distribution, the counterfactual policy dramatically improve on the actual policy for the intensive-margin objective at the same cost in terms of the wage bonuses (see the light-blue line in Panel A). Only 5% of the schools in the bottom decile of the proximity distribution have a high-quality teacher under the actual policy (see Panel B), compared to 100% of these schools in the counterfactual scenario.

There is, therefore, large scope for improvement in the actual policy by reallocating resources towards specific locations when it comes to fulfilling both policy objectives. Evidence on the concavity of the achievement production function with respect to the appointment of a high-quality teacher (see Figure 2.6) documents that students in the most remote schools are likely to be those that benefit the most from policies aimed at leveling the playing field. In these schools, a back-of-the-envelope calculation based on the RD estimates suggests that the share of students in the bottom two deciles of the test score distribution would decrease from 80% under the actual policy to less than 50% in the counterfactual policy regime at the same total cost for the government.[32]

2.7 Conclusion

Teachers are a central input to the education production function and better teachers have been shown to positively affect student outcomes, both in the short term and in the medium term (Chetty et al., 2014a,0). Providing qualified teachers with the right set of incentives to (re-)locate across the country may be one promising alternative to improve education opportunities in relatively disadvantaged areas.

Three distinctive features of our setting allow us to study teacher compensation policies and their potential for mitigating the deep and historical inequality in Peru; a large developing country characterized by a wide array of heterogeneity in geography, language, and ethnicity. First, the government uses a centralized matching platform that acts as a market clearinghouse between prospective teachers and school vacancies. Second, we rely on high-quality administrative data that link information on (i) job openings for all public schools in the country, (ii) detailed records on job applications for the universe of public-sector teachers, and (iii) student achievement in standardized tests. Third, the introduction of a wage bonus policy for positions in hard-to-staff schools with replicable and arbitrary cutoff rules provides a credible source of variation to study the effects of teacher compensation on geographic sorting.

Our first contribution is to show causal evidence that increasing teacher pay at disadvantaged locations has important selection effects. We find that unconditional wage increases are successful in attracting more competent teachers to public schools. We also document that students in schools that offer higher wages perform significantly better in standardized achievement tests. This effect can be mostly explained by large improvements at the bottom of the distribution of student test scores, and it is entirely driven by the inflow of new teachers across schools. In fact, the policy effect on student outcomes is large and significant

[32]This estimate is likely to be a lower bound since the corresponding simulated effects on teacher quality of the counterfactual policy vis-a-vis the actual policy are 2-3 times larger than the corresponding threshold-crossing effects reported in Section 2.4. The average of the standardized teacher competency scores for the schools in the first decile of the proximity distribution is -0.72 under the actual policy and it goes up to 0.50 in the counterfactual policy. Analogously, the average share of filled vacancies for the schools in these remote locations goes from 36.2% under the actual policy to 96.8% in the counterfactual policy.

for schools that had openings during the period when the policy was in place, while it is estimated to be a precise zero in schools where no new openings were available, reinforcing the argument that the selection mechanism is the driver of the results.

We then turn to quantify the way teachers trade off wages with local school and community amenities by leveraging geocoded data on applications and job postings from the centralized assignment system. The model estimates shed light on the channels through which teachers sort across locations and provide key insights on alternative policy levers beyond wage incentives that may be effective in reducing inequality in access to qualified teachers. In our model, teachers have heterogeneous preferences for locality and school amenities that are unequally distributed throughout the country. While wage profiles are rigid and do not fully take take into account these trade offs, more competent teachers seem to be more sensitive to compensation.

Overall, the evidence presented here suggests that policymakers can increase equity in the market for public-sector teachers through wage policies that take into account teacher heterogenous preferences while at the same time enhancing the efficiency of the overall system. We implement this insight by recasting the current assignment algorithm in a more general matching framework in which schools can sequentially post higher wages in order to achieve a more equal access to (high-quality) teachers across the country. The resulting alternative wage schedules are more cost-effective than the actual policy implemented in Perú and can help reduce structural inequality in access to learning opportunities. In comparison, a rigid system that ignores teacher preferences will indirectly reinforce such inequalities.

Many organizations routinely employ algorithmic pricing strategies that effectively account for demand and supply considerations in real time. Our study illustrates the untapped potential of leveraging this approach in the context of the public sector. By incentivizing sorting toward jobs or locations where working conditions are less appealing compensation policies can feasibly alter the spatial distribution of public-sector employees. These considerations can be relevant in a variety of other settings that typically feature rigid wage profiles, whereby such reallocation process is likely consequential for the quality, equity, and efficiency of public good provision.

Appendices

2.A Additional Figures and Tables

2.A.1 Descriptive Evidence

Table 2.A.1: Applicant Characteristics

	Only contract		Contract + permanent		Only Permanent	
	Mean	Sd	Mean	Sd	Mean	Sd
Age	37.72	6.934	34.50	5.802	34.48	5.465
Female	0.698	0.459	0.837	0.369	0.696	0.460
Indigenous	0.300	0.458	0.119	0.324	0.189	0.391
University degree	0.289	0.453	0.454	0.498	0.415	0.493
Curricular knowledge	40.29	13.17	67.95	6.578	70.04	7.408
Competency score	89.81	24.65	145.2	11.11	148.1	12.45
New entrant	0.344	0.475	0.313	0.464	0.166	0.372
Experience Private > 0	0.776	0.417	0.737	0.440	0.868	0.339
Experience Public > 0	0.448	0.497	0.739	0.439	0.619	0.486
Previous school: Urban	0.321	0.467	0.673	0.469	0.499	0.500
Previous school: Extremely rural	0.291	0.454	0.0852	0.279	0.189	0.391
Previous school: Rural	0.255	0.436	0.133	0.340	0.188	0.391
Previous school: Moderately rural	0.132	0.339	0.108	0.311	0.124	0.330
Number of teachers	119490		7630		8916	

NOTES. This table reports the summary statistics for the applicants to the 2015 and 2017 centralized teacher assignment system. Applicants are split in three groups: i) applicants to the contract teaching positions only; ii) unassigned applicants to the permanent teaching positions who applied to a contract teaching position; iii) applicants to the permanent teaching positions (assigned). The information on whether the applicant speaks a Peruvian indigenous language (*Indigenous*) is available for the first round of the assignment system only (2015). *Newentrant* is a dummy variable that takes value 1 if the teacher has not been employed as public sector teacher (i.e. she was not observed in *NEXUS* teacher occupation and payroll system) before the teacher assignment process. The (self-reported) information on applicants' prior teaching experience in public and private schools is collected at the time of the application.

Table 2.A.2: School and Locality Characteristics

	Rural schools		Urban Schools	
	Mean	Std. Dev.	Mean	Std. Dev.
Panel A: School characteristics				
Number of students	40.16	(45.89)	339.9	(262.0)
Bilingual school	0.249	(0.432)	0.00864	(0.0926)
Single-teacher school	0.393	(0.488)	0.0151	(0.122)
Multigrade school	0.466	(0.499)	0.0868	(0.282)
Number of teachers	5.092	(4.050)	24.59	(13.58)
% of permanent teachers	0.677	(0.468)	0.807	(0.394)
% of certified contract teachers	0.164	(0.371)	0.114	(0.317)
% of non-certified contract or other teachers	0.158	(0.365)	0.0790	(0.270)
% of competent teachers	0.210	(0.407)	0.386	(0.487)
Panel B: Student characteristics				
Math test scores (std)	-0.438	(1.005)	0.125	(0.962)
Math test scores: % Below basic	0.233	(0.423)	0.0681	(0.252)
Math test scores: % Proficient	0.147	(0.354)	0.285	(0.452)
Spanish test scores (std)	-0.568	(0.924)	0.162	(0.961)
Spanish test scores: % Below basic	0.223	(0.416)	0.0513	(0.221)
Spanish test scores: % Proficient	0.141	(0.348)	0.368	(0.482)
Panel C: School infrastructure				
No water	0.311	(0.463)	0.0355	(0.185)
No electricity	0.233	(0.423)	0.0127	(0.112)
Cafeteria	0.284	(0.451)	0.211	(0.408)
Computer	0.619	(0.486)	0.932	(0.252)
Kitchen	0.392	(0.488)	0.372	(0.483)
Internet	0.186	(0.389)	0.912	(0.283)
Library	0.207	(0.405)	0.564	(0.496)
Sport facility	0.190	(0.392)	0.614	(0.487)
Gym	0.0126	(0.111)	0.118	(0.323)
Stadium	0.00268	(0.0517)	0.0419	(0.200)
Panel D: Locality infrastructure				
Electricity	0.803	(0.398)	0.997	(0.0553)
Sewage	0.259	(0.438)	0.915	(0.279)
Library	0.0166	(0.128)	0.430	(0.495)
Doctor	0.324	(0.468)	0.869	(0.338)
Internet access point	0.0554	(0.229)	0.845	(0.362)
Village phone	0.0498	(0.218)	0.0928	(0.290)
Drinking water	0.582	(0.493)	0.945	(0.228)

Notes. This table reports the summary statistics for the universe of rural and urban primary schools in Peru over the period 2016-2018. The first panel describes the baseline characteristics of each type of school (size, bilingual spanish/indigenous language curriculum) for the year 2016, and the teaching staff composition (pooling together the post-recruitment drives years 2016 and 2018). The second panel summarizes students' achievement in the 2016 and 2018 standardized test. The third and the fourth panel describes the quality and quantity of school infrastructures and locality amenities, as measured by the 2016 school census.

Table 2.A.3: Applicant Survey (Participation and Choice Attributes)

	All Teachers Rank				Score in Top Quartile Rank			
	1^{st}	2^{nd}	3^{rd}	In Top 3	1^{st}	2^{nd}	3^{rd}	In Top 3
Panel A: Why did you apply to the centralized assignment mechanism? (% of respondents)								
Career	33.77	30.35	20.57	84.69	33.73	29.97	21.35	85.05
Stability	51.08	17.04	14.76	82.88	50.66	18.26	13.92	82.84
Formation Opportunities	9.63	29.15	21.81	60.59	9.57	26.73	20.32	56.62
Better Wage Opportunities	2.08	9.51	23.84	35.43	2.14	11.41	22.75	36.30
Social Benefits	1.04	7.78	7.96	16.78	1.10	7.00	7.58	15.68
Prestige	1.71	4.28	7.19	13.18	1.62	3.24	7.73	12.59
18 mil Soles Incentive	0.69	1.89	3.87	6.45	1.18	3.39	6.33	10.90
Panel B: What are the most important characteristic for your ranked choices? (% of respondents)								
Close to House	44.17	11.66	8.00	63.83	49.77	13.22	8.76	71.75
Safe	10.66	24.19	19.25	54.10	7.65	24.50	19.35	51.50
Well Connected	9.69	20.62	20.20	50.51	8.23	18.70	19.67	46.60
Prestige	17.92	14.12	12.29	44.33	21.13	15.77	12.68	49.58
Cultural Reasons	10.61	9.67	12.31	32.59	7.58	9.45	12.61	29.64
Good Infrastructure	2.02	8.40	12.86	23.28	1.81	7.23	11.83	20.87
Good Students	1.24	4.52	6.08	11.84	0.84	4.36	5.95	11.15
Possibility other Jobs	1.93	3.72	4.90	10.55	1.62	4.10	4.71	10.43
Career	1.76	3.10	4.09	8.95	1.36	2.67	4.44	8.47

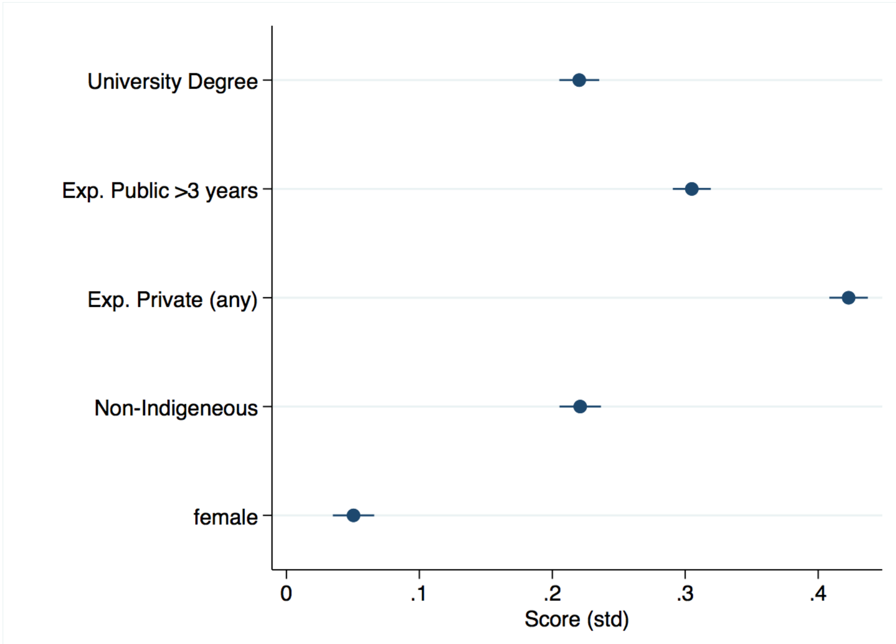
NOTES. This table displays the share of the 5,553 survey respondents that chose the corresponding answers to Question A and B. The first three columns show which answer they chose and how they ranked them (by order of importance) while column 4 shows the share of respondents that listed the corresponding choice in their top 3 reasons. The last four columns display the same results for respondents that scored above the top quartile of the test score distribution for tenured teachers.

Table 2.A.4: Applicant Survey (Strategy and Information)

	All	Score in Top Quartile
Panel A: Strategic behavior (% of respondents)		
Preferred school in concurso	63.36	61.37
If preferred school in concurso, which rank?		
Ranked 1^{st}	84.26	88.93
Ranked 2^{nd}	6.28	3.51
Ranked 3^{rd}	2.31	1.32
Ranked 4^{th}	0.71	0.66
Ranked 5^{th}	0.95	0.66
Not Ranked	5.48	4.93
If not ranked first, why?		
High demand and score too low	64.91	41.82
Remuneration not attractive	3.51	5.45
Other	31.58	52.73
Panel B: Information about first choice (% of respondents)		
Had prior information about first choice	50.97	54.01
Does your first choice benefit from wage bonus?		
Yes	16.42	15.08
No	54.53	62.69
Do not know	29.04	22.23
Expected wage - actual wage (in %)	-11.02	-8.97

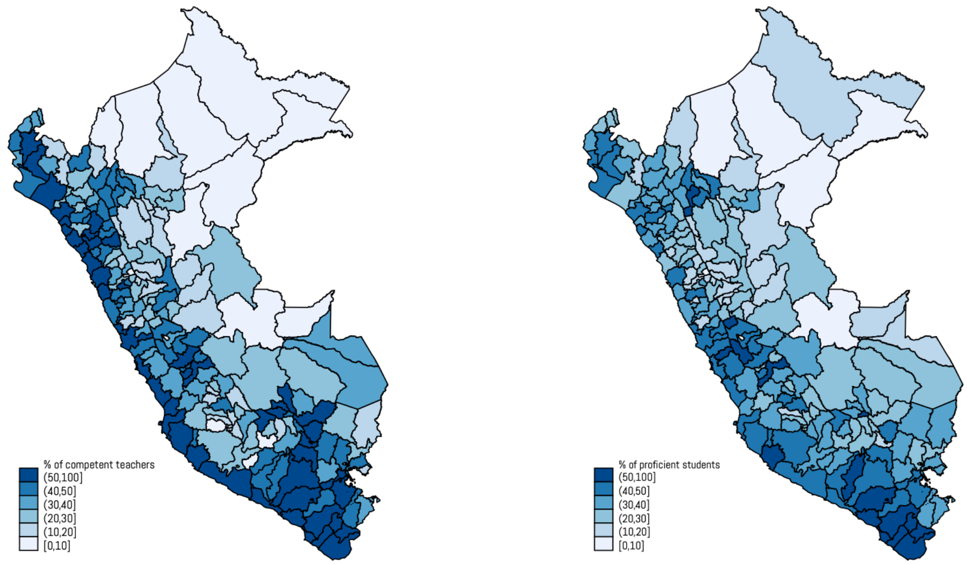
NOTES. This table displays the answers of the 5,553 survey respondents to the corresponding questions. The last columns displays the same results for respondents that scored above the top quartile of the test score distribution for tenured teachers.

Figure 2.A.1: Teacher Characteristics and Standardized Competency Scores



Notes: This figure shows OLS estimates and the associated 95 percent confidence intervals of the effect of individual teacher characteristics on the standardized competency score undertaken by all the applicants for a primary school vacancy in the context of the national recruitment drive in 2015 (see Section 2.3.2).

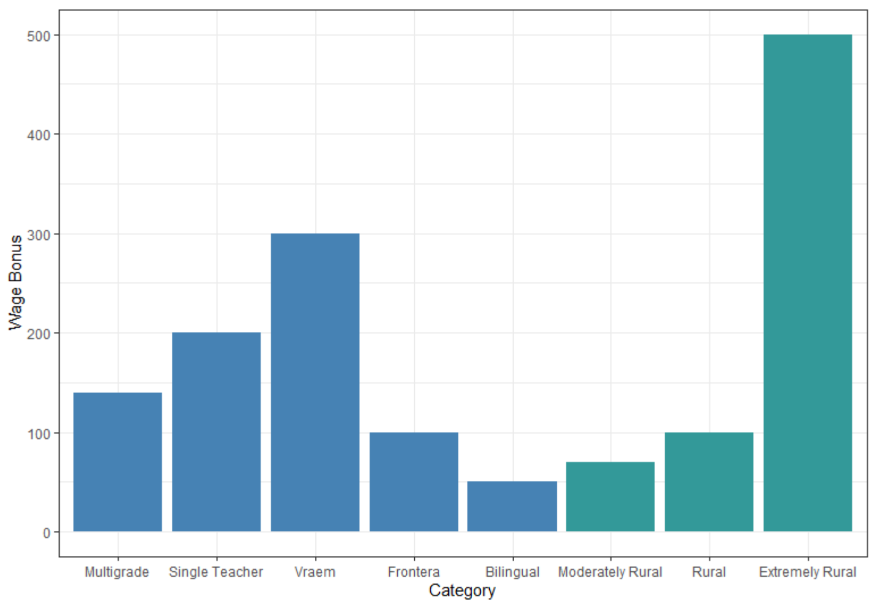
Figure 2.A.2: Geographic Distribution of Teacher Competency and Student Achievement



a) % Competent Teachers **b) % of Proficient Students**

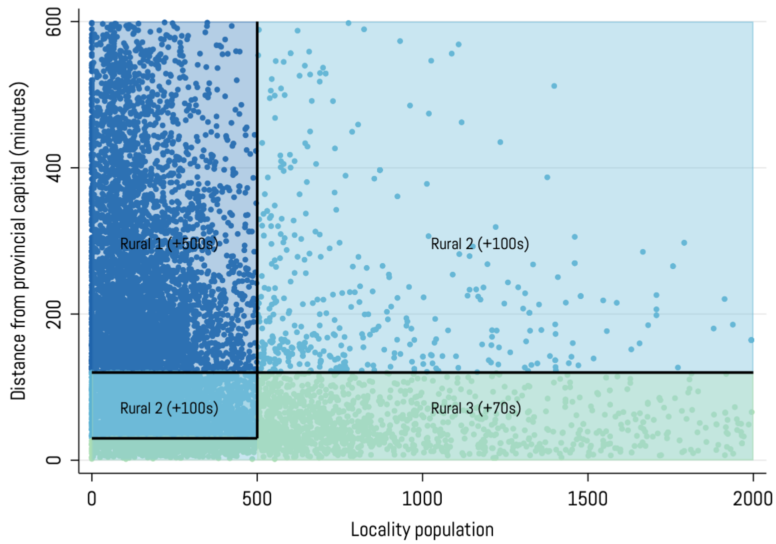
NOTES: This figure depicts the geographical variation in the share of competent teachers (panel A) and the share of proficient students (panel B) within each province of Peru. Proficient students are defined as those who attain a proficient (*Satisfactorio*) achievement level in Math and/or Spanish. Similarly, competent teachers are defined as those who attain at least 60% of correct answers in the curricular and pedagogical knowledge module of the standardized test. The reported shares are obtained by pooling the data across two school years (2016 and 2018).

Figure 2.A.3: The Different Wage Bonuses for Disadvantaged Schools

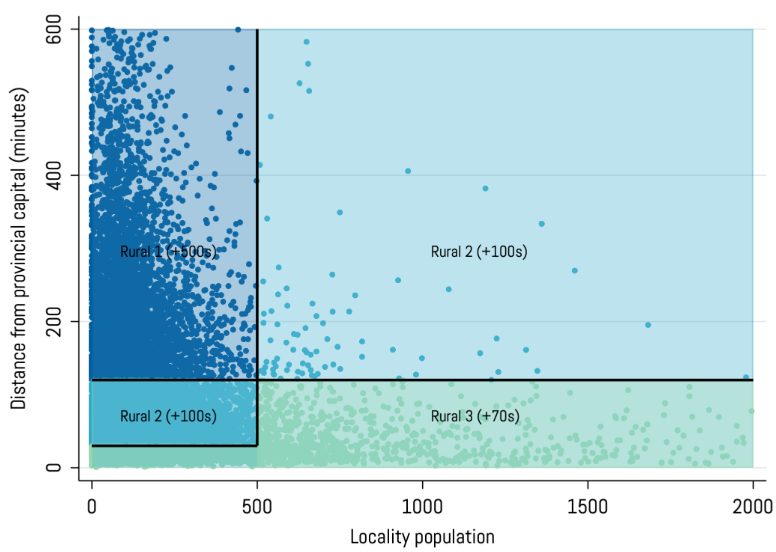


NOTES. This figure shows the monetary amount in Peruvian Soles for the different wage bonuses implemented by the Government as of December 2015. Vraem correspond to schools located in the Valle de los Rios Apurimac, Ene y Mantaro which is extremely poor and under the control of drug cartels. Frontera categorizes schools that are close to the frontier of the country.

Figure 2.A.4: The Distribution of Rural Schools over Population and Remoteness



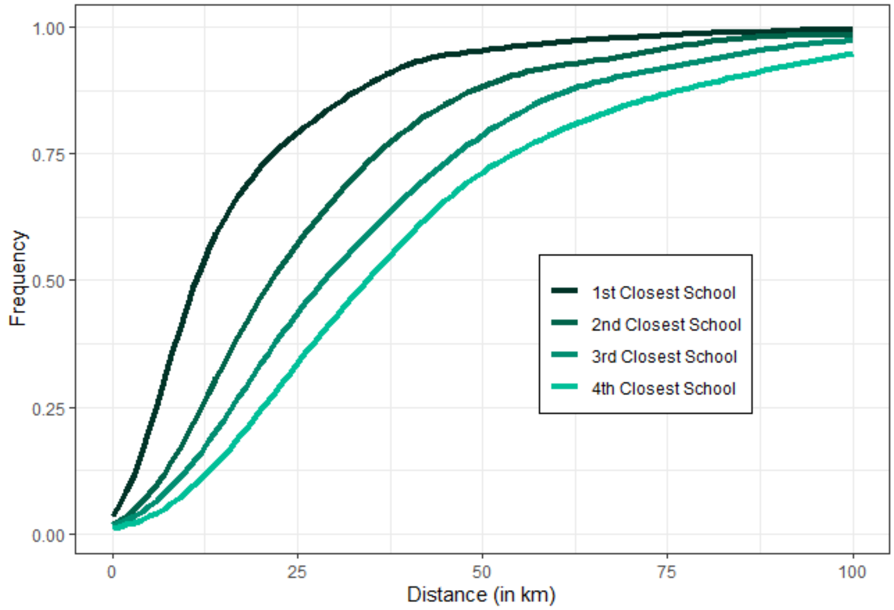
a) Schools with Vacancies in 2015-2017



b) Schools without Vacancies in 2015-2017

NOTES: This figure shows the spatial distribution of rural primary schools along the two dimensions that determine the assignment of the wage bonus. *Extremely Rural* schools are the purple dots, *Rural* are light blue and *Moderately Rural* schools are green.

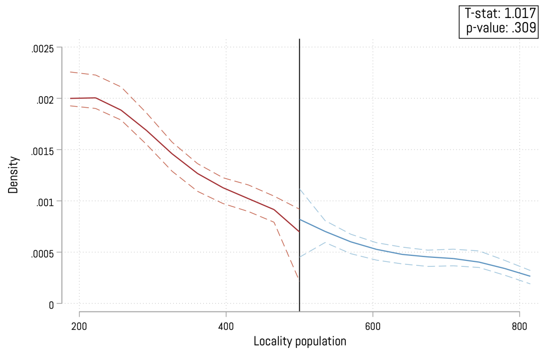
Figure 2.A.5: Distance from Schools Just Above the Population Cutoff



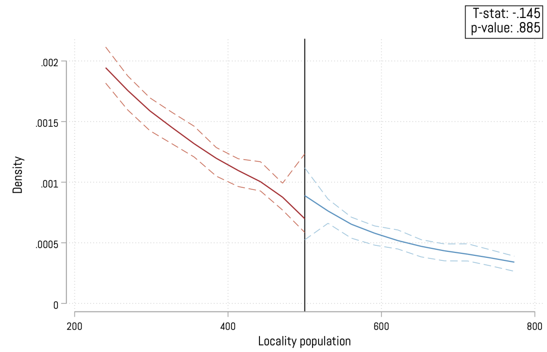
Notes: This figure plots the CDF of the distance in Kilometers for the four closest below-cutoff schools from schools just above the cutoff. The sample includes schools with an open position for contract teachers during the 2015 and 2017 recruitment drives.

2.A.2 RD Evidence

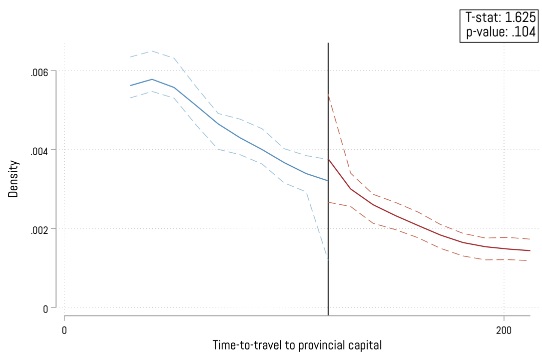
Figure 2.A.6: Manipulation charts



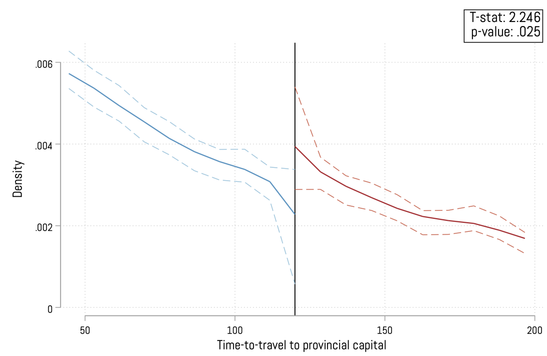
a. Population (2015)



b. Population (2017)



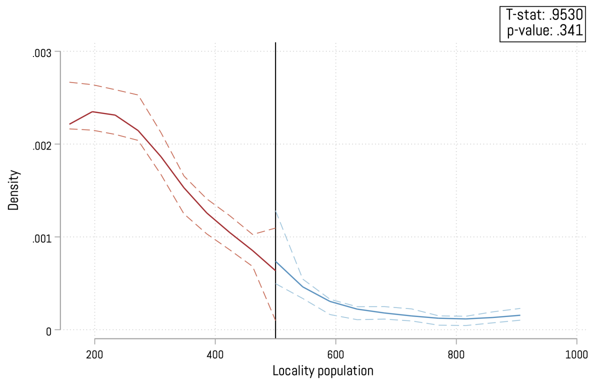
c. Time-to-travel (2015)



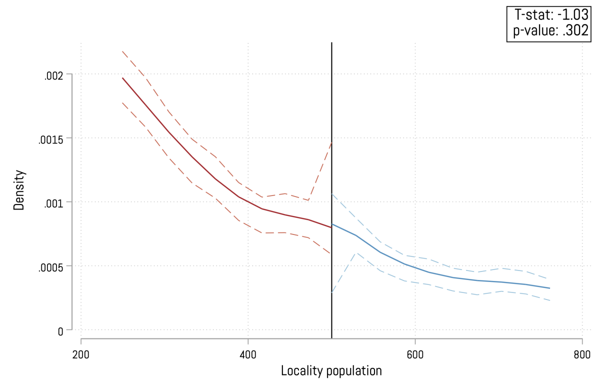
d. Time-to-travel (2017)

NOTES.
The figure displays the empirical densities with the corresponding confidence intervals for two running variables (population and time-to-travel) for each of the years in which the teacher recruitment drive was conducted (2015 and 2017). The density is computed using the local-polynomial estimator proposed in Cattaneo et al. (2020), and the figures show the 95% confidence intervals. The sample includes all schools with a permanent or contract teacher opening in the corresponding year.

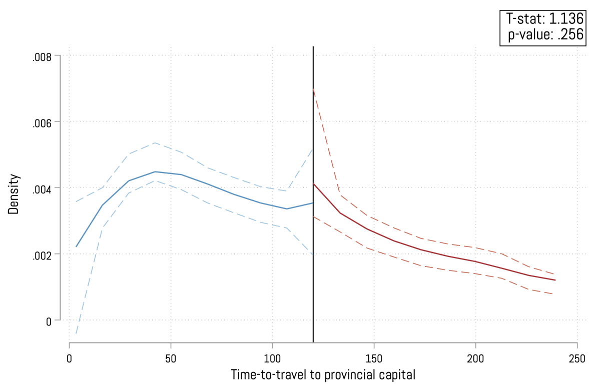
Figure 2.A.7: Manipulation Charts - Schools with a Vacancy for Permanent Teachers



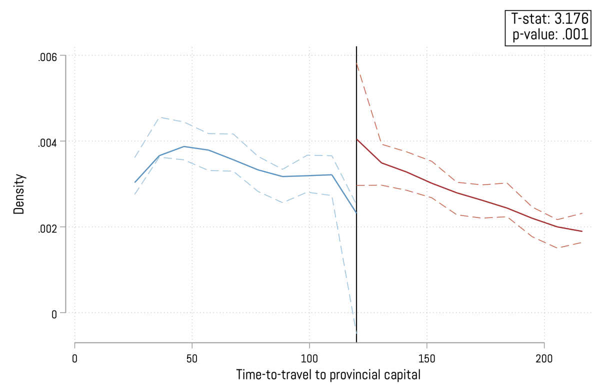
a. Population (2015)



b. Population (2017)



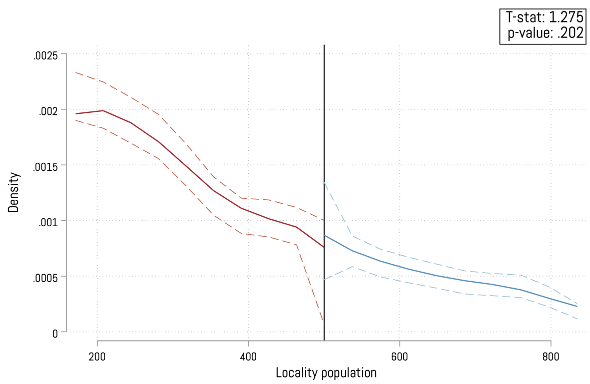
a. Time-to-travel (2015)



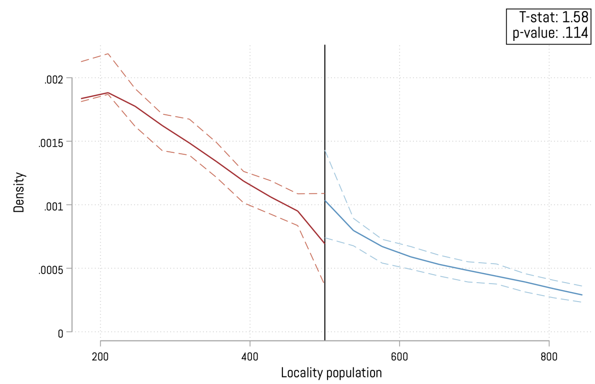
b. Time-to-travel (2017)

Notes. The figure displays the empirical densities with the corresponding confidence intervals for two running variables (population and time-to-travel) for each of the years in which the teacher recruitment drive was conducted (2015 and 2017). The density is computed using the local-polynomial estimator proposed in Cattaneo et al. (2020), and the figures show the 95% confidence intervals. The sample includes only schools with a permanent teacher opening in the corresponding year.

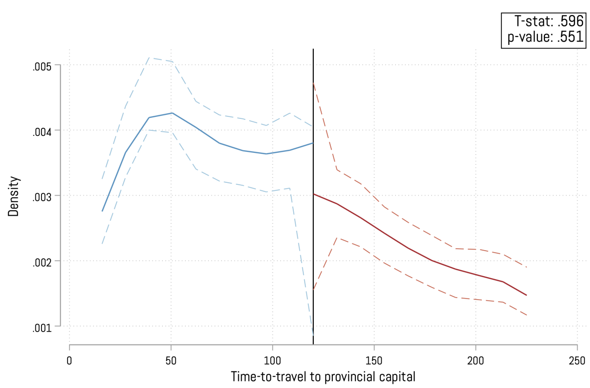
Figure 2.A.8: Manipulation Charts - Schools with a Vacancy for Contract Teachers



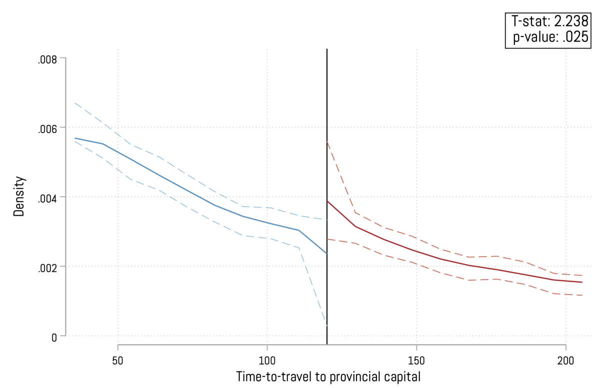
a. Population (2015)



b. Population (2017)



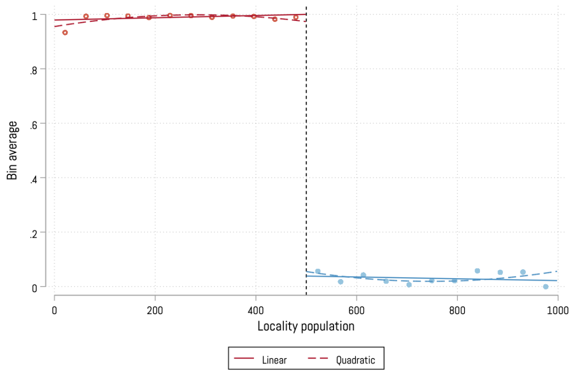
a. Time-to-travel (2015)



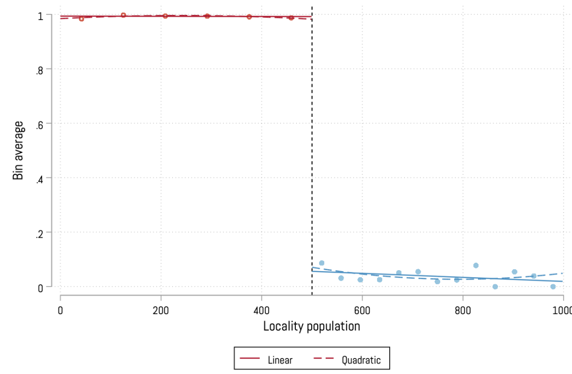
b. Time-to-travel (2017)

NOTES. The figure displays the empirical densities with the corresponding confidence intervals for two running variables (population and time-to-travel) for each of the years in which the teacher recruitment drive was conducted (2015 and 2017). The density is computed using the local-polynomial estimator proposed in Cattaneo et al. (2020), and the figures show the 95% confidence intervals. The sample includes only schools with a contract teacher opening in the corresponding year.

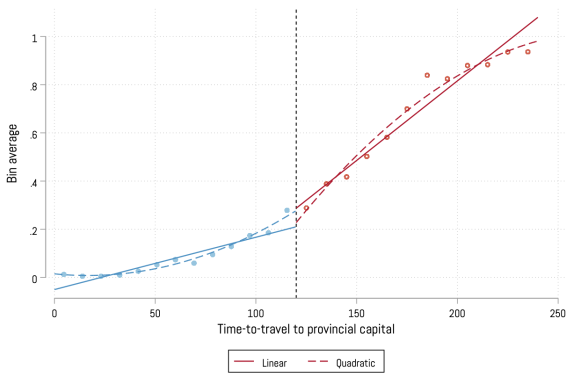
Figure 2.A.9: First Stage for Different Years and Treatment Status



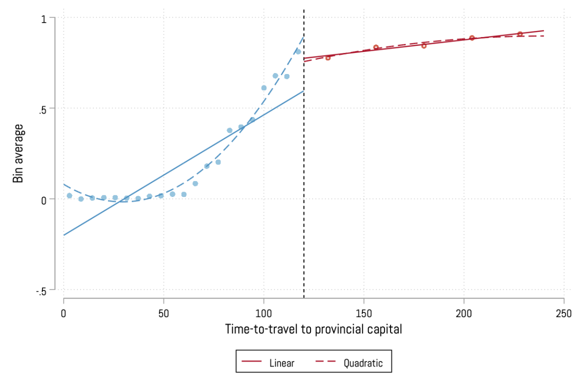
a. Treatment 2017; RV: population 2015



b. Treatment 2015; RV: population 2017



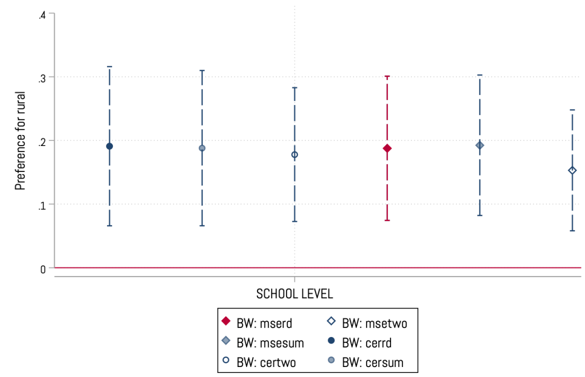
c. Treatment 2015; RV: time-to-travel 2017



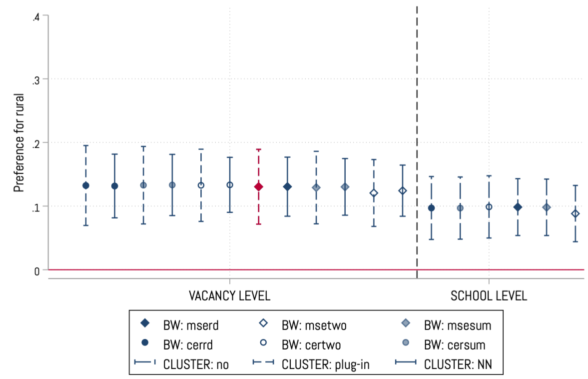
d. Treatment 2017; RV: time-to-travel 2015

Notes. The figures show the probability that a school is classified as *Extremely Rural* in each year (2015 and 2017) plotted against the two different running variables (Population and time-to-travel) for the opposite year (2017 and 2015, respectively). The regression lines are computed using linear and quadratic polynomials.

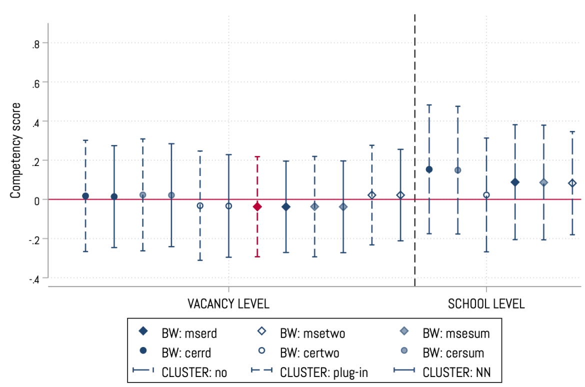
Figure 2.A.10: Robustness to Alternative RD Specifications – Teacher Outcomes



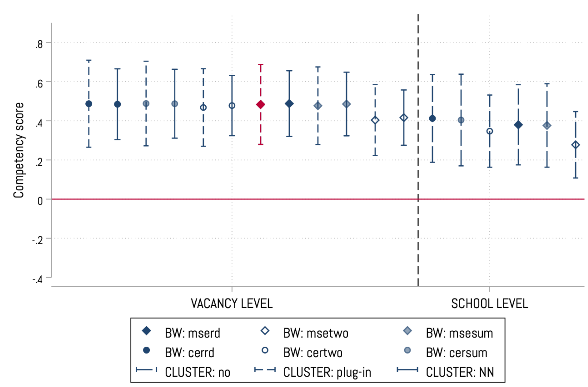
a. Stated Preferences, Permanent teachers



b. Revealed Preferences, Contract teachers



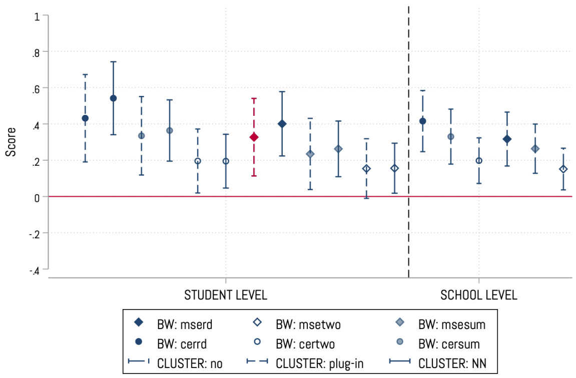
c. Competency score, Permanent teachers



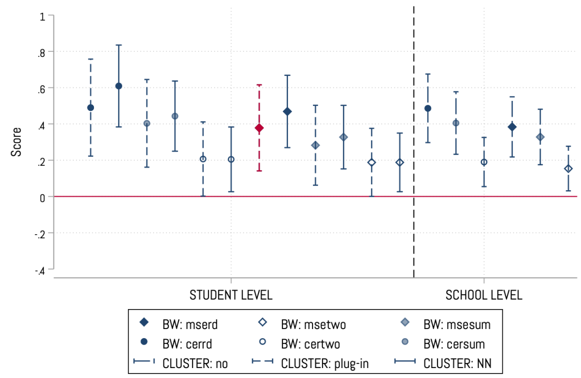
d. Competency score, Contract teachers

Notes. The figure shows how the applicants' preferences and quality vary based on the distance from the population threshold. Panels A and C focus on the assignment process of permanent teachers. In Panel A the outcome variable is a dummy equal to one if a school was mentioned in at least one application, while in Panel C the outcome variable is the standardized (total) score obtained in the centralized test by the newly-assigned permanent teacher. Panels B and D are analogous to A and C for the assignment process of contract teachers. Panel B uses as outcome variable the rank in which a vacancy was chosen in the serial dictatorship mechanism (normalized so that it takes value from zero to one), while Panel D uses the standardized score obtained in the centralized test by the newly-assigned contract teacher. Markers indicate the robust bias-corrected regression-discontinuity estimates obtained using the robust estimator proposed in Calonico et al. (2014). Regressions are defined within different specifications for the optimal bandwidths. These are: *i.* one common mean-square error (MSE) optimal bandwidth (BW: mserd); *ii.* two different MSE-optimal bandwidths, above and below the cutoff (BW: msetwo); *iii.* one common MSE-optimal bandwidth for the sum of regression estimates (BW: msesum); *iv.* one common coverage error rate (CER) optimal bandwidth (BW: cerrd); *v.* two different CER-optimal bandwidths, above and below the cutoff (BW: certwo); *vi.* one common CER-optimal bandwidth for the sum of regression estimates (BW: cersum). Vertical lines indicate confidence intervals (at the 95% level) obtained from different estimation procedures: heteroskedasticity-robust plug-in residuals (CLUSTER: no); cluster-robust plug-in residuals (CLUSTER: plug-in); cluster-robust nearest neighbor (CLUSTER: NN). The vertical dotted line separates estimates based on whether they are obtained from regressions where the unit of observation is the student (on the left) or the school (on the right). In the latter case, the outcome variables are school-level averages

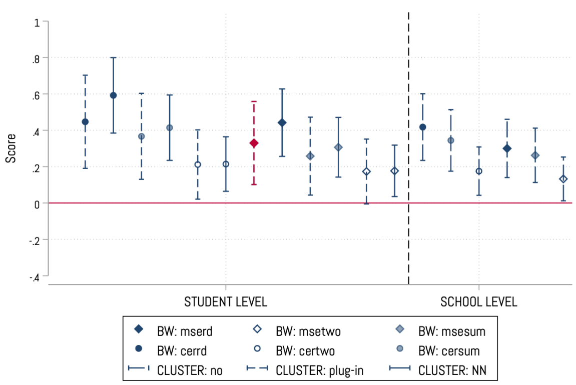
Figure 2.A.11: Robustness to Alternative RD Specifications – Student Outcomes



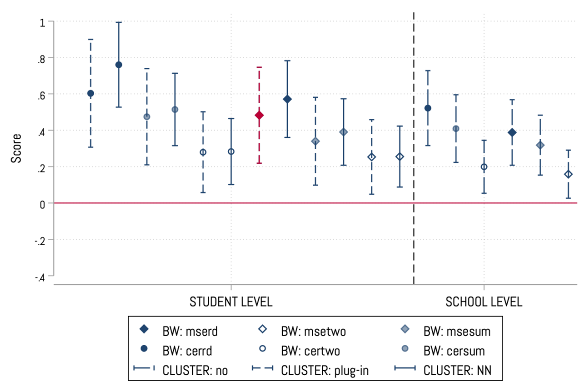
a. Spanish scores, Any vacancy



b. Math scores, Any vacancy



c. Spanish scores, Short-Term vacancy



d. Math scores, Short-Term vacancy

NOTES. This figures shows the effect of crossing the population threshold on student achievement under different specifications. The outcome variable is the average of the standardized 2018 test scores in Math and Spanish for students in the fourth grade. The sample includes schools that had an open vacancy for contract teachers the 2015 or 2017 centralized recruitment drive. Markers indicate the robust bias-corrected regression-discontinuity estimates obtained using the robust estimator proposed in Calonico et al. (2014). Regressions are defined within different specifications for the optimal bandwidths. These are: *i.* one common mean-square error (MSE) optimal bandwidth (BW: mserd); *ii.* two different MSE-optimal bandwidths, above and below the cutoff (BW: msetwo); *iii.* one common MSE-optimal bandwidth for the sum of regression estimates (BW: msesum); *iv.* one common coverage error rate (CER) optimal bandwidth (BW: cerrd); *v.* two different CER-optimal bandwidths, above and below the cutoff (BW: certwo); *vi.* one common CER-optimal bandwidth for the sum of regression estimates (BW: cersum). Vertical lines indicate confidence intervals (at the 95% level) obtained from different estimation procedures: heteroskedasticity-robust plug-in residuals (CLUSTER: no); cluster-robust plug-in residuals (CLUSTER: plug-in); cluster-robust nearest neighbor (CLUSTER: NN). The vertical dotted line separates estimates based on whether they are obtained from regressions where the unit of observation is the student (on the left) or the school (on the right). In the latter case, the outcome variables are school-level averages

Table 2.A.5: Wage increases around the population cutoff

Panel A: Permanent teacher			
	(1)	(2)	(3)
	Low bonus	High bonus	Average
High Bonus	23.321	369.796	224.931
	(17.861)	(27.099)	(29.931)
Mean dep. var. (Low Bonus)	2012.572	2107.689	2061.620
Bandwidth	149.828	307.458	222.189
Schools	361	1146	1181
Observations	599	2340	2365
Panel B: Contract teacher			
	(1)	(2)	(3)
	Low bonus	High bonus	Average
High Bonus	45.537	386.965	255.993
	(11.026)	(33.834)	(34.418)
Mean dep. var. (Low Bonus)	1906.026	1956.918	1928.570
Bandwidth	144.376	183.205	178.720
Schools	467	537	1042
Observations	827	1462	2434

Notes. This table reports the effect of crossing the population threshold on the wages of permanent (Panel A) and contract teachers (Panel B). In all columns, the outcome variable is the gross salary, which includes both the baseline wage and the bonuses. In Column (1), the sample includes only schools in rural locations whose travel time to the provincial capital is between 30 and 120 minutes, so that crossing the 500 inhabitant cutoff from above implies moving from a Moderately Rural to a Rural area. Similarly, in Column (2) the sample includes only schools in rural locations whose travel time to the provincial capital is above 120 minutes, so that crossing the 500 inhabitant cutoff from above implies moving from a Rural to an Extremely Rural area. In Column (3), the sample is the union of that in Column (1) and (2): it includes all schools in rural locations whose travel time to the provincial capital is above 30 minutes. Cells report the bias-corrected regression-discontinuity estimates obtained using the robust estimator proposed in Calonico et al. (2014). Regressions are defined within a mean-square error optimal bandwidth (BW), reported at the bottom part of the table. The table also reports the mean of the dependent variable computed within the intervals $(0, +BW)$ (right-hand-side of the cutoff) and $(-BW, 0]$ (left-hand-side of the cutoff). SE are clustered at the school level. *** p< 0.01, ** p<0.05, and *p<0.10.

Table 2.A.6: Covariate Smoothness around the Population Cutoff

	2015			2017		
	(1) Any vac.	(2) Permanent	(3) Contract	(4) Any vac.	(5) Permanent	(6) Contract
School characteristics						
Number of students	-2.912	5.555	-18.543	-1.045	-4.498	-3.479
	(10.290)	(11.990)	(11.635)	(6.499)	(8.513)	(6.736)
Indigenous language students	-0.038	-0.052	-0.056	0.017	-0.042	0.014
	(0.097)	(0.143)	(0.108)	(0.067)	(0.087)	(0.075)
% indigenous language students	-0.022	0.028	-0.030	-0.008	-0.040	0.015
	(0.085)	(0.112)	(0.103)	(0.046)	(0.066)	(0.065)
% proficient students (math)	3.863	-0.939	4.796	1.331	-4.160	2.993
	(3.144)	(7.601)	(3.305)	(3.477)	(3.511)	(3.722)
% proficient students (spanish)	6.294	5.182	8.202**	-2.264	-5.437	0.278
	(4.070)	(5.609)	(4.114)	(3.775)	(4.073)	(4.049)
Village amenities						
Electricity	0.062	0.011	0.012	0.026	-0.043	0.058
	(0.090)	(0.126)	(0.083)	(0.053)	(0.064)	(0.068)
Drinking water	0.260**	0.231	0.309**	0.110	0.174	0.144
	(0.132)	(0.173)	(0.150)	(0.083)	(0.115)	(0.101)
Sewage	0.179	0.067	0.171	-0.022	-0.030	-0.001
	(0.115)	(0.153)	(0.127)	(0.070)	(0.097)	(0.080)
Medical clinic	0.056	0.030	0.066	0.000	-0.069	0.001
	(0.107)	(0.151)	(0.122)	(0.082)	(0.100)	(0.091)
Meal center	0.186**	0.246**	0.146	0.069	0.113	0.075
	(0.087)	(0.117)	(0.101)	(0.081)	(0.093)	(0.085)
Community phone	-0.007	-0.059	-0.036	-0.034	-0.033	-0.086
	(0.093)	(0.135)	(0.114)	(0.069)	(0.091)	(0.075)
Internet access point	0.054	0.153*	0.070	0.022	-0.004	0.024
	(0.058)	(0.084)	(0.079)	(0.051)	(0.059)	(0.062)
Bank	0.023*	0.000	0.031*	0.010	0.005	0.013
	(0.013)	(0.000)	(0.016)	(0.007)	(0.008)	(0.009)
Public library	0.018	-0.059	0.019	-0.004	0.002	0.006
	(0.032)	(0.049)	(0.043)	(0.023)	(0.030)	(0.016)
Police	-0.079	-0.161	-0.094	-0.056	-0.124	-0.078
	(0.082)	(0.118)	(0.097)	(0.063)	(0.089)	(0.067)
School amenities						
Distance from district municipality (min.)	-27.579	99.432	-17.468	78.389	83.076	101.385
	(112.029)	(171.377)	(128.940)	(138.805)	(173.709)	(169.936)
Teachers room	-0.033	0.016	-0.095	-0.074	-0.177**	-0.069
	(0.072)	(0.095)	(0.084)	(0.066)	(0.075)	(0.072)
Sport pitch	-0.033	0.023	-0.041	0.002	-0.033	0.020
	(0.087)	(0.098)	(0.090)	(0.059)	(0.067)	(0.069)
Courtyard	-0.061	-0.010	-0.096	-0.116	-0.074	-0.104
	(0.092)	(0.107)	(0.100)	(0.080)	(0.087)	(0.081)
Administrative office	-0.010	-0.130	-0.094	0.056	0.032	0.048
	(0.101)	(0.155)	(0.128)	(0.077)	(0.102)	(0.094)
Courtyard	0.002	0.001	0.001	-0.009	-0.025	0.002
	(0.004)	(0.001)	(0.005)	(0.014)	(0.023)	(0.004)
Computer lab	-0.004	-0.023	-0.048	0.050	0.006	0.070
	(0.087)	(0.122)	(0.113)	(0.074)	(0.099)	(0.083)
Workshop	-0.002	-0.006	-0.020	0.010	-0.013	0.002
	(0.036)	(0.066)	(0.037)	(0.029)	(0.033)	(0.033)
Science lab	0.030	0.043	0.029	0.040	0.006	0.049
	(0.062)	(0.090)	(0.076)	(0.042)	(0.043)	(0.050)
Library	0.044	-0.115	0.007	0.094	0.076	0.044
	(0.104)	(0.159)	(0.134)	(0.071)	(0.102)	(0.095)
At least a personal computer	0.030	0.045	0.043	0.075	0.125	0.103
	(0.082)	(0.118)	(0.094)	(0.073)	(0.091)	(0.074)
Electricity	0.173	0.145	0.179	0.106	0.072	0.124
	(0.114)	(0.147)	(0.132)	(0.075)	(0.093)	(0.083)

NOTES. This table studies whether schools in localities just above or below the population threshold differ in terms of village and school amenities (as of 2013). Columns (1) to (3) focus on the 2015 assignment process, with schools split based on whether they had at least a permanent (column 2) or contract (column 3) vacancy (the sample in column 1 is the union of column 2 and 3). Columns (4) to (6) are the analogous of columns (1)-(2) but focus on the 2017 assignment process. Cells report the bias-corrected regression-discontinuity estimates obtained using the robust estimator proposed in Calonico et al. (2014). Regressions are defined within a mean-square error optimal bandwidth. Robust SE in parentheses.*** p< 0.01, ** p<0.05, and *p<0.10.

Table 2.A.7: Probability of Openings around the Population Cutoff

	All		Permanent teacher		Contract teacher	
	(1)	(2)	(3)	(4)	(5)	(6)
	Vacancy	# Vacancies	Vacancy	# Vacancies	Vacancy	# Vacancies
High Bonus	-0.012	-0.119	0.004	-0.046	-0.011	-0.116
	(0.041)	(0.138)	(0.041)	(0.091)	(0.043)	(0.134)
Mean dep. var. (Low Bonus)	0.480	0.954	0.253	0.461	0.399	0.764
Bandwidth	237.233	184.699	165.436	173.385	228.761	183.478
Observations	5912	4221	3763	3929	5612	4195

Notes. This table reports the effect of crossing the population threshold on the probability that vacancy is posted (and their number) in the 2015 or 2017 assignment process. In column (1) the outcome variable is a dummy equal to 1 if the school had at least a vacancy (of any type), while in column (2) is the number of open vacancies. Columns (3)-(4) and (5)-(6) are the analogous of columns (1)-(2) but focus only on permanent and contract teachers vacancies, respectively. Cells report the bias-corrected regression-discontinuity estimates obtained using the robust estimator proposed in Calonico et al. (2014). Regressions are defined within a mean-square error optimal bandwidth (BW), reported at the bottom part of the table. The table also reports the mean of the dependent variable computed within the intervals $(0, +BW)$ (right-hand-side of the cutoff) and $(-BW, 0]$ (left-hand-side of the cutoff). SE are clustered at the school level. *** p< 0.01, ** p<0.05, and *p<0.10.

Table 2.A.8: Monetary Incentives and Teacher Selection (2015)

Panel A: Permanent teacher			
	(1)	(2)	(3)
	Stated Preferences	Vacancy filled	Competency score
High Bonus	0.095	-0.108	0.372
	(0.085)	(0.148)	(0.384)
Bounds			[.246; .246]
Mean dep. var. (Low Bonus)	0.793	0.526	0.245
Bandwidth	238.248	209.055	152.735
Schools	552	445	170
Observations	552	604	215
Panel B: Contract teacher			
	(1)	(2)	(3)
	Revealed Preferences	Vacancy filled	Competency score
High Bonus	0.153	0.101	0.664
	(0.062)	(0.073)	(0.199)
Bounds	[.118; .181]		[.466; .74]
Mean dep. var. (Low Bonus)	0.616	0.869	-0.113
Bandwidth	156.897	200.982	144.348
Schools	402	587	365
Observations	667	978	614

Notes. This table reports the effect of crossing the population threshold on different outcomes. Panel A uses the sample of permanent teachers. In Column (1) the outcome variable is a dummy equal to one if a school was mentioned in at least one application, while in Column (2) is an indicator for whether the vacancy was filled by a certified teacher in the assignment process for permanent teachers. The regression displayed in the last column uses as outcome variable the standardized total score obtained by the teachers in the centralized test. In Columns (3) the sample is restricted to vacancies that were actually filled by a certified teacher. Panel B focuses on the selection process of contract teachers. Column (1) shows the effects on the rank in which a vacancy was chosen in the deferred acceptance mechanism (normalized so that it takes value from zero to one), while Columns (2) to (3) are analogous to those from Panel A. Cells report the bias-corrected regression-discontinuity estimates obtained using the robust estimator proposed in Calonico et al. (2014) and their bounds estimated using the procedure developed in Gerard et al. (2020). Regressions are defined within a mean-square error optimal bandwidth (BW), reported at the bottom part of the table. The table also reports the mean of the dependent variable computed within the interval $(-BW, 0]$ (left-hand-side of the cutoff). Standard errors are clustered at the school×year level. *** p< 0.01, ** p<0.05, and *p<0.10.

Table 2.A.9: Monetary Incentives and Teacher Selection (2017)

Panel A: Permanent teacher			
	(1)	(2)	(3)
	Stated Preferences	Vacancy filled	Competency score
High Bonus	0.258	0.084	-0.044
	(0.090)	(0.083)	(0.218)
Bounds			[-.517; .408]
Mean dep. var. (Low Bonus)	0.735	0.329	-0.169
Bandwidth	151.059	166.276	160.587
Schools	603	669	328
Observations	603	1240	446
Panel B: Contract teacher			
	(1)	(2)	(3)
	Revealed Preferences	Vacancy filled	Competency score
High Bonus	0.119	0.020	0.380
	(0.042)	(0.059)	(0.151)
Bounds	[.111; .119]		[.359; .362]
Mean dep. var. (Low Bonus)	0.642	0.912	0.169
Bandwidth	165.307	158.194	178.439
Schools	815	805	866
Observations	1401	1438	1482

Notes. This table reports the effect of crossing the population threshold on different outcomes. Panel A uses the sample of permanent teachers. In Column (1) the outcome variable is a dummy equal to one if a school was mentioned in at least one application, while in Column (2) is an indicator for whether the vacancy was filled by a certified teacher in the assignment process for permanent teachers. The regression displayed in the last column uses as outcome variable the standardized total score obtained by the teachers in the centralized test. In Columns (3) the sample is restricted to vacancies that were actually filled by a certified teacher. Panel B focuses on the selection process of contract teachers. Column (1) shows the effects on the rank in which a vacancy was chosen in the deferred acceptance mechanism (normalized so that it takes value from zero to one), while Columns (2) to (3) are analogous to those from Panel A. Cells report the bias-corrected regression-discontinuity estimates obtained using the robust estimator proposed in Calonico et al. (2014) and their bounds estimated using the procedure developed in Gerard et al. (2020). Regressions are defined within a mean-square error optimal bandwidth (BW), reported at the bottom part of the table. The table also reports the mean of the dependent variable computed within the interval $(-BW, 0]$ (left-hand-side of the cutoff). Standard errors are clustered at the school×year level. *** p< 0.01, ** p<0.05, and *p<0.10.

Table 2.A.10: Monetary Incentives and Teaching Staff Composition

	Permanent Vacancy			Short-term Vacancy		
	(1)	(2)	(3)	(4)	(5)	(6)
	# Teachers	Student/Teacher	% Permanent	# Teachers	Student/Teacher	% Contract
High Bonus	0.124	-0.095	0.083	-0.537	0.052	-0.045
	(0.345)	(0.182)	(0.043)	(0.372)	(0.184)	(0.036)
Mean dep. var. (Low Bonus)	6.572	2.668	0.543	6.562	2.598	0.411
Bandwidth	172.891	144.740	238.703	147.124	164.956	193.869
Observations	1033	835	1599	1152	1282	1568

Notes. This table reports the effect of crossing the population threshold on the number and the composition of teaching staff in schools that had an open vacancy in the 2015 or 2017 assignment process. The sample in columns (1) to (3) includes schools that had vacancies for permanent teachers. In column (1) the outcome variable is the total number of teachers, in column (2) is the students to teachers ratio, while in column (3) is the share of permanent teachers. Columns (4) to (6) are the analogous of columns (1)-(3) for schools that had vacancies for contract teachers. Cells report the bias-corrected regression-discontinuity estimates obtained using the robust estimator proposed in Calonico et al. (2014). Regressions are defined within a mean-square error optimal bandwidth (BW), reported at the bottom part of the table. The table also reports the mean of the dependent variable computed within the intervals $(0, +BW)$ (right-hand-side of the cutoff) and $(-BW, 0]$ (left-hand-side of the cutoff). SE are clustered at the school level. *** p< 0.01, ** p<0.05, and *p<0.10.

Table 2.A.11: Monetary Incentives and Teachers' Retention

	Permanent teachers		Contract teachers	
	(1)	(2)	(3)	(4)
	Within-year	Between-years	Within-year	Between-years
High Bonus	0.014	0.012	0.003	-0.005
	(0.020)	(0.026)	(0.007)	(0.013)
Mean dep. var. (Low Bonus)	0.905	0.099	0.970	0.919
Bandwidth	200.427	150.910	174.360	142.533
Schools	1366	998	2021	1613
Observations	5606	4187	19553	15908

NOTES. This table reports the effect of crossing the population threshold on the within- and between-years retention of contract and permanent teachers. In column (1) the outcome variable is a dummy equal to one if the teaching position is filled by the same permanent teacher at the beginning (March) and the end (December) of a school year. In column (2) it is a dummy equal to one if the position is filled by the same teacher for two consecutive years (the teacher in school year t is the same teacher observed in year $t-1$). Columns (3) and (4) are the analogous of columns (1) and (2) for contract teaching positions. The sample includes all the teaching positions in rural Peru over the period 2016-2018 that are observed for at least two consecutive years. Cells report the bias-corrected regression-discontinuity estimates obtained using the robust estimator proposed in Calonico et al. (2014). Regressions are defined within a mean-square error optimal bandwidth (BW), reported at the bottom part of the table. The table also reports the mean of the dependent variable computed within the interval (?BW,0] (left-hand-side of the cutoff). SE are clustered at the school×year level. * p¡ 0.01, p¡0.05, and *p¡0.10.

Table 2.A.12: Monetary Incentives and the Characteristics of Contract Teachers

	(1) Female	(2) Age	(3) Experience	(4) Indigenous	(5) Univ. Degree
High Bonus	0.109	-1.302	-0.009	-0.006	0.075
	(0.060)	(0.864)	(0.024)	(0.127)	(0.054)
Mean dep. var. (Low Bonus)	0.578	37.363	0.950	0.358	0.294
Bandwidth	138.955	158.719	170.756	192.227	182.079
Schools	794	930	1007	1149	1072
Observations	1761	2115	2165	853	2306

Notes. This table reports the effect of crossing the population threshold on several teachers' characteristics. These are a female dummy (column 1), age (column 2), a dummy taking value 1 for teachers with at least 3 years of teaching experience (column 3), a dummy equal to 1 if the teacher speaks a Peruvian indigenous language (column 4), an indicator for university or technical institute education (column 5). The sample includes all contract teacher vacancies assigned in the 2015 and 2017 processes, regardless of whether they were assigned to certified or non-certified teachers. In column (4) the sample includes only vacancies assigned during the 2015 assignment process, as the same information is not available for 2017. Cells report the bias-corrected regression-discontinuity estimates obtained using the robust estimator proposed in Calonico et al. (2014). Regressions are defined within a mean-square error optimal bandwidth (BW), reported at the bottom part of the table. The table also reports the mean of the dependent variable computed within the intervals $(0, +BW)$ (right-hand-side of the cutoff) and $(-BW, 0]$ (left-hand-side of the cutoff). SE are clustered at the school×year level. *** p< 0.01, ** p<0.05, and *p<0.10.

2.A.3 Teacher School Choice Model

Table 2.A.13: Preference Estimates

Panel A: School/Locality Characteristics

	Wage		Poverty Score		Infrastructure		Multigrade		Single Teacher	
	0.815	(0.120)	-0.201	(0.035)	-0.054	(0.054)	-0.237	(0.119)	-0.786	(0.192)
$\times$ Male	0.611	(0.157)	0.115	(0.032)	-0.060	(0.048)	0.019	(0.099)	0.519	(0.137)
$\times$ Experience ≥ 4	0.070	(0.053)	0.097	(0.036)	0.132	(0.052)	-0.284	(0.118)	0.020	(0.181)
$\times$ Urban	0.115	(0.061)	-0.060	(0.044)	0.036	(0.068)	0.009	(0.170)	-0.125	(0.242)
$\times$ Competent	0.170	(0.067)	-0.065	(0.047)	0.198	(0.076)	-0.782	(0.185)	-0.752	(0.351)
Std. Deviation	0.560	(0.053)								

	Bilingue		Vraem		Frontier	
	-0.747	(0.123)	-0.409	(0.284)	-0.747	(0.123)
$\times$ Male	0.011	(0.113)	-0.234	(0.187)	0.270	(0.142)
$\times$ Experience ≥ 4	-0.290	(0.112)	0.009	(0.247)	0.047	(0.155)
$\times$ Urban	-0.050	(0.166)	0.017	(0.404)	-0.135	(0.319)
$\times$ Competent	-0.732	(0.473)	-0.233	(1.063)	-0.048	(0.299)
$\times$ Lives in Vraem			0.521	(0.208)		

Rural Wage Bonus Determinants (polynomial)

log(Pop)	0.228	(0.301)	Time^3	-0.000	(0.000)
Time	-0.207	(0.097)	Time $\times$ log(Pop)	-0.002	(0.028)
$\log(\text{Pop})^2$	-0.054	(0.031)	$\text{Time}^2 \times \log(\text{Pop})$	-0.002	(0.000)
Time^2	0.011	(0.003)	$\text{Time} \times \log(\text{Pop})^2$	0.007	(0.002)
$\log(\text{Pop})^3$	0.002	(0.001)			

Panel B: Teacher-School Match Effects

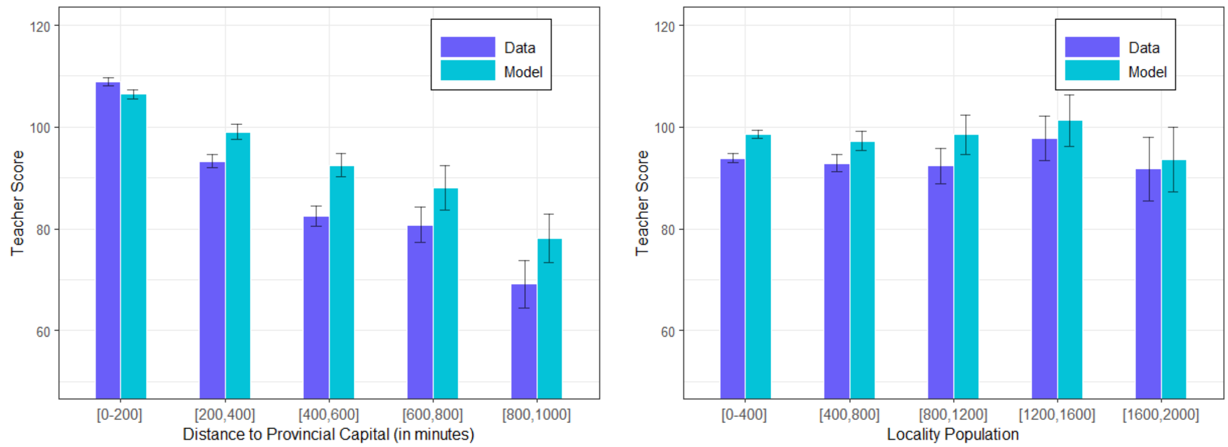
	Ethnolinguistic Match		Geographical Proximity (spline)		
Quechua $\times$ Quechua	1.488	(0.158)	Distance < 20km	-0.187	(0.003)
Aymara $\times$ Aymara	1.375	(0.537)	20km < Distance < 100km	-0.033	(0.001)
Ashaninka $\times$ Ashaninka	2.243	(0.558)	100km < Distance < 200km	-0.018	(0.001)
Awajun $\times$ Awajun	2.086	(1.020)	200km < Distance < 300km	-0.017	(0.002)
Other $\times$ Other	0.995	(0.113)	Distance > 300km	-0.002	(0.000)

Panel C: Outside Option

Constant	2.740	(1.197)	Quechua	0.527	(0.116)
Male	0.840	(0.271)	Aimara	0.214	(0.454)
Score	-0.205	(0.036)	Ashaninka	-0.564	(0.646)
Age	0.019	(0.005)	Awajun	-0.026	(0.913)
Experience	-0.043	(0.005)	Other Amazonas	-0.473	(0.067)
Private Exp > 0	0.195	(0.054)	Time	-0.059	(0.008)
			log(Pop)	0.115	(0.011)

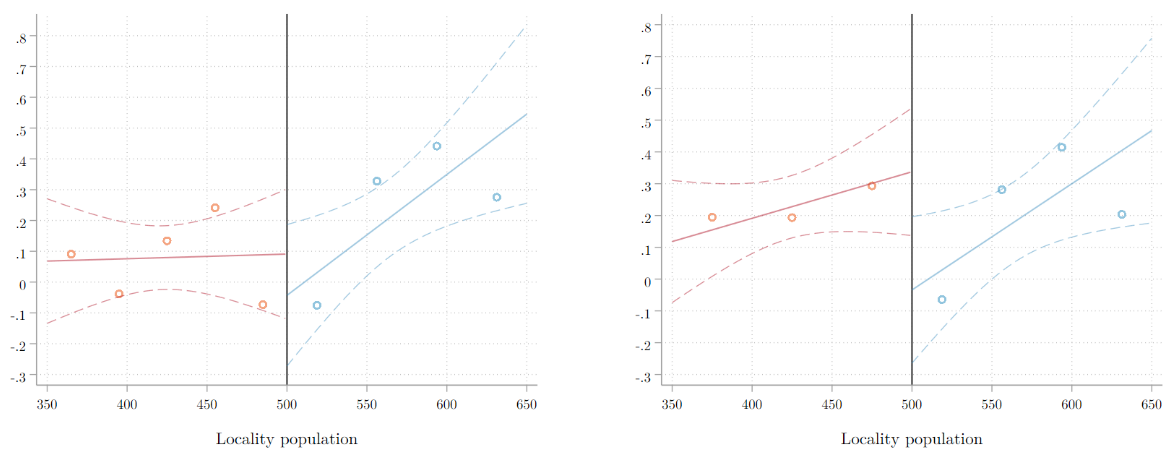
Notes. This table displays estimates and standard errors (in parentheses) of the parameters of the model described in Equation 2.2. Panel A shows the estimated coefficients associated to a selected set of schools/locality characteristics while Panel B shows estimated preferences for geographical proximty as well as the interaction between schools' language of instruction and teachers own native language. The data used contains choices of the pool of 59,949 applicants (note that 500 applicants are left out due to missing data) that participated in the allocation of short-term contracts for public primary schools in 2015. Estimation is done via maximizing the likelihood described in Equation 2.4 where the integral is computed numerically in an inner loop via a Gaussian-Hermite quadrature.

Figure 2.A.12: Model Fit with Respect to the Competency Scores of the Assigned Teachers



Notes. This figure uses simulated assignment data which is generated by running the serial dictatorship algorithm using predicting utilities computed from the estimates from Table 2.4 as well as a randomly drawn set of taste shocks ϵ_{ij}. It then compares the average score of teachers assigned to vacancies observed in the actual data and the simulated data depending on the associated school's distance to the provincial capital and locality population.

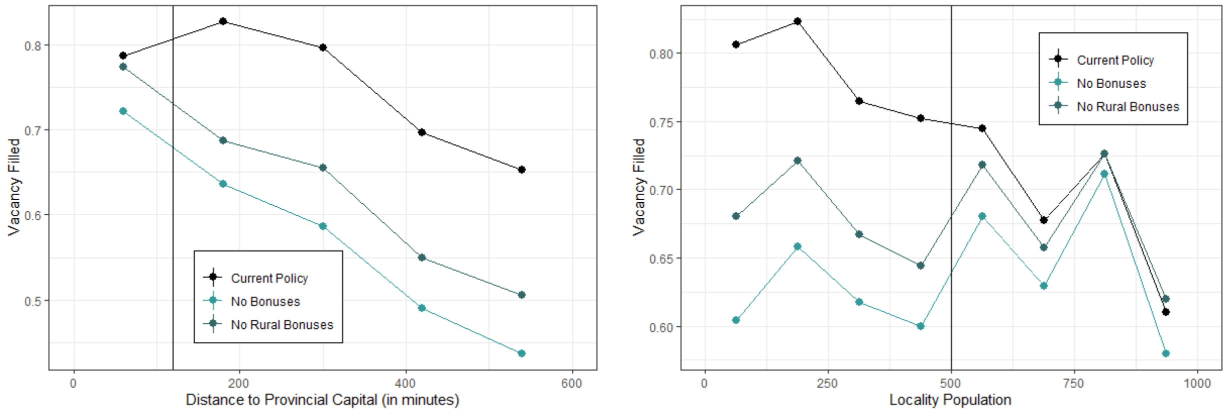
Figure 2.A.13: Simulated Threshold-Crossing Effects With and Without Wage Bonus



a) Without Rural Wage Bonus **b) With Rural Wage Bonus**

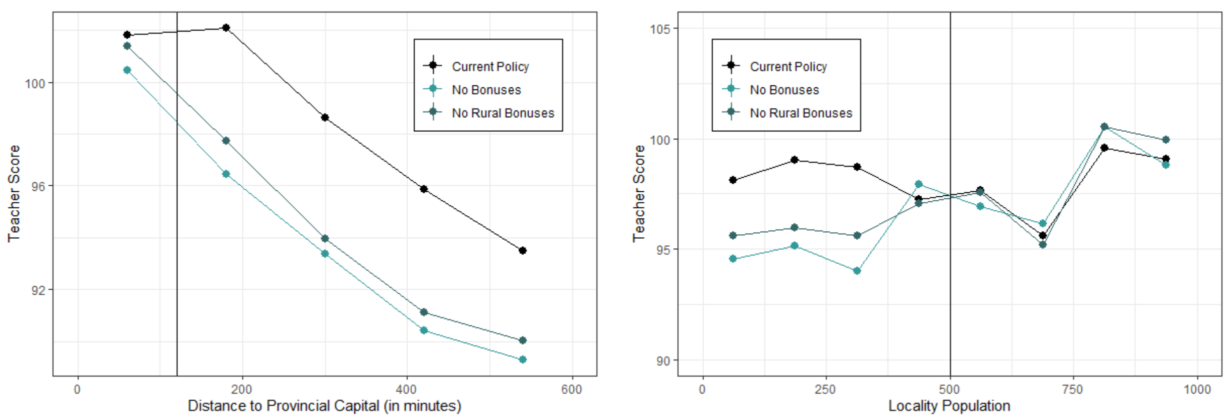
Notes. This figure uses simulated assignment data which is generated by running the serial dictatorship algorithm using predicting utilities computed from the estimates from Table 2.4 as well as a randomly drawn set of taste shocks ϵ_{ij}. The counterfactual scenario depicted in Panel A is computed assuming the presence of all the existing wage bonuses except the S/ 500 rural wage bonus for localities with population smaller than 500 inhabitants and time-t-o-travel distance to the provincial capital higher than 120 minutes.

Figure 2.A.14: The Effect of the Wage Bonus on Vacancy Filled



Notes. This figure uses simulated assignment data which is generated by running the serial dictatorship algorithm using predicting utilities computed from the estimates from Table 2.4 as well as a randomly drawn set of taste shocks ϵ_{ij}. It then compares, along the population and distance to provincial capital dimension, the average score of teachers assigned to vacancies under three counterfactual scenarios: (a) under the current policy, (b) in the absence of all wage bonuses, (c) in the absence of rural wage bonuses only.

Figure 2.A.15: The Effect of the Wage Bonus on Teachers' Competency Scores



Notes. This figure uses simulated assignment data which is generated by running the serial dictatorship algorithm using predicting utilities computed from the estimates from Table 2.4 as well as a randomly drawn set of taste shocks ϵ_{ij}. It then compares, along the population and distance to provincial capital dimension, the share of vacancies filled under three counterfactual scenarios: (a) under the current policy, (b) in the absence of all wage bonuses, (c) in the absence of rural wage bonuses only.

2.B Proofs

2.B.1 School Preferences Satisfy the Substitute Condition

Denote the set of all possible contracts $X = S \times T \times W$ where S is the set of schools, T the set of teachers we consider and W the set of wages that schools can propose. Under objective (i), we assume that T is the set of all teachers whereas we restrict T to be the set of high quality teachers under objective (ii). We assume that wages range discretely from the minimum wage proposed to teachers in Peru to an arbitrarily large upper bound.

Consider X' a subset of X. Define $C_s(X')$ and $R_s(X')$ the chosen set and the rejected set of school s. We assume WLOG that $C_s(X')$ is not empty. Otherwise this would imply, under (A1), that X' is also empty. Define w^* the wage offered in $C_s(X')$ and define $\underline{t}^*$ as the teacher with the lowest test score in $C_s(X')$. Under (A2), we know that w^* has to be the lowest wage offered in any of the contracts in X'. Consider now that we add an additional contract to X' such that $X'' = X' \cup \{(s,t,w)\}$. Under (A2), we know that if $w < w^*$ the new chosen set will be $C_s(X'') = \{(s,t,w)\}$ and the rejected set will be $R_s(X'') = R_s(X') \cup C_s(X')$. If $w > w^*$, the chosen set does not change $C_s(X'') = C_s(X')$ and the rejected set becomes $R_s(X'') = R_s(X') \cup \{(s,t,w)\}$.

If $w = w^*$, two cases may arise.

- If the size of $C_s(X')$ is strictly smaller than school s capacities, under (A1), we have that $C_s(X'') = C_s(X') \cup \{(s,t,w)\}$ and $R_s(X') = R_s(X'')$.

- If the size of $C_s(X')$ is equal to school s capacities (school s is at max capacity), under (A1) we have: (i) $C_s(X'') = C_s(X')$ and $R_s(X'') = R_s(X') \cup \{(s,t,w)\}$ if t is ranked lower than teacher $\underline{t}^*$, or (ii) $C_s(X'') = C_s(X') \setminus \{(s,\underline{t}^*,w)\} \cup \{(s,t,w)\}$ and $R_s(X'') = R_s(X') \cup \{(s,\underline{t}^*,w)\}$ if t is ranked higher than $\underline{t}^*$.

In any case, $R_s(X') \subseteq R_s(X'')$.

2.B.2 Proposition 1

Under (A1)-(A2) stability implies that every school fills at least one vacancy for policy objective (i) and every school is matched with at least one high-quality teacher for policy objective (ii). Assuming that a given school has not reached the targeted policy objective would contradict stability given that schools would be willing to increase wages until they do so. Also, we know that the school-proposing generalized DA algorithm gives the stable allocation maximizing the individual welfare of the schools. This means that, conditional on stability, the sum of the wages offered is minimal, which proves part (i) of Proposition 1.

Given the wages offered, the matching outcome is stable also with respect to the priorities used in the initial mechanism. This implies that the same allocation can be implemented in the initial mechanism by fixing wages to the derived accepted wages, which proves part (ii) of Proposition 1.[33]

[33]Under policy objective (ii), a similar argument applies when restricting the set of applicants to high quality teachers. However, given that low quality teachers have a lower priority than high quality teachers in the current mechanism, we can treat the allocation of the remaining vacancies to low quality applicants separately in order to simulate the equilibrium.

Chapter 3

Labor Market Dynamics and Teacher Spatial Sorting

Abstract. This paper provides a unifying explanation for the lack of supply of skilled teachers in remote locations. I build an empirical model of dynamic two-sided matching to link teachers' and schools' preferences with equilibrium sorting and job-to-job flows. I show that this mapping is invertible such that preferences can be identified and estimated from observed matches. Taking these tools to panel data on the assignment of public teachers in Peru, I show that the spatial disaggregation of labor demand coupled with the concentration of labor supply in cities imply the existence of a *spatial job ladder*. Low quality teachers get displaced in remote schools and move toward urban schools by climbing up the ladder once they have accumulated experience and skills. Labor mobility thus magnifies the urban-rural gap in teacher quality by one third. Dynamic wage contracts that foster retention can largely mitigate this effect.

3.1 Introduction

Many public and private services are provided locally and require the presence of a skilled workforce on-site. In such labor markets, the distribution of workers across locations has important welfare consequences. Unequal access to essential services such as education, childcare or healthcare directly contributes to spatial inequalities. Moreover, geographical differences in the overall quality of local services and amenities are key drivers of the spatial distribution of human capital, creating a feedback loop that would reinforce existing inequalities (Diamond and Gaubert, 2022). Understanding what drives worker sorting and mobility across locations is thus a first-order concern.

This paper studies this question in the context of the provision of an essential local public service: education. Teachers are key inputs of school quality (Rivkin et al., 2005) and strong predictors of students' later outcomes (Chetty et al., 2014b). Evidence of heterogeneous teacher effects further reveals that low ability students can potentially benefit more from being exposed to good teachers (Ahn et al., 2021; Bobba et al., 2022). This implies that an unequal access to skilled teachers can harm both equity and efficiency. However, analyzing sorting and mobility in teachers' labor markets is challenging as (i) wages are often set through collective bargaining and do not adjust to local labor market conditions and (ii) positions are often allocated through frictionless centralized clearinghouses. Job search or spatial equilibrium models are not tailored to such settings as they rely on wages to clear local labor markets or search frictions to rationalize sorting and job-to-job flows (Diamond, 2016; Moscarini and Postel-Vinay, 2018). Instead, an emerging literature has relied on empirical models of two-sided matching to study the role of workers' idiosyncratic preferences over job attributes in shaping sorting when prices are fixed (Agarwal, 2015; Bobba et al., 2022; Bates et al., 2022). Yet, these papers abstract away from labor market dynamics, making the analysis of sorting and mobility incomplete.

This paper bridges these literatures by incorporating dynamics into an empirical model of two-sided matching. It then applies these novel tools to study the causes of teacher spatial sorting and mobility and their consequences on spatial inequalities in access to skilled teachers.

I make several methodological and empirical contributions. First, I build a model of dynamic two-sided matching with non-transferable utility where forward-looking agents repeatedly meet in a single market and form matches according to their idiosyncratic preferences and expectations about their future matching opportunities. I propose a tractable large market approximation yielding an analytical solution to the model which directly maps agents' preferences into sorting and job-to-job transitions. Second, I show that this mapping is invertible such that the preferences of participating agents can be nonparametrically identified from data on realized matches. Third, I take this methodology to panel data on the allocation of public teachers in Peru and show that (i) the spatial disaggregation of labor demand,

(ii) the concentration of labor supply in cities and (iii) the presence of home bias in teachers' preferences, lead to the existence of a *spatial job ladder*. As a result, low quality teachers get displaced in remote locations, creating a wide urban-rural gap in teacher quality, and move toward urban schools once they have accumulated experience and skills, which further magnifies this gap by one third. Finally, I show that dynamic wage contracts can reduce inequalities in access to skilled teachers by incentivizing teacher retention.

I start the analysis by leveraging countrywide panel data on the centralized allocation of public teachers in Peru. I document that remoteness is highly predictive of teacher sorting as high-skilled teachers concentrate in urban schools, while low-skilled teachers mostly work in remote locations. Teachers working in remote locations switch from job-to-job at a high rate to get closer to urban centers. This implies that teacher attrition rates in remote villages are three times greater than in cities. Movers are, on average, of higher quality than those who replace them. Labor market dynamics thus seem to largely reinforce spatial inequalities in teacher quality and student achievement.

To understand what drives local labor demand and supply and how they translate into equilibrium sorting and career paths, I develop an empirical model of dynamic two-sided matching without transfers. Teachers and schools meet repeatedly in a single market over several time periods. The observed characteristics of both sides evolve endogenously according to their matching decisions. Agents are forward-looking and form preferences over observed and unobserved job/teacher attributes. I impose few assumptions on preferences and beliefs: (i) the systematic and unobserved part of the payoff functions are additively separable, (ii) the unobserved taste shocks are *iid* with a type-I upper tail and (iii) agents have rational expectations about their future match payoffs. I extend the concept of stability, widely used in static empirical models of two-sided matching, to this dynamic setting. I assume that the observed match in each period is stable with respect to teachers' and schools' lifetime utility and that beliefs about future aggregate states are consistent with their realizations.

To map preferences into sorting, I build on the static framework of Menzel (2015) and leverage the implications of stability in a large market setting where the number of agents on both sides grows to infinity. Stability implies that, in each period, each teacher is matched to her preferred job among the set of jobs that would be willing to hire her and vice versa. We can thus reinterpret the realized matches as the outcome of two dynamic discrete choice models with unobserved and endogenous choice sets. Under the assumption that shocks have a type-I upper tail, I show that the information contained in choice sets, that is necessary to characterize conditional choice probabilities, can be summarized into sufficient statistics called *inclusive values*. In the limit economy, inclusive values converge to the unique solution of a fixed-point problem, which explicitly models the dependence between preferences and choice sets. This allows us to derive an analytical expression for the equilibrium conditional choice probabilities and map preferences into sorting.

I show that the mapping between preferences and observed sorting is invertible. The joint surplus function can be nonparametrically identified from data on realized matches. Under appropriate exclusion restrictions or with the availability of additional data, preferences can be separately identified from the joint surplus. I provide these results in two settings: (i) finite horizon and nonstationarity of preferences and aggregate states and (ii) infinite horizon and stationarity. I then propose a maximum likelihood estimator that can be tractably used for a parametric version of this framework.

Equipped with this methodology, I identify and estimate teachers' and schools' preferences from data on observed matches within the centralized assignment procedure in Peru. To separately identify preferences from the joint surplus, I use additional data on how schools rank the applicants they interview. The estimated preference parameters indicate that (i) geographical proximity to home is highly predictive of teachers' preferences and (ii) schools highly value observed measures of teacher quality, such as experience. This results in the existence of a spatial job ladder. As labor demand is widely scattered while teachers' home location is concentrated in cities, fact (i) implies that teachers have a strong distaste for remote locations putting rural schools at the bottom of the ladder and urban schools at the top. As the number of jobs located in urban centers is limited, fact (ii) implies that excess supply is rationed based on quality such that high-skilled teachers concentrate in cities while low-skilled teachers are matched to remote schools. The spatial job ladder also has important consequences on labor market dynamics. Teachers accumulate experience and human capital throughout their career and climb up the ladder by matching closer to home. As a consequence, rural schools fail to retain skilled teachers and sustain disproportionately low levels of teaching experience and quality. Overall, I estimate that teacher mobility along the spatial job ladder explains one third of the urban-rural gap in teacher quality.

I then investigate the effectiveness of dynamic wage contracts aimed at slowing down labor mobility and mitigating its adverse effects on spatial inequalities through retention bonuses. To do so, I simulate the equilibrium response to a policy that would impose a minimum contract length in exchange for appropriate compensation to prevent teachers from moving up the ladder. If compensation is too low, this policy creates large shortages as it forces teachers to commit and prevents them from rematching ex-post. This highlights a key trade off between recruitment and retention in the presence of a job ladder. Bonuses that would negate this adverse sorting effect amount to a 20-40% wage increase depending on the contract length.

I conclude the analysis with a thought experiment simulating the equilibrium in a counterfactual scenario where teachers' home locations would be scattered across the country instead of being concentrated in cities. As proximity to home is no longer associated with proximity to cities, the spatial job ladder collapses. Teachers still aim to match close to home but face little competition for these positions. Consequently, high quality teachers are no longer disproportionately matched to urban schools. The rate at which teachers switch jobs drops by half. Job-to-job flows are no longer directed from rural schools toward urban schools which shuts down urban-rural inequalities in attrition. This suggests that designing policies targeting the root causes of the existence of the spatial job ladder, such as investing in training local teachers, might be more effective than aiming at slowing down its symptoms through recruitment or retention policies.

Related literature

This paper relates and contributes to several strands of the literature. First, I contribute to a growing literature at the intersection of industrial organization and econometrics studying the empirical content of two-sided matching models with non-transferable utility (NTU).[1] Several papers investigate, in a static setting, how preferences of participating agents can be identified from reported preferences (Fack et al., 2019; Agarwal and Somaini, 2020) or realized matches (Menzel, 2015; Diamond and Agarwal, 2017; He et al., 2021; Agarwal and Somaini, 2022; Ederer, 2022). Yet, there are few equivalent results for models of dynamic two-sided matching, despite being increasingly studied in the matching theory literature.[2] A handful

[1]Following the seminal work of Choo and Siow (2006), a large literature on empirical models of two-sided matching with transferable utility (TU) has evolved separately (Fox, 2010; Galichon and Salanié, 2022; Gualdani and Sinha, 2019).

[2]See Baccara and Yariv (2021) for a survey of this rapidly growing literature.

of papers study waitlist mechanisms (Agarwal et al., 2021; Waldinger, 2021; Verdier and Reeling, 2022) or include dynamics in college admissions/school choice models (Larroucau and Rios, 2020). However, these papers study priority-based assignment mechanisms where the preferences of one side of the market are known ex-ante. This paper contributes to this literature by building an empirical model of dynamic two-sided matching where the preferences of both sides of the market are unknown. It extends the concept of stability to a dynamic setting to map preferences into sorting and show that preferences can be nonparametrically identified from data on realized matches.

Second, I contribute to a large literature in labor and urban economics studying the causes and welfare consequences of spatial skill sorting (Moretti, 2013; Diamond, 2016; Diamond and Gaubert, 2022). I provide a unifying explanation for the lack of access to local services requiring skilled labor in remote areas. As labor demand is inherently spatially scattered in these markets while human capital concentrates in cities, the presence of home bias generates the existence of a spatial job ladder, which has drastic consequences on spatial sorting and mobility. The tools provided in this paper could help understand the causes and welfare consequences of important phenomenons such as the existence of medical deserts. I also contribute to the literature studying sorting and labor mobility through on-the-job search models (Moscarini and Postel-Vinay, 2018) by showing that labor market dynamics can alternatively be rationalized by a frictionless dynamic two-sided matching model.[3]

Third, I contribute to a recent literature on equilibrium models of the teachers' labor market (Tincani, 2021b; Biasi et al., 2021b; Bates et al., 2022; Bobba et al., 2022). These papers study teacher sorting through static models of two-sided matching. I provide a general framework nesting the existing approaches and derive conditions under which preferences are nonparametrically identified from realized matches. I also show the importance of labor market dynamics in shaping teacher sorting, which is typically ignored in this literature.

Fourth, I relate to a large body of work in the economics of education studying the causes and consequences of teacher attrition (Boyd et al., 2005; Falch and Strøm, 2005; Falch, 2011; Hanushek et al., 2016; Bonhomme et al., 2016). This paper provides a unifying framework to study teacher sorting and mobility. I show that attrition is mostly caused by teachers leaving rural schools by climbing up the spatial job ladder. I then provide new evidence on

[3]This raises the question of whether search frictions and idiosyncractic preferences over job attributes can be separately identified from matched employer-employee data in typical search models. I plan to investigate this in future work.

the costs of attrition by quantifying its role in shaping urban-rural inequalities in access to skilled teachers.

Finally, this paper relates to a literature in public economics studying the design of incentives to recruit and retain civil servants in underprivileged areas. Several papers explored the role of wage incentives on recruitment, effort and retention but found mixed results on retention (Deserranno, 2019b; Leaver et al., 2021b; Bobba et al., 2022). Instead, I explore the effect of dynamic wage contracts designed to increase retention. I show that these policies can have strong adverse effects on recruitment if teachers are not properly compensated for the implied lack of flexibility. This highlights a trade off between recruiting and retaining workers in the presence of a job ladder.

Overview

Section 3.2 briefly describes the institutional setting and the data. Section 3.3 presents relevant descriptive evidence. Section 3.4 introduces the equilibrium model and characterizes the mapping between preferences and sorting. Section 3.5 states the main identification results. Sections 3.6 and 3.7 discuss the empirical strategy and the results. Section 3.8 concludes.

3.2 Context and Data

In this section, I briefly describe the different types of contracts under which teachers can be employed and how the centralized clearinghouse allocating teaching positions is organized. I then give a short summary of the different sources of data used throughout the paper.

3.2.1 Institutional Setting

Public teachers in Peru can be hired under two types of contracts. *Temporary* contracts last at least one year and can be renewed up to a second year. *Permanent* contracts can last indefinitely and are akin to usual civil servant contracts. Temporary contracts are paid a fixed rate that does not vary with experience. Permanent teachers can get promoted throughout their career to higher ranks in the civil servant scale system to get higher wages.[4] On the

[4]Promotions are awarded through a national standardized evaluation and a decentralized evaluation made by a committee evaluating teachers' performance and professional career.

lowest scale, permanent teachers are paid the same wage as temporary teachers. On the highest scale, permanent teachers are paid 75% more. In an effort to make remote schools and schools with difficult teaching conditions more attractive, the Ministry of Education provides wage bonuses to teachers working in schools belonging to a predetermined set of categories (see Appendix 3.B.1 for more details). However, the overall spatial variation in wages induced by this bonus scheme remains very limited.[5]

Since 2015, the allocation of new teaching positions is organized through a biennial centralized clearinghouse. All teachers without a permanent contract seeking a position have to go through this process.[6] The allocation is organized into three steps that take place at the end of the academic year from November to January. First, all applicants participate in the *national competency exam*, which assesses their skills and curricular knowledge. If their score falls above a given threshold, teachers become eligible for permanent contracts. Second, eligible applicants participate in the allocation of permanent positions. Teachers form an unconstrained list of choices within the same province and are then interviewed by their three top schools.[7] Schools then make offers to their preferred candidates. Finally, all remaining teachers participate in the allocation of temporary contracts. In this step, schools are passive and cannot express their preferences. Teachers are ranked according to their test score and choose among the set of available positions by order of priority. Finally, schools which did not manage to recruit anyone can resort to hiring non-certified teachers through temporary contracts. More details about the test and the timing of the allocation mechanism are available in Appendix 3.B.1.

3.2.2 Data

I combine several sets of administrative data provided by the Ministry of Education in Peru to create a unique record of teachers' movements across schools throughout their careers. Most importantly, I observe teachers repeatedly applying through the centralized assignment

[5] Table 3.A.2 shows summary statistics on various job characteristics. One standard deviation in wages corresponds to only 16% of the minimum wage.

[6] Permanent teachers seeking to get transferred to another school need to go through a separate decentralized procedure.

[7] As schools cannot interview more than ten applicants, capacity constraints are rationed using test scores as priorities.

platform, allowing for a deeper investigation of the causes of these movements.[8] I briefly describe these data sources below.[9]

Teacher assignment data: I observe a panel including all positions and teachers employed in the public sector in Peru from 2015 to 2021. For each teacher in each year, I know in which position they work, which type of contract they hold and which wage they receive. I supplement these data with additional sources of information on jobs and teachers (see Appendix 3.B.2 for details on the data construction). First, I link teachers national ID to the Household Targeting System (SISFOH) data containing information about their poverty status, education level and, most importantly, their home location. It also allows me to link each teacher to other members of their household and know their marital status, whether they have children and whether they live with their parents. Second, I link each job to the School Census containing a wide set of locality and school characteristics. I observe whether a given locality has access to basic amenities such as water and electricity. I also have information about the precise geolocalization of the school and the level of poverty and rurality of its locality.

Centralized assignment data: I have access to detailed information about the biennial countrywide centralized assignment of new teaching positions from 2015 to 2019. I observe the universe of participating applicants and positions in each step of the mechanism. The dataset contains information on applicants' test scores at the national competency exam. I also have access to detailed information on the allocation of permanent positions. In particular, I observe the set of applicants each school interviews and how they rank them. The dataset also records the final match for both temporary and permanent contracts in each year. Finally, and key to my analysis, this dataset can also be linked to the teacher assignment data in order to track applicants and positions across years.

Note that the teacher assignment data and centralized assignment data do not necessarily overlap. The centralized assignment data contains information about the set of applicants and vacancies that end up staying unmatched and thus do not appear in the teacher assignment data. The teacher assignment data contains information about applicants already holding a permanent contract and non-certified teachers who are not allowed to participate in the

[8] Bobba et al. (2022) use similar data but do not exploit the panel dimension of the data and abstract away from the role of labor dynamics.

[9] I restrict the analysis to public primary education. Primary schools are evenly distributed across the country while secondary schools are sometimes missing in remote locations. Teachers' spatial sorting is thus a more salient concern for primary education.

centralized allocation mechanism.

3.3 Descriptive Evidence

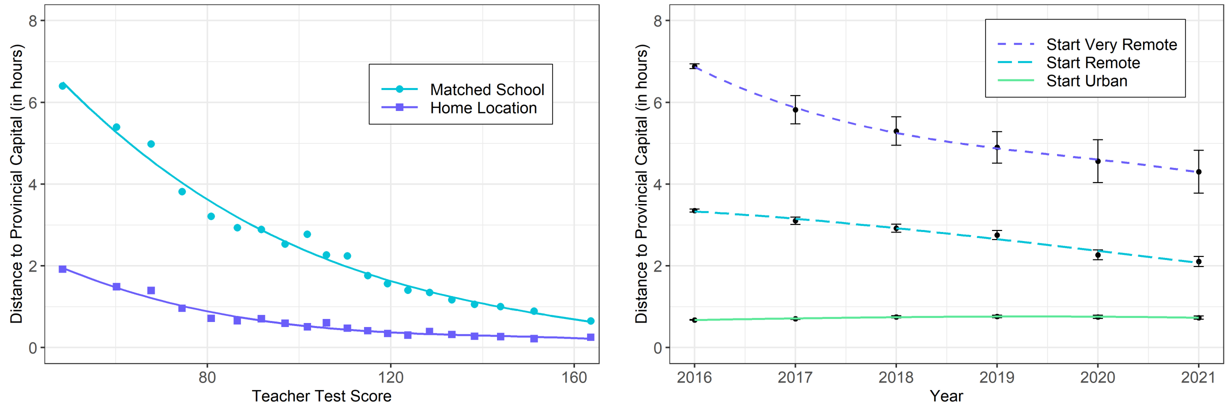
Jobs are geographically scattered across locations which greatly differ in their level of remoteness and amenities (see Table 3.A.2). One quarter of positions are located more than four hours away from the provincial capital. One third of the available positions are located in schools that have no access to electricity or water. In contrast, teachers' home locations are concentrated in cities: 82% of applicants live in a provincial capital (see Table 3.A.3). In this section, I provide suggestive evidence that this creates an imbalance between local supply and demand, which shapes teacher spatial sorting and mobility and translates into spatial inequalities in teaching quality.

3.3.1 Spatial Sorting and Mobility

I first leverage data on the centralized assignment mechanism to document how teachers sort across locations, in the cross-section, based on observed measures of teacher quality. Panel A of Figure 3.1 plots the relationship between teachers' test scores and the distance between their matched school and the provincial capital. I find that high scoring teachers are disproportionately matched to schools located close to urban centers. Specifically, teachers in the top decile of the score distribution work on average 45 minutes away from the provincial capital, while teachers in the bottom decile work 6 hours away. This pattern is not driven by spatial disparities in the quality of local workers as low scoring teachers live close to urban centers, on average.

I then document how teachers move across locations throughout their careers using the panel structure of the data. Among the set of teachers who started a new job in 2016, 40% switched jobs at least twice over the period 2016-2021. This number decreases to 25% for teachers starting in urban areas in 2016 while it increases to 60% for teachers starting in remote locations. Panel B of Figure 3.1 plots the time trend of the remoteness of teachers' matched schools. I find that as teachers switch jobs, they also switch locations and progressively move closer to urban centers. The rate at which they move increases with the remoteness of their starting job. Teachers who start in remote locations get closer to the provincial capital by almost three hours. In contrast, teachers who already start in proximity

Figure 3.1: Sorting and Movements Across Locations



a) Sorting **b) Job-to-Job Transitions**

NOTES. This figure uses the teacher assignment data. Panel A plots binned averages of the distance (in hours) between applicants' home location and the provincial capital as well as applicants' matched location and the provincial capital. Each bin is equally spaced using vigintiles of the distribution of teachers' test scores. Panel B plots the evolution of the distance between teachers' matched schools and the provincial capital over the period 2016-2021 for three groups of teachers starting at different levels of remoteness in 2016.

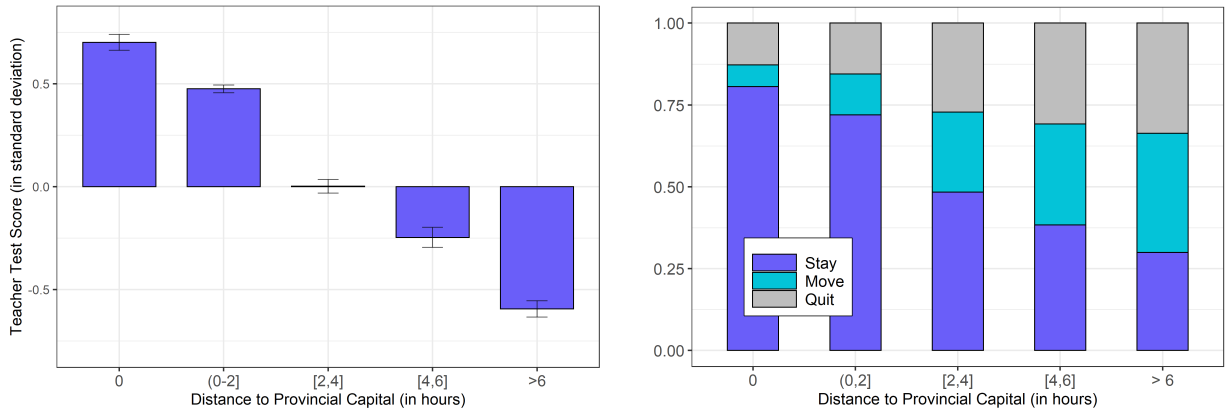
to urban centers do not get closer by switching jobs.

These patterns suggest that teachers have a distaste for remoteness, potentially creating an imbalance between local labor supply and demand. Excess supply in urban locations seems to be rationed through observed measures of teacher quality such as test scores. As a result, low-quality teachers work temporarily far from urban centers and switch from job-to-job at a high rate to move closer to cities.

3.3.2 Spatial Inequalities

Teacher spatial sorting and movements across locations have direct consequences on the distribution of teaching quality across space. The sorting patterns described in Figure 3.1 directly imply that teachers working in remote schools are less qualified than teachers working in cities. Panel A of Figure 3.2 shows the resulting urban-rural gap in teacher test scores. Teachers working in the provincial capital score on average 1.3 standard deviations higher than teachers working in very remote schools located more than 6 hours away from the provincial capital. Similarly, the magnitude and direction of the job-to-job flows described in Figure 3.1 imply that schools located in rural areas face high attrition rates. Panel B of Figure 3.2 shows that between 2016 and 2018, the teacher attrition rate in schools located in remote villages is 50 percentage points higher than in schools located in the provincial

Figure 3.2: Spatial Inequalities



a) Teacher Test Score **b) Transition 2016-2018**

Notes. This figure uses the teacher assignment data and documents urban-rural inequalities in teacher test scores and in the type of job transitions between 2016 and 2018. Panel A shows the average test score of matched teachers for several bins of the distance to the provincial capital. Panel B shows the share of teachers that stayed in the same school, moved to another school or quit teaching in the public sector for several bins of the schools' distance to the provincial capital.

capital.

It has been widely documented that teacher attrition negatively affects student learning through disruption and the resulting loss of experience (Hanushek et al., 2016). I provide descriptive evidence in line with these results. I compare movers with the teachers who replaced them in 2018 over several dimensions. Table 3.1 shows that movers are significantly more experienced than newcomers. Eleven percent of newcomers have no prior experience. Newcomers are 6 percentage points more likely to be non-certified. I also find that movers score on average 0.16 standard deviations higher at the national exam compared to newcomers. This is quite substantial as this corresponds to 12% of the urban-rural gap in teacher test scores.

As the literature points out that observable measures of teacher quality can be poor predictors of teacher value added (Rockoff, 2004), I also provide additional evidence in Appendix 3.C that movers are of significantly higher value added than newcomers. To do so, I follow Chetty et al. (2014a) and estimate teacher value added using matched teacher-classroom data. I find that movers' value added is 0.10 standard deviations higher than newcomers. This corresponds to 50% of a standard deviation in value added which is quite substantial. This result is consistent with evidence of large value added gains through experience in the early stages of teachers' careers (Rockoff, 2004; Rivkin et al., 2005; Araujo et al., 2016).

Overall, these findings suggest that teacher sorting and mobility have important conse-

Table 3.1: Movers vs. Newcomers

	Movers	Newcomers	Difference
Competency Score	0.545	0.389	0.156 (0.017)
Non-certified	0.125	0.181	0.056 (0.006)
Value Added	0.057	-0.042	0.099 (0.031)
Experience			
No Experience	0	0.109	-0.109 (0.004)
Between 1 and 2 years	0.166	0.190	-0.024 (0.006)
Between 3 and 5 years	0.308	0.250	0.058 (0.007)
Between 6 and 10 years	0.271	0.187	0.084 (0.007)
Above 10 years	0.123	0.083	0.040 (0.005)

NOTES. This table uses the centralized assignment data to compare the temporary teachers that moved to a different school between 2016 and 2018 to the teachers that were hired to replace them in 2018 over several dimensions. Details on how value added is estimated are in Appendix 3.C.

quences on spatial inequalities. Schools located in remote areas fail to attract high-quality teachers and face high attrition rates. As movers are replaced by teachers of lower experience and quality, labor market dynamics sustain and exacerbate spatial inequalities in teaching quality and student achievement.

The suggestive evidence presented in this section highlights the need for further investigation on the causes of teacher spatial sorting and mobility. More specifically, it is crucial to understand (i) how teachers trade off geographical proximity against other job/locality characteristics and (ii) how schools ration excess labor supply. To do so, I develop next a general model of dynamic two-sided matching mapping teachers' and schools' preferences into equilibrium sorting and job-to-job flows.[10]

3.4 Empirical Model of Dynamic Two-Sided Matching

In this section, I build on Menzel (2015) and develop a general model of dynamic two-sided matching with non-transferable utility incorporating the following features. First, an empirical model of teachers' and schools' preferences able to quantify how agents trade off a potentially large set of job and teacher attributes. Second, state variables that evolve over

[10] As reallocation entails costly migration decisions, embedding these decisions within a dynamic framework is crucial to disentangle moving costs from taste for specific locality characteristics such as amenities or remoteness (Kennan and Walker, 2011).

time depending on agents' matching decisions. Third, forward-looking agents that anticipate the effect of their current action on the future. Finally, an equilibrium concept mapping these elements into sorting and job-to-job flows.

This section is divided into two parts. I first describe the environment, the preference model and introduce the equilibrium concept. Then, I characterize the mapping between preferences and realized sorting.

3.4.1 Model

Throughout this section, I refer to one side of the market as teachers and the other side as schools. I assume that matching is one-to-one meaning that each school only opens one vacancy. Alternatively, we can consider jobs as separate entities such that matching is one-to-one by design. To simplify the analysis, I use a large market approximation to obtain a tractable analytical expression linking primitives to equilibrium sorting. I start by introducing the relevant parts of the model in the finite economy before defining the asymptotic sequence that characterizes the limit economy.

Timing

I consider a repeated matching game where a set of schools and teachers meet in a single market in each period. An extension considering the opposite polar case where matches are irreversible is in Appendix 3.F. Time is discrete and indexed by $t = 1, ..., T$. I assume that $T \in [1, \infty]$ meaning that the model nests both the static case $T = 1$, which corresponds to Menzel (2015), and the infinite horizon case $T = \infty$. For simplicity, I assume that the set of participating agents and schools is fixed over time. However, this framework can be extended to settings where agents enter and exit the market sequentially in an exogenous way. Teachers are indexed by $i \in \mathcal{I} = \{1, ..., n_w\}$ and schools are indexed by $j \in \mathcal{J} = \{1, ..., n_m\}$. In each period t, a matching is formed summarized by the functions μ_{wt}, which maps $\mathcal{I}$ to $\mathcal{J} \cup \{0\}$ and μ_{mt}, which maps $\mathcal{J}$ to $\mathcal{I} \cup \{0\}$ where 0 is the option of staying unmatched. The resulting matching is summarized in $\boldsymbol{\mu} = (\mu_{wt}, \mu_{mt})_{t=1}^{T}$.

Teacher i and school j are characterized in each period t by a set of observed characteristics which are collected into two vectors $\boldsymbol{x}_{it}$ and $\boldsymbol{z}_{jt}$. I fix the probability distribution functions of their initial value $\boldsymbol{x}_{i1}$ and $\boldsymbol{z}_{j1}$ as $w_1(\boldsymbol{x})$ and $m_1(\boldsymbol{z})$ with support $\mathcal{X}_1$ and $\mathcal{Z}_1$ and assume that they are exogenous. Individual states evolve stochastically over time depending on agents' matching decisions $\boldsymbol{\mu}$ through the Markov transition probability distribution functions $w_{t+1}(\boldsymbol{x}_{it+1}|\boldsymbol{x}_{it}, \boldsymbol{z}_{\mu_{wt}(i)t})$ and $m_{t+1}(\boldsymbol{z}_{jt+1}|\boldsymbol{x}_{\mu_{mt}(j)t}, \boldsymbol{z}_{jt})$. I denote separately $m_{0t+1}(\boldsymbol{z}_{jt+1}|\boldsymbol{z}_{jt})$ and $w_{0t+1}(\boldsymbol{x}_{it+1}|\boldsymbol{x}_{it})$ the transition probability distribution functions for agents choosing to stay unmatched. Throughout the rest of the paper, I drop the index t from the functions w, w_0, m and m_0 for simplicity. Finally, individual matching decisions in period t aggregate into the probability distribution functions of observed states w_{t+1} and m_{t+1} as follows:

$$w_{t+1}(\boldsymbol{x}, \boldsymbol{\mu}) = \int_{\mathcal{X}_t} \int_{\mathcal{Z}_t} w(\boldsymbol{x}|s,h) f_t(s,h) dh ds + \int_{\mathcal{X}_t} w_0(\boldsymbol{x}|s) f_t(s,*) ds$$

$$m_{t+1}(\boldsymbol{z}, \boldsymbol{\mu}) = \int_{\mathcal{X}_t} \int_{\mathcal{Z}_t} m(\boldsymbol{z}|s,h) f_t(s,h) dh ds + \int_{\mathcal{Z}_t} m_0(\boldsymbol{z}|h) f_t(*,h) dh$$

where $f_t(x,z)$, $f_t(x,*)$ and $f_t(*,z)$ are, respectively, the joint probability distribution function of the characteristics of matched teachers and schools, of unmatched teachers and of unmatched schools in period t. A formal definition of these functions is in the next subsection.

Preferences and Beliefs

Agents are forward looking and anticipate how their current decision affects their lifetime utility. I define the lifetime utility that teacher i gets from being matched with school j in period t as:

$$U_{ijt} = U_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{jt}) + \sigma\eta_{ijt} + \beta_w \int \overline{U}_{it+1}(\boldsymbol{x}_{it+1}) w(\boldsymbol{x}_{it+1}|\boldsymbol{x}_{it}, \boldsymbol{z}_{jt}) d\boldsymbol{x}_{it+1}$$

whereas the lifetime utility that school j gets from being matched with teachers i in period t is defined as:

$$V_{ijt} = V_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{jt}) + \sigma\epsilon_{ijt} + \beta_m \int \overline{V}_{jt+1}(\boldsymbol{z}_{jt+1}) m(\boldsymbol{z}_{jt+1}|\boldsymbol{x}_{it}, \boldsymbol{z}_{jt}) d\boldsymbol{z}_{jt+1}$$

Agents' lifetime utility is first composed of a *flow utility*, which agents enjoy from their match in period t. It includes a systematic part $U_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{jt})$ and $V_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{jt})$, where the functions (U_t, V_t) are unknown, and unobserved shocks $(\eta_{ijt}, \epsilon_{ijt})$ which are assumed to enter additively. σ is a normalizing sequence which is defined later. I impose the following assumptions on these objects.

Assumption 1 *(i) U_t and V_t are uniformly bounded in absolute value and $p \geq 1$ times differentiable with uniformly bounded partial derivatives in $\mathcal{X} \times \mathcal{Z}$ for all t.*
(ii) ϵ_{ijt} and η_{ijt} are drawn independently from $\boldsymbol{x}_{it}$ and $\boldsymbol{z}_{jt}$ from a distribution with absolutely continuous c.d.f. $G(s)$ and density $g(s)$. The upper tail of the distribution $G(s)$ is of type I with auxiliary function $a(s) = \frac{1-G(s)}{g(s)}$.

Assumption 1.(i) is a standard regularity condition which ensures that the functions U_t and V_t are well-behaved. Assumption 1.(ii) imposes restrictions on the upper tail of the distribution of ϵ_{ijt} and η_{ijt} but leaves the lower tail unrestricted. As the number of teachers and schools grows to infinity, the number of independent draws of ϵ_{ijt} and η_{ijt} also grows. All draws of ϵ_{ijt} and η_{ijt} from the lower tail of their distribution thus become inconsequential in determining which is the most preferred school or teacher. As in Menzel (2015), I assume that G belongs to a class of distributions which has a type I extreme value distributed upper tail.[11] Note that this class of functions encompasses most of the parametric distributions traditionally used in discrete choice models. For the Gamma distribution or the Gumbel distribution, Assumption 1.(ii) holds for $a(s) = 1$. For the standard normal distribution, it holds for $a(s) = \frac{1}{s}$.

Agents' lifetime utility is then composed of a *continuation value*. Teachers and schools internalize that their matching decisions affect their future states and thus their future payoffs. This continuation value is the discounted sum of future expected payoffs. I assume that teachers discount future utility at a rate β_w, while schools discount at a rate β_m. I define $\overline{U}_{it+1}$ and $\overline{V}_{jt+1}$ as agents' expectations about $U_{i\mu_{t+1}(i),t+1}$ and $V_{\mu_{t+1}(j)j,t+1}$ conditional on their future state variables. As agents only observe their current states, I integrate this object over the transition distribution functions m and w. I impose the following assumptions on $\overline{U}_{it+1}$ and $\overline{V}_{jt+1}$.

Assumption 2 *For each period t, each teacher $i = 1, ..., n_w$ and each school $j = 1, ..., n_m$:*

$$\overline{U}_{it+1}(x) = \mathbb{E}_{\mathcal{S}_t}[U_{i\mu_{t+1}(i),t+1} | x_{i,t+1} = x] \quad \textit{and} \quad \overline{V}_{jt+1}(z) = \mathbb{E}_{\mathcal{S}_t}[V_{\mu_{t+1}(j)j,t+1} | z_{j,t+1} = z]$$

[11] This class of distribution is also called the domain of attraction of the Gumbel distribution (Resnick (1987))

where $\mathcal{S}_t$ is the information set of participating agents in period t:

$$\mathcal{S}_t = \{(\tilde{m}_s)_{s=t}^T, (\tilde{w}_s)_{s=t}^T, G, (U_s)_{s=t}^T, (V_s)_{s=t}^T\}$$

Assumption 2 states that agents have rational expectations about the lifetime utility they will get from their future match conditional on their future state. Agents have incomplete information about the exact realization of the future observed and unobserved states of other participants. Instead, I assume that they know the distribution of taste shocks and the payoff functions for all subsequent periods. I also assume that they form beliefs $(\tilde{m}_s)_{s=t}^T, (\tilde{w}_s)_{s=t}^T$ about the probability distribution functions of future aggregate states $(m_s(\boldsymbol{\mu}))_{s=t}^T, (w_s(\boldsymbol{\mu}))_{s=t}^T$. I assume that individual agents are atomistic and internalize that their decisions only influence their own future state and not the future aggregate states.

Normalizations

For the limit economy to predict sorting patterns that are consistent with the finite economy, I make a few technical assumptions. First, I specify the utility of the outside option as follows:

$$U_{i0t} = \sigma \max_{k=1,\dots,J} \eta_{i0,k} + \beta_w \int \overline{U}_{it+1}(\boldsymbol{x}_{it+1}) w_0(\boldsymbol{x}_{it+1}|\boldsymbol{x}_{it}) d\boldsymbol{x}_{it+1}$$

$$V_{0jt} = \sigma \max_{k=1,\dots,J} \epsilon_{0j,k} + \beta_m \int \overline{V}_{jt+1}(\boldsymbol{z}_{jt+1}) m_0(\boldsymbol{z}_{jt+1}|\boldsymbol{z}_{jt}) d\boldsymbol{z}_{jt+1}$$

I then assume that the size of the market is denoted by n and impose the following normalizations on the asymptotic sequence:

Assumption 3 *The asymptotic sequence is controlled by $n = 1, 2, ...$ and we define:*
(i) $n_w = [\exp(\gamma_w)n]$, $n_m = [\exp(\gamma_m)n]$
(ii) $J = [n^{1/2}]$
(iii) $\sigma = \frac{1}{a(b_n)}$ *where* $b_n = G^{-1}(1 - n^{-1/2})$

Assumption 3.(i) allows to flexibly control the relative sizes of each side of the market through the parameters γ_w and γ_m. Assumption 3.(ii) guarantees that, in each period t, the probability that teachers or schools stay unmatched does not degenerate to zero in the limit. If the size of the outside option does not grow with the size of the market, the probability that it becomes dominated by an alternative option will tend to one given that taste shocks have unbounded support. Assumption 3.(iii) controls the scale of the unobserved shocks such that both the unobserved and systematic parts of the payoffs jointly determine agents' choices in the limit. Given that U_t and V_t are bounded and that the support of taste shocks is unbounded, U_t and V_t would become irrelevant in the limit without this restriction. More specifically, if G is Gumbel, then $b_n \asymp \frac{1}{2}\log(n)$ and $\sigma_n = 1$. If taste shocks are standard normal, $b_n \asymp \sqrt{\log n}$ and $\sigma_n \asymp b_n$ and for Gamma distributed taste shocks, $b_n \asymp \log(n)$ and $\sigma_n = 1$.

Equilibrium

To rationalize the observed matching and link it to the primitives of the model, I impose the following equilibrium assumptions.

Assumption 4 *The match $\boldsymbol{\mu}$ is such that, for all $i = 1, ..., n_w$ and $j = 1, ..., n_m$ in each period t:*
(i) Individually rational in period t: $U_{i\mu_{wt}(i)t} \geq U_{i0t}$ and $V_{\mu_{mt}(j)jt} \geq V_{0jt}$.
(ii) No blocking pairs in period t: There exists no pair (i, j) such that $U_{ijt} > U_{i\mu_{wt}(i)t}$ and $V_{ijt} > V_{\mu_{mt}(j)jt}$.
(iii) Consistent beliefs about aggregate states:

$$\tilde{w}_{t+1}(\boldsymbol{x}) = w_{t+1}(\boldsymbol{x}, \boldsymbol{\mu}) = \int_{\mathcal{X}_t} \int_{\mathcal{Z}_t} w(\boldsymbol{x}|s, h) f_t(s, h) dh ds + \int_{\mathcal{X}_t} w_0(\boldsymbol{x}|s) f_t(s, *) ds$$

$$\tilde{m}_{t+1}(\boldsymbol{z}) = m_{t+1}(\boldsymbol{z}, \boldsymbol{\mu}) = \int_{\mathcal{X}_t} \int_{\mathcal{Z}_t} m(\boldsymbol{z}|s, h) f_t(s, h) dh ds + \int_{\mathcal{Z}_t} m_0(\boldsymbol{z}|h) f_t(*, h) dh$$

Assumption 4 (i) and (ii) impose that the outcome of the match is stable in each period t given agents' lifetime utility. This means that there should exist no teacher-school pair that would prefer to break their current match to rematch together instead. Note that I impose no restriction on preferences such that within a single period there could exist many different stable outcomes (Roth and Sotomayor, 1992). I define the teacher-optimal stable match in period t as μ_t^W and the firm-optimal stable match in period t as μ_t^M. Assumption 4 (iii) imposes that agents' beliefs about the distribution of future aggregate states are consistent with the actual realized equilibrium distributions.

3.4.2 Linking Primitives to Equilibrium Sorting

Equilibrium sorting and job-to-job transitions are summarized by the joint distributions of matched characteristics in each period t. I define this distribution for a given random matching μ_t from a finite economy indexed by n as follows:

$$F_{nt}(x_{it}, z_{jt}|\mu_t) = \frac{1}{n}\sum_{i=0}^{n_w}\sum_{j=0}^{n_m}\mathbb{P}(x_{it} \leq x, z_{jt} \leq z, \mu_{wt}(i) = j)$$

I then denote F_t the limit of the distribution function F_{nt} as the size of the market n grows to infinity. I also define the joint density of matched characteristics as f_t. The goal of this section is to express f_t as a function of the primitives of the model.

The proof is divided in four steps. First, I show that stability implies that the realized matches in each period can be interpreted as the outcome of two dynamic discrete choice models with endogenous and unobserved choice sets called *opportunity sets*. Second, I consider a simplified economy with observed and exogenous choice sets and derive the limit of conditional choice probabilities. Third, I show that the information contained in opportunity sets which is necessary to characterize conditional choice probabilities can be summarized into sufficient statistics called *inclusive values*. Finally, I show that, in the limit, these inclusive values converge to the unique solution of a fixed point problem. This allows to characterize conditional choice probabilities and, in turn, f_t as a function of agents' payoff functions.

Opportunity Sets

Given an arbitrary match μ, I define the opportunity set of a teacher in period t as the set of schools that would be willing to hire her instead of its currently matched employee in the same period. Similarly, the opportunity set of a school is the set of teachers that would be willing to quit their current employer to work there. Formally, I define the opportunity set faced by a given teacher $i \in \mathcal{I}$ in period t given a match μ as:

$$M_{it}(\mu) = \{j \in \mathcal{J} : V_{ijt} \geq V_{\mu_{mt}(j)jt}\}$$

Similarly, I define the opportunity set of school $j \in \mathcal{J}$ as:

$$W_{jt}(\mu) = \{i \in \mathcal{I} : U_{ijt} \geq U_{i\mu_{mt}(i)t}\}$$

I state the first important result:

Proposition 1 *Consider a match μ^* satisfying Assumption 4, for all $i = 1, ..., n_w$ and $j = 1, ..., n_m$:*
(i) For all $t = 1, ..., T$:

$$U_{i\mu^*_{wt}(i)t} = \max_{k \in M_{it}(\mu^*) \cup \{0\}} U_{ikt} \quad \textit{and} \quad V_{\mu^*_{mt}(j)jt} = \max_{l \in W_{jt}(\mu^*) \cup \{0\}} V_{ljt}$$

(ii) Under Assumption 2, for all $t < T$:

$$\overline{U}_{it+1}(x) = \mathbb{E}_{\mathcal{S}_t}\left[\max_{k \in M_{it+1}(\mu^*) \cup \{0\}} U_{ikt+1} | x_{it+1} = x\right]$$

$$\overline{V}_{jt+1}(z) = \mathbb{E}_{\mathcal{S}_t}\left[\max_{l \in W_{jt+1}(\mu^*) \cup \{0\}} V_{ljt+1} | z_{jt+1} = z\right]$$

See Appendix 3.D.1 for a proof of this result. Proposition 1.(i) states that a match μ^*_t is stable if and only if each teacher $i \in \mathcal{I}$ is matched to her preferred school among her opportunity set and each school $j \in \mathcal{J}$ is matched to its preferred teacher among its opportunity set. Proposition 1.(ii) thus follows immediately from (i). This result implies that an equilibrium match μ^* can be rewritten as the outcome of two dynamic discrete choice models where each agent's choice set is its opportunity set. The characterization of optimal choices within dynamic discrete choice models has been extensively studied and used in a variety of settings. However, existing results cannot be transposed to this problem as opportunity sets (i) depend on agents' preferences and are thus unobserved and (ii) depend on the overall equilibrium match and are thus potentially endogenous. The rest of the proof shows that these two issues can be circumvented thanks to a large market approximation.

Conditional Choice Probabilities

To simplify the analysis, I start by characterizing the limit of conditional choice probabilities (CCPs) and expected future payoffs under arbitrary exogenous choice sets and by fixing the aggregate states distributions. I assume that $M_{it} = \{1, ..., J\}$ and $W_{jt} = \{1, ..., J\}$ for all t and I fix m_t and w_t for all t.

Proposition 2 *Consider a given teacher $i \in \mathcal{I}$. Under Assumption 1-3 we have:*

(i) For all t, as $J \to \infty$:

$$J\mathbb{P}(U_{ijt} \geq U_{ikt}, k = \{0,1,...,J\}|\boldsymbol{x}_{it}, \boldsymbol{z}_{jt}) \;\longrightarrow$$
$$\frac{\exp\left\{U_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{jt}) + \beta_w \int \overline{U}_{t+1}(s) w(s|\boldsymbol{x}_{it}, \boldsymbol{z}_{jt}) ds\right\}}{\exp\left\{\beta_w \int \overline{U}_{t+1}(s) w_0(s|\boldsymbol{x}_{it}) ds\right\} + \int \exp\left\{U_t(\boldsymbol{x}_{it}, h) + \beta_w \int \overline{U}_{t+1}(s) w(s|\boldsymbol{x}_{it}, h) ds\right\} m_t(h) dh}$$

$$\mathbb{P}(U_{i0t} \geq U_{ikt}, k = \{0,1,...,J\}|\boldsymbol{x}_{it}) \;\longrightarrow$$
$$\frac{\exp\left\{\beta_w \int \overline{U}_{t+1}(s) w_0(s|\boldsymbol{x}_{it}) ds\right\}}{\exp\left\{\beta_w \int \overline{U}_{t+1}(s) w_0(s|\boldsymbol{x}_{it}) ds\right\} + \int \exp\left\{U_t(x_{it}, h) + \beta_w \int \overline{U}_{t+1}(s) w(s|\boldsymbol{x}_{it}, h) ds\right\} m_t(h) dh}$$

(ii) For all t:

$$\begin{aligned}\overline{U}_{t+1}(x) = \log\Bigg(&\exp\left\{\beta_w \int \overline{U}_{t+2}(s) w_0(s|x) ds\right\} \\ &+ \int \exp\left\{U_{t+1}(x,h) + \beta_w \int \overline{U}_{t+2}(s) w(s|x,h)\right\} m_{t+1}(h) dh\Bigg) + \log(J) + \gamma + o(1)\end{aligned}$$

where $\gamma \approx 0.5772$ is Euler's constant. See Appendix 3.D.2 for a proof of this result. The same result holds symmetrically for the school side. Proposition 2 shows that, under the assumption that unobservables have a type-I upper tail, CCPs converge to the usual Logit formula when the number of alternatives grow to infinity. Similarly, expectations about future payoffs can be computed using the logsum formula commonly used in dynamic discrete choice models with type-I errors.[12]

Note that the conditional choice probability of choosing a particular alternative j would converge to zero if we do not weight it by J, the rate at which the total number of alternatives increases. Lemma 1 in Appendix 3.D.4 establishes that the size of opportunity sets increases at a rate $\sqrt{n}$ which justifies Assumption 3.(ii).

[12]These CCPs exhibit the independence of irrelevant alternatives (IIA) property which limits the model's ability to allow for flexible substitution patterns. Introducing unobserved discrete types or random coefficients to relax this assumption is possible.

Inclusive Values

I now introduce that opportunity sets are unobserved and endogenous and show that the implications of Proposition 2 allow us to tackle both of these issues.

Endogeneity arises as shifting teacher i's unobserved preferences in a given period t could affect her own opportunity set by triggering a chain of rematches. As in Menzel (2015), I find that, as the size of the market increases, the probability for such an event to occur vanishes to zero. This result stems from two implications of Proposition 2: (i) the probability that school j rematches with a specific teacher i vanishes to zero as the size of opportunity sets increases to infinity and (ii) the probability of choosing the outside option instead, which would terminate such a chain of rematches, is nondegenerate in the limit. This implies that the dependence between taste shocks and opportunity sets vanishes in the limit. Note that this claim can only be proven for the opportunity sets derived from the school-optimal and teacher-optimal stable matchings μ_t^M and μ_t^W. The distribution of taste shocks conditional on opportunity sets is only well defined for the extremal matchings, given that they are the only stable matchings that always exist irrespective of the size of the market. This result is formalized in Lemmas 2 and 3 in Appendix 3.D.4.

I now consider a sequence of school-optimal stable matches μ^M. As opportunity sets' endogeneity vanishes in the limit for extremal matchings, we can then use Proposition 2 (i) to bound[13] teachers' CCPs in period t, assuming that we would observe the corresponding opportunity set $M_{it}(\mu_t^M)$ and future expected payoff function $\overline{U}_{it+1}^M$:

$$n^{1/2}\mathbb{P}(U_{ijt} \geq \max_{k \in M_{it}(\mu_t^M) \cup \{0\}} U_{ikt} | x_{it}, z_{jt}, (z_{kt})_{k \in M_{it}(\mu_t^M)}, M_{i\tau}(\mu_t^M), \overline{U}_{it+1}^M) \leq \tag{3.1}$$
$$\frac{\exp\left\{U_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{jt}) + \beta_w \int \overline{U}_{it+1}^M(s) w(s|\boldsymbol{x}_{it}, \boldsymbol{z}_{jt}) ds\right\}}{\exp\left\{\beta_w \int \overline{U}_{it+1}^M(s) w_0(s|\boldsymbol{x}_{it}) ds\right\} + n^{-1/2} \sum_{k \in M_{it}(\mu_t^M)} \exp\left\{U_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) + \beta_w \int \overline{U}_{it+1}^M(s) w(s|\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) ds\right\}} + o(1)$$

Similar bounds can be computed for a sequence of teacher-optimal stable match μ^W where the direction of the inequality is reversed. The same result also holds for the school side with the direction of the inequality reversed. Using Proposition 2 (ii), we can also bound agents' expectations about their match payoff under a sequence of school-optimal stable matches μ^M

[13]Note that I only provide bounds given that there are several potential stable matches μ_t^* such that $M_{it}(\mu_t^*) = M_{it}(\mu_t^M)$ and $W_{jt}(\mu_t^*) = W_{jt}(\mu_t^M)$.

as follows:

$$\overline{U}_{it}^{M}(x) \geq \log\Bigg(\exp\left\{\beta\int\overline{U}_{it+1}^{M}(s)w_0(s|x)ds\right\} + n^{-1/2}\sum_{k\in M_{it}(\mu_t^M)}\exp\left\{U_t(x,\boldsymbol{z}_{kt})+\beta\int\overline{U}_{it+1}^{M}(s)w(s|x,\boldsymbol{z}_{kt})ds\right\}\Bigg)+\frac{1}{2}\log(n)+\gamma+o(1) \tag{3.2}$$

where again similar bounds can be computed for the teacher-optimal stable match and for the school side with the direction of the inequality reversed.

In Equations (3.1) and (3.2), $n^{-1/2}\sum_{k\in M_{it}(\mu_t^M)}\exp\left\{U_t(\boldsymbol{x}_{it},\boldsymbol{z}_{kt})+\beta_w\int\overline{U}_{it+1}^{M}(s)w(s|\boldsymbol{x}_{it},\boldsymbol{z}_{kt})ds\right\}$ serves as a sufficient statistic that collapses all the information contained in opportunity sets which is needed to approximate CCPs and expectations about future payoffs. These objects are called *inclusive values*. More generally, I define teacher i's inclusive value given a sequence of realized matches μ^* as:

$$I_{wit}^{*} = n^{-1/2}\sum_{k\in M_{it}(\mu_t^*)}\exp\left\{U_t(\boldsymbol{x}_{it},\boldsymbol{z}_{kt})+\beta_w\int\overline{U}_{it+1}^{*}(s)w(s|\boldsymbol{x}_{it},\boldsymbol{z}_{kt})ds\right\}$$

Similarly, I define school j's inclusive value given μ^* as:

$$I_{mjt}^{*} = n^{-1/2}\sum_{l\in W_{jt}(\mu_t^*)}\exp\left\{V_t(\boldsymbol{x}_{lt},\boldsymbol{z}_{jt})+\beta\int\overline{V}_{jt+1}^{*}(s)m(s|\boldsymbol{x}_{lt},\boldsymbol{z}_{jt})ds\right\}$$

I also define I_{wit}^{M} and I_{mjt}^{M} as the inclusive values that would arise under a sequence of school-optimal stable matches μ^M in period t and I_{wit}^{W} and I_{mjt}^{W} as the inclusive values that would arise under a sequence of teacher-optimal stable matches μ^W in period t.

Fixed point characterization

Inclusive values are unobserved as we do not observe opportunity sets, we do not know which stable match is selected and we do not know agents' expectations about future match payoffs. The rest of the proof shows that the inclusive values arising from an equilibrium match μ^* can be approximated by the solution of a fixed point problem.

I first show that, as in the static case (Menzel (2015)), inclusive values arising from a sequence of school-optimal and teacher-optimal stable matches in a given period t can be approximated by *expected inclusive value functions.* I rewrite I^M_{wit} as:

$$
\begin{aligned}
I^M_{wit} &= \frac{1}{n}\sum_{k=1}^{n_m} \exp\left\{U(\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) + \beta_w \int \overline{U}^M_{it+1}(s) w(s|\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) ds\right\} \times \sqrt{n}\mathbb{1}\{k \in M_{it}(\mu^M_t)\} \\
&= \frac{1}{n}\sum_{k=1}^{n_m} \exp\left\{U(\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) + \beta_w \int \overline{U}^M_{it+1}(s) w(s|\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) ds\right\} \sqrt{n}\mathbb{1}\{V_{ikt} \geq \max_{l \in W_{kt}(\mu^M_t)\cup\{0\}} V_{lkt}\}
\end{aligned}
$$

The inclusive value of a given teacher is determined by the set of schools that would accept her, which in turn depends on the preferences of all schools as well as their opportunity sets. Using the school analogous of Equation (3.1), I thus show that:

$$
I^M_{wit} \geq \hat{\Gamma}^M_{wt}(x_{it}) + o_p(1) \quad \text{and} \quad I^M_{mjt} \leq \hat{\Gamma}^M_{mt}(z_{jt}) + o_p(1)
$$

where $\hat{\Gamma}^M_{wt}$ and $\hat{\Gamma}^M_{mt}$ are the school-optimal expected inclusive value function of teachers and schools in period t which are defined as:

$$
\hat{\Gamma}^M_{wt}(x_{it}) = \frac{1}{n}\sum_{k=1}^{n_m} \frac{\exp\left\{U_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) + V_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) + \beta\int \overline{U}^M_{it+1}(s) w(s|\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) ds + \beta\int \overline{V}^M_{kt+1}(s) m(s|\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) ds\right\}}{\exp\left\{\beta\int \overline{V}^M_{kt+1}(s) m_0(s|\boldsymbol{z}_{kt}) ds\right\} + I^M_{mkt}}
$$

$$
\hat{\Gamma}^M_{mt}(z_{jt}) = \frac{1}{n}\sum_{l=1}^{n_w} \frac{\exp\left\{U_t(\boldsymbol{x}_{lt}, \boldsymbol{z}_{jt}) + V_t(\boldsymbol{x}_{lt}, \boldsymbol{z}_{jt}) + \beta\int \overline{U}^M_{lt+1}(s) w(s|\boldsymbol{x}_{lt}, \boldsymbol{z}_{jt}) ds + \beta\int \overline{V}^M_{jt+1}(s) m(s|\boldsymbol{x}_{lt}, \boldsymbol{z}_{jt}) ds\right\}}{\exp\left\{\beta\int \overline{U}^M_{lt+1}(s) w_0(s|\boldsymbol{x}_{lt}) ds\right\} + I^M_{wlt}}
$$

where I define $\overline{U}^M_{it+1}$ and $\overline{V}^M_{jt+1}$ as follows:

$$
\overline{U}^M_{it+1}(x) = \log\left(\exp\left\{\beta\int \overline{U}^M_{it+2}(s) w_0(s|x) ds\right\} + I^M_{wit+1}\right)
$$

$$
\overline{V}^M_{jt+1}(z) = \log\left(\exp\left\{\beta\int \overline{V}^M_{jt+2}(s) m_0(s|z) ds\right\} + I^M_{mjt+1}\right)
$$

Note that similar bounds can be established for the inclusive values that would arise under the teacher-optimal stable match:

$$
I^W_{wit} \leq \hat{\Gamma}^W_{wt}(x_{it}) + o_p(1) \quad \text{and} \quad I^W_{mjt} \geq \hat{\Gamma}^W_{mt}(z_{jt}) + o_p(1)
$$

A formal exposition and proof of this result can be found in Lemma 4 in Appendix 3.D.4. The inclusive value of a given teacher can be approximated by a function of schools' preferences and inclusive values. Similarly, the inclusive value of a given school can be approximated by a function of teachers' preferences and inclusive values. Hence, the two-sided nature of the problem gives rise naturally to a fixed point problem characterizing these inclusive values. Dynamics add a layer of complexity as expectations about future payoffs depend on future inclusive values. There is thus dependence between inclusive values within and across periods.

The rest of the proof entails characterizing this fixed point problem and showing that inclusive values arising from an equilibrium match μ^* can be approximated by its solution. I define the fixed point mappings as follows:

$$\hat{\Psi}_{wt}[\mathbf{\Gamma}](x) = \frac{1}{n}\sum_{k=1}^{n_m} \frac{\exp\left\{U_t(x, \boldsymbol{z}_{kt}) + V_t(x, \boldsymbol{z}_{kt}) + \beta\int \overline{U}_{t+1}[\mathbf{\Gamma}](s)w(s|x, \boldsymbol{z}_{kt})ds + \beta\int \overline{V}_{t+1}[\mathbf{\Gamma}](s)m(s|x, \boldsymbol{z}_{kt})ds\right\}}{\exp\left\{\beta\int \overline{V}_{t+1}[\mathbf{\Gamma}](s)m_0(s|\boldsymbol{z}_{kt})ds\right\} + \Gamma_{mt}(z_{kt})}$$

$$\hat{\Psi}_{mt}[\mathbf{\Gamma}](z) = \frac{1}{n}\sum_{l=1}^{n_w} \frac{\exp\left\{U_t(\boldsymbol{x}_{lt}, z) + V_t(\boldsymbol{x}_{lt}, z) + \beta\int \overline{U}_{t+1}[\mathbf{\Gamma}](s)w(s|\boldsymbol{x}_{lt}, z)ds + \beta\int \overline{V}_{t+1}[\mathbf{\Gamma}](s)m(s|\boldsymbol{x}_{lt}, z)ds\right\}}{\exp\left\{\beta\int \overline{U}_{t+1}[\mathbf{\Gamma}](s)w_0(s|\boldsymbol{x}_{lt})ds\right\} + \Gamma_{wt}(x_{lt})}$$

$$\overline{U}_{t+1}[\mathbf{\Gamma}](x) = \log\left(\exp\left\{\beta\int \overline{U}_{t+2}[\mathbf{\Gamma}](s)w_0(s|x)ds\right\} + \Gamma_{wt+1}(x)\right)$$

$$\overline{V}_{t+1}[\mathbf{\Gamma}](z) = \log\left(\exp\left\{\beta\int \overline{V}_{t+2}[\mathbf{\Gamma}](s)m_0(s|z)ds\right\} + \Gamma_{mt+1}(z)\right)$$

I then show that for a given equilibrium match μ^*, for any $x \in \mathcal{X}$ and $z \in \mathcal{Z}$ in each period t:

$$\hat{\Gamma}^*_{wt}(x) = \hat{\Psi}_{wt}[\hat{\mathbf{\Gamma}}^*](x) + o_p(1) \quad \text{and} \quad \hat{\Gamma}^*_{mt}(z) = \hat{\Psi}_{mt}[\hat{\mathbf{\Gamma}}^*](z) + o_p(1) \tag{3.3}$$

meaning that inclusive values in period t arising from an equilibrium match μ^* can be approximated by fixed points of the mappings $\hat{\Psi}_{wt}, \hat{\Psi}_{mt}$. To characterize the limit of inclusive values, I then consider the limit version of this fixed point problem:

$$\Gamma_{wt} = \Psi_{wt}[\mathbf{\Gamma}] \quad \text{and} \quad \Gamma_{mt} = \Psi_{mt}[\mathbf{\Gamma}] \quad \forall t \tag{3.4}$$

where Ψ_{wt} and Ψ_{mt} are defined in Appendix 3.D.3. The final step of the proof shows that this population fixed point problem has a unique solution and that the approximate solution of the finite sample fixed point problem converges to it. This is stated in the following result:

Theorem 1 *Under Assumption 1-4:*

(i) The mapping $(\log \boldsymbol{\Gamma_w}, \log \boldsymbol{\Gamma_m}) \mapsto (\log \boldsymbol{\Psi_m}[\boldsymbol{\Gamma}], \log \boldsymbol{\Psi_w}[\boldsymbol{\Gamma}])$ *is a contraction.*

(ii) The fixed point problem described in Equation (3.4) always has a unique solution $\boldsymbol{\Gamma}^*_{\boldsymbol{m}}, \boldsymbol{\Gamma}^*_{\boldsymbol{w}}$.

(iii) For any equilibrium μ^*, $I^*_{wit} \longrightarrow \Gamma^*_{wt}(x_{it})$ *and* $I^*_{mjt} \longrightarrow \Gamma^*_{mt}(z_{jt})$ *for all* i, j *and* t.

The complete proof of this result can be found in Appendix 3.D.4. Theorem 1 has several implications. First, it implies that for any arbitrary equilibrium match μ^*, inclusive values converge to the same limit. Consequently, even if there might exist several matches which satisfy the equilibrium conditions in Assumption 4, all are observationally equivalent in the limit. Second, it implies that we can easily characterize conditional choice probabilities as inclusive value functions can be derived by iterating a contraction mapping.

Main result

From Theorem 1 and Proposition 2, we can fully characterize analytically the equilibrium of the model as a function of teachers' and schools' payoff functions. The limit joint density of matched characteristics f_t can be derived from the limit of conditional choice probabilities and has the following expression:

$$\frac{f_t(x,z)}{w_t(x)m_t(z)} = \frac{\exp\left\{U_t(x,z) + V_t(x,z) + \beta\int \overline{U}^*_{t+1}(s)w(s|x,z)ds + \beta\int \overline{V}^*_{t+1}(s)m(s|x,z)ds + \gamma_w + \gamma_m\right\}}{\left(\exp\left\{\beta\int \overline{U}^*_{t+1}(s)w_0(s|x)ds\right\} + \Gamma^*_{wt}(x)\right)\left(\exp\left\{\beta\int \overline{V}^*_{t+1}(s)m_0(s|z)ds\right\} + \Gamma^*_{mt}(z)\right)}$$

$$\frac{f_t(x,*)}{w_t(x)} = \frac{\exp\left\{\beta\int \overline{U}^*_{t+1}(s)w_0(s|x)ds + \gamma_w\right\}}{\left(\exp\left\{\beta\int \overline{U}^*_{t+1}(s)w_0(s|x)ds\right\} + \Gamma^*_{wt}(x)\right)}$$

$$\frac{f_t(*,z)}{m_t(z)} = \frac{\exp\left\{\beta\int \overline{V}^*_{t+1}(s)m_0(s|z)ds + \gamma_m\right\}}{\left(\exp\left\{\beta\int \overline{V}^*_{t+1}(s)m_0(s|z)ds\right\} + \Gamma^*_{mt}(z)\right)}$$

where $f_t(x,*)$ and $f_t(*,z)$ are, respectively, the density of the characteristics of unmatched teachers and unmatched schools. I define the equilibrium expected future payoff functions $\overline{U}^*_{t+1}$ and $\overline{V}^*_{t+1}$ recursively as:

$$\overline{U}^*_{t+1}(x) = \log\left(\exp\left\{\beta\int \overline{U}^*_{t+2}(s)w_0(s|x)ds\right\} + \Gamma^*_{wt+1}(x)\right)$$

$$\overline{V}^*_{t+1}(z) = \log\left(\exp\left\{\beta\int\overline{V}^*_{t+2}(s)m_0(s|z)ds\right\} + \Gamma^*_{mt+1}(z)\right)$$

and the equilibrium aggregate states distribution w^*_t and m^*_t as:

$$w^*_t(x) = \int_{\mathcal{X}_t}\int_{\mathcal{Z}_t} w(x|s,h)f_{t-1}(s,h)dhds + \int_{\mathcal{X}_t} w_0(x|s)f_{t-1}(s,*)ds$$

$$m^*_t(z) = \int_{\mathcal{X}_t}\int_{\mathcal{Z}_t} m(z|s,h)f_{t-1}(s,h)dhds + \int_{\mathcal{Z}_t} w_0(x|s)f_{t-1}(*,h)dh$$

To simulate the equilibrium in practice, one first needs to solve for inclusive values given the specified payoff functions, the initial aggregate distribution of states m_1 and w_1 and the transition distribution functions. From there, it is then possible to construct U_{ijt} and V_{ijt} given a simulated set of taste shocks ϵ and η. To reach a stable match and simulate the equilibrium in a given period t, any version of the Deferred Acceptance algorithm can be used as they are observationally equivalent. Monte Carlo simulations testing the validity of the convergence results derived in this section can be found in Appendix 3.E.

3.5 Identification and Estimation

The previous section built an equilibrium model of dynamic two-sided matching and provided a tractable way to map preferences into sorting. This section shows that this mapping can be inverted such that one can identify and estimate preferences from observed sorting.

3.5.1 Sampling Process

I assume that the available data is a random sample of a panel of individuals from the population regardless of whether they are schools or teachers. One observation in a given period t is thus composed of this individual alone, in the case where it is unmatched, or along with its matched partner otherwise. The probability that a matched individual is selected by this sampling process is thus twice the probability that an unmatched individual is selected. The joint density function of matched characteristics h_t arising from this sampling process relates to f_t in the following way:

$$h_t(x,z) = \frac{2f_t(x,z)}{\exp\{\gamma_{wt}\} + \exp\{\gamma_{mt}\}}$$

where $h_t(x,z)$ is the mass of schools with observed characteristics z matched with teachers with observed characteristics x in period t arising from the sampling scheme defined above and $\exp\{\gamma_w\} + \exp\{\gamma_m\}$ is the total mass of teachers and schools available in this economy. Similarly, I define:

$$h_t(x,*) = \frac{f_t(x,*)}{\exp\{\gamma_w\} + \exp\{\gamma_m\}}$$

$$h_t(*,z) = \frac{f_t(*,z)}{\exp\{\gamma_w\} + \exp\{\gamma_m\}}$$

where $h_t(*,z)$ is the mass of unmatched schools with observed characteristics z and $h_t(x,*)$ is the mass of unmatched teachers.

I also assume that we observe the aggregate distribution of observed states m_t and w_t as this can be easily recovered from f_t as follows:

$$\int_{\mathcal{Z}_t} f_t(x,z)dz + f_t(x,*) = w_t(x)\exp\{\gamma_w\}$$

$$\int_{\mathcal{X}_t} f_t(x,z)dx + f_t(*,z) = m_t(z)\exp\{\gamma_m\}$$

Finally, I assume that the Markov transition density functions m, m_0, w and w_0 can be directly identified from data on observed state transitions.

3.5.2 Identification

The primitives of the model that we do not observe and wish to identify and estimate from the data are the payoff functions $(U_t)_{t=1}^T$ and $(V_t)_{t=1}^T$ and the discount factors β_w and β_m. We know from the literature on dynamic discrete choice models that intertemporal preferences cannot be identified from observed choices without further assumptions (Magnac and Thesmar (2002)).[14] I thus fix the value of the discount factors from now onward. Similarly, I cannot allow for $T = \infty$ while having a nonstationary setting. I thus consider two polar cases: (i) $T < \infty$ and nonstationarity and (ii) $T = \infty$ and stationarity.

[14]Similarly, the flow utility of one alternative needs to be fixed in each period. This is already done through normalizing the option of staying unmatched.

Finite horizon

Given the recursive structure of the problem, the identification argument in the finite horizon case can be done by backward induction. Starting from the last period T, we can identify the joint surplus as follows:

$$U_T(x,z) + V_T(x,z) = \log\left(\frac{f_T(x,z)}{f_T(x,*)f_T(*,z)}\right)$$

We can also identify Γ^*_{wT} and Γ^*_{mT} from the distribution of unmatched teachers and schools:

$$\Gamma^*_{wT}(x) = \frac{w_T(x)\exp(\gamma_{wT})}{f_T(x,*)} - 1$$

$$\Gamma^*_{mT}(z) = \frac{m_T(z)\exp(\gamma_{mT})}{f_T(*,z)} - 1$$

$\overline{U}_T$ and $\overline{V}_T$ can then be computed by backward induction:

$$\overline{U}_T(x) = \log(1 + \Gamma^*_{wT}(x)) + \gamma$$

$$\overline{V}_T(z) = \log(1 + \Gamma^*_{mT}(z)) + \gamma$$

From there, we can then repeat the same steps to identify the inclusive value functions and the joint surplus in period $T-1$. Finally, we iterate the procedure to identify the joint surplus and the inclusive value functions in all periods t. This results in the following proposition.

Proposition 3 *Under Assumption 1-4 and for $T < \infty$:*
(i) The joint surplus function $U_t + V_t$ and the inclusive value functions Γ^*_{wt} and Γ^*_{mt} are identified for all t from f_t, the limiting joint distribution of matched characteristics in period t.
(ii) Without further restrictions, we cannot separately identify U_t and V_t for all t.

We face a similar identification challenge as in the static case (Menzel, 2015) as preferences are not separately identified from the joint surplus. However, note that this is not necessarily a negative result. Given that the joint distribution of matched characteristics is solely driven by the joint surplus, knowing the joint surplus is enough to perform counterfactuals where we would change the distribution of teachers' and schools' observed attributes. Nevertheless, we might be interested in identifying and estimating preferences as these might be objects of interest. Exclusion restrictions might be useful to disentangle preferences from the joint surplus, as in the static case (Ederer, 2022). In the empirical analysis, I use additional data on how schools rank the applicants they interview to disentangle teachers' and schools' preferences from the joint surplus.

Infinite horizon

To allow for $T = \infty$, I impose the following assumptions.

Assumption 5 *(i) Stationarity of preferences: $U_t = U$ and $V_t = V$ for all t.*
(ii) Stationarity of aggregate states distribution: $m_t = m$ and $w_t = w$ for all t.

Assumption 5 has the direct implication that inclusive value functions are also stationary $\Gamma_{mt} = \Gamma_m$ and $\Gamma_{wt} = \Gamma_w$ for all t. As a consequence, $\overline{U}_t = \overline{U}$ and $\overline{V}_t = \overline{V}$. However, Assumption 5 (ii) is fairly restrictive as it forces aggregate states to remain on a predetermined stationary path which might not be consistent with what the model predicts. Showing existence of a stationary equilibrium which would satisfy consistency requirements is left for future work. Assumption 5 then implies that we can write:

$$\frac{f(x,*)}{w(x)} = \frac{\exp\left\{\beta \int \overline{U}^*(s) w_0(s|x) ds\right\}}{\left(\exp\left\{\beta \int \overline{U}^*(s) w_0(s|x) ds\right\} + \Gamma_w^*(x)\right)}$$
$$= \frac{\exp\left\{\beta \int \overline{U}^*(s) w_0(s|x) ds\right\}}{\exp\left\{\overline{U}^*(x) - \gamma\right\}} = \exp\left\{\beta \int \overline{U}^*(s) w_0(s|x) ds - \overline{U}^*(x) + \gamma\right\}$$

From there, we can invert this mapping to recover $\overline{U}^*$. We can follow the same steps to recover $\overline{V}$ from $f(*, z)$. It is then immediate to see that we can identify $U + V$ from $f(x, z)$.

Proposition 4 *Under Assumption 1-5 and for $T = \infty$:*
(i). The joint surplus function $U + V$ and the inclusive value functions Γ_w^* and Γ_m^* are identified from the limiting joint distribution of matched characteristics in each period f.
(ii). Without further restrictions, we cannot separately identify U and V.

Note that in the stationary case, a single cross section is sufficient to identify and estimate $U + V$ as the joint distribution of matched characteristics does not depend on t anymore. However, this does not mean that dynamics do not play a role as agents still make forward looking decisions.

3.5.3 Estimation

I consider a parametric version of this framework where I define the payoff functions as $U(x, z; \boldsymbol{\theta_t})$ and $V(x, z; \boldsymbol{\theta_t})$ such that U and V are known for all (x, z) up to a vector of unknown parameters $\boldsymbol{\theta_t}$. I assume that we observe a random sample of K individuals over each period t, drawn from the sampling scheme described in Section 5.1, along with their respective matches. For a given observation k in period t, we observe a vector $(x_t(k), z_t(k))$ which is encoded differently depending on the type of match we observe. For an unmatched teacher, indexed by $w_t(k) = 0$, I record its characteristics in $x_t(k)$ and encode $z_t(k)$ as missing. Similarly, for an unmatched school, indexed by $m_t(k) = 0$, I record its characteristics in $z_t(k)$ and encode $x_t(k)$ as missing. For a matched teacher or school, indexed by $m_t(k) = w_t(k) = 1$, I record their characteristics in $(x_t(k), z_t(k))$. We can then construct the following sample average log-likelihood:

$$
\begin{aligned}
L(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}) = \frac{1}{KT} \sum_{t=1}^{T} \sum_{k=1}^{K} \log \big[& \mathbb{1}\{w_t(k) = 0\} h_t(x_t(k), *, \boldsymbol{\theta_t}) + \mathbb{1}\{m_t(k) = 0\} h_t(*, z_t(k), \boldsymbol{\theta_t}) \\
& + \mathbb{1}\{m_t(k) = 1, w_t(k) = 1\} h_t(x_t(k), z_t(k), \boldsymbol{\theta_t}) \big]
\end{aligned}
$$

Calculating the likelihood function for a given parameter vector $\boldsymbol{\theta}$ first involves solving the fixed point problem described in Equation 3.4 to derive the inclusive values. This can be achieved by setting up an inner loop which will apply the contraction mapping until convergence. The estimator proposed is then defined as:

$$
\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta})
$$

Asymptotic inference for $\hat{\boldsymbol{\theta}}$ is then standard if the size of the sample is not too large relative to the size of the overall economy. As noted in Menzel (2015) and Diamond and Agarwal (2017), the inherent structure of matching markets could introduce dependence between observations.

A bootstrap procedure could then be used for inference otherwise (Diamond and Agarwal, 2017; Menzel, 2021). Monte Carlo simulations testing the validity of the proposed estimation strategy can be found in Appendix 3.E.

3.6 Empirical Strategy

The rest of the paper leverages the above general methodology to identify and estimate teachers' and schools' preferences and investigate the determinants of the observed spatial sorting and job-to-job flows. Before showing the results of the empirical analysis, I briefly describe how I adapt the model to the context under study by defining the estimation sample, how model primitives are parameterized and discussing the identification strategy.

Estimation Sample: Throughout the empirical analysis, I consider one side as being teachers and the other side as jobs such that matching is one-to-one.[15] I use several parts of the centralized assignment data for identification and estimation. First, I use information on the universe of applicants and positions that participate in the centralized allocation for the academic years 2016, 2018 and 2020. This allows me to identify directly from the data the distribution of aggregate states m_t and w_t for $t = \{2016, 2018, 2020\}$. I then use data on realized matches following the sampling process described in Section 3.5 in order to identify the joint distribution of matched characteristics f_t for $t = \{2016, 2018, 2020\}$. Finally, I supplement the analysis with additional data on how schools rank the applicants they interview. This allows me to overcome the negative result highlighted in Proposition 3 by separately identifying preferences from the joint surplus. As the horizon of the data is limited, I fix the distribution of aggregate states and the payoff functions to be stationary from 2020 onward and set the horizon of the model to $T = \infty$. This avoids assuming that the continuation value of a match in 2020 is zero.

Permanent vs. Temporary Contracts: I consider the joint allocation of permanent and temporary positions. Permanent contracts have several non-standard features that the model needs to account for. Teachers are forced to stay at least three years in the first permanent

[15] In a static setting, observing the same school making several choices brings additional identification power to pin down schools' unobserved preference heterogeneity (Ederer, 2022). Investigating whether this result also holds in a dynamic setting is left for future work.

job they accept. Once a teacher accepts a permanent position, it can no longer participate in the centralized allocation mechanism (see Appendix 3.B for more details). This has several implications. First, this implies that choosing a permanent contract is a commitment to stay at least three years in the same location. It is thus crucial to model how agents sort between these two types of contracts in order to explain teachers movements across locations.[16] Second, this means that choosing a permanent position is a terminating action as teachers exit the market if they do so. I thus specify the lifetime utility that teacher i gets from choosing a permanent position j as follows:

$$U_{ijt} = U(x_{it}, z_{jt}, \boldsymbol{\theta}_{\text{perm}}) + \sigma\eta_{ijt}$$

while the utility that teacher i gets from choosing a temporary position k is defined as:

$$U_{ikt} = U(\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}, \boldsymbol{\theta}_{\text{temp}}) + \sigma\eta_{ikt} + \beta_w \int \overline{U}_{it+1}(\boldsymbol{x}_{it+1}) w(\boldsymbol{x}_{it+1}|\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) d\boldsymbol{x}_{it+1}$$

Similarly, on the school side, I assume that accepting to match with a teacher with a permanent contract is a terminating action. I thus define the utility that a school with a permanent vacancy j gets from being matched with teacher i as:

$$V_{ijt} = V(\boldsymbol{x}_{it}, \boldsymbol{z}_{jt}, \boldsymbol{\gamma}) + \sigma\epsilon_{ijt}$$

As for temporary jobs, the allocation mechanism is priority-based and schools cannot express their preferences, I assume that the utility that a school with a temporary vacancy j gets from being matched with teacher l is:

$$V_{ljt} = s_{lt}$$

where s_{lt} is teacher l's test score in period t. This slightly simplifies the problem as we can directly observe which temporary jobs are in teachers' opportunity sets.

Parametrization Payoffs: As one time period spans two academic years, I first set the discount factor β_w to 0.9. The model aims to capture (i) how teachers trade off geographical

[16]Figure 3.A.2 shows that sorting across permanent and temporary contracts explains a large part of the observed attrition patterns.

proximity with other job characteristics such as wages and amenities and (ii) how schools value observed measures of teacher quality. I thus parametrize teachers' payoff function as follows:

$$U(\boldsymbol{x_{it}}, \boldsymbol{z_{jt}}; \boldsymbol{\theta}) = \theta_0 + \theta_1 w_{jt} + \boldsymbol{a'_{jt}\theta_2} + \boldsymbol{d'_{ij}\theta_3} + \boldsymbol{m'_{ijt}\theta_4} + \boldsymbol{z'_{jt}\theta_5} + \boldsymbol{x'_{it}\theta_6}$$

Where w_{jt} is the monthly wage offered in school j in year t, $\boldsymbol{a_{jt}}$ is a vector of indicators measuring the local level of amenities through the availability of a range of services such as electricity, sewage, medical centers, internet and libraries, $\boldsymbol{d_{ij}}$ is a spline of the distance between school j and teacher i's home location and $\boldsymbol{m_{ijt}}$ is a set of dummies indicating if teacher i's current location is in the same region or province as school j. I also include other teacher characteristics $\boldsymbol{x_{it}}$ such as experience, marital status, gender and age as well as other school/locality characteristics $\boldsymbol{z_{jt}}$ part of the bonus scheme driving the variation in wages. I then parametrize schools' payoff function as:

$$V(\boldsymbol{x_{it}}, \boldsymbol{z_{jt}}; \boldsymbol{\gamma}) = \gamma_0 + \boldsymbol{s'_{it}\gamma_1} + \boldsymbol{e'_{it}\gamma_2} + \boldsymbol{z'_{jt}\gamma_3} + \boldsymbol{x'_{it}\gamma_4}$$

Where s_{it} is a vector of the different components of teacher i's test score in period t, e_{it} is a vector of dummies dividing the experience level of teacher i in period t in discrete categories. I also include various additional teacher and school characteristics in $\boldsymbol{x_{it}}$ and z_{jt}. Note that I exclude wages, amenities and geographical proximity from schools' preferences. I directly test for these exclusion restrictions by estimating schools' preferences separately and find that we cannot reject that the parameters associated to these characteristics are jointly equal to zero.

The main parameters of interest on the teacher side are $(\theta_1, \theta_2, \boldsymbol{\theta_3}, \boldsymbol{\theta_4})$. They quantify the trade offs between wages, amenities and geographical proximity which drive how mobile labor supply is. On the school side, the main parameters of interest are $\boldsymbol{\gamma_1}, \boldsymbol{\gamma_2}$ as they are likely to explain how the demand side rations excess supply and thus how teacher quality is distributed across locations.

Transition processes: I separate state variables evolving over time in several groups. I assume that age and experience evolve deterministically and exogenously by getting incremented by one every year. The competency score s_{it}, which is contained in x_{it}, evolves stochastically and exogenously. I assume that the transition distribution function of s_{it} is conditionally normal such that:

$$s_{it+1}|\boldsymbol{x_{it}}, \boldsymbol{z_{jt}} \sim \mathcal{N}(\boldsymbol{x'_{it}}\boldsymbol{\beta_x} + \boldsymbol{z'_{jt}}\boldsymbol{\beta_z}, \sigma^2)$$

I estimate $(\boldsymbol{\beta_x}, \boldsymbol{\beta_z})$ via an auxiliary linear regression and report the estimates in Table 3.A.4. Finally, $\boldsymbol{m_{ijt}}$ evolves deterministically and endogenously. Each move across provinces or regions updates teachers' current location such that they internalize that moving again in the next period might be costly.

Additional identifying variation: To overcome the negative result of Proposition 3 and separately identify teachers' and schools' preferences from the joint surplus, I use additional data on how schools rank the applicants they interview. As the set of interviewees in each school is determined by teachers' rank-ordered list and priority index, schools' choice sets are independent of schools' unobserved preferences by construction. Schools do not have incentives to misreport their preferences at this stage as job offers are automatically sent in order of the reported ranks. I thus assume that these rankings are truthful and use them to construct the corresponding exploded logit conditional choice probabilities. The log of these CCPs then enters additively in the log-likelihood derived in Section 3.5.[17]

Discussion of the stability assumption: While I cannot directly test whether the observed matching is stable, several properties of the allocation mechanism, described in Section 3.2, limit the presence of frictions that could lead to the existence of blocking pairs. First, the allocation of temporary positions is implemented via serial dictatorship by sequentially asking teachers to choose their preferred position by order of priority. This procedure is equivalent to Deferred Acceptance and thus leads to a stable allocation. Second, the allocation of interviews for permanent positions is also done via serial dictatorship, which ensures that teachers get interviewed by their preferred schools by order of priority, if they reveal their true preferences. Still, as the number of interviews per teacher is limited to three, two issues might arise: (i) teachers might end up unmatched because they failed all their interviews while schools with unfilled positions might be willing to hire them and (ii) teachers might anticipate this

[17]As this likelihood is not exact, I correct for standard errors using the standard formula for the asymptotic variance of QMLE.

possibility and try to avoid it by being strategic when forming their rank-ordered lists.[18]

To mitigate the first concern, the Ministry implemented an aftermarket such that all unassigned permanent positions and teachers can meet and match in order to minimize justified envy. Modeling potential mismatches generated by the second concern would be challenging as this would entail having access to data on teachers' beliefs about their chances to succeed at the interviews and developing a dynamic model of strategic reporting where preferences of both sides of the market are unknown. This is beyond the scope of the available data and the proposed methodology and is left for future research. Instead, I propose a test assessing the validity of the estimated parameters by leveraging the cutoffs determining eligibility to permanent positions in a regression discontinuity design.[19] Teachers just above the cutoff have both permanent and temporary positions in their choice sets while teachers just below the cutoff only have temporary positions in their choice sets. Comparing the matching outcomes of these two groups allows to pin down how teachers trade off job attributes depending on whether the position is permanent or temporary. I thus verify whether the estimated model can replicate the threshold crossing effect on the characteristics of teachers' matched schools. Figure 3.A.4 shows that the model predictions match the observed responses at the threshold. Eligible teachers are more likely to choose a permanent position and are willing to trade off the benefits of permanent contracts with geographical proximity.

3.7 Results

3.7.1 Preferences and the Spatial Job Ladder

I report in Panel A of Table 3.2 the estimated willingness to pay of teachers for amenities, proximity to home and moving away from their current location. Consistently with the migration literature, I find a large distaste for moving (Kennan and Walker, 2011). Teachers would be willing to give up 309 USD from their monthly wage to avoid moving 10 kilometers

[18]Additionally, teachers might have ex-post justified envy if they refuse a permanent position and realize ex-post that the available temporary positions are worse. To solve this issue, one could instead model the allocation of permanent and temporary contracts sequentially and not jointly. However, the results derived in Appendix 3.F show that these two models are observationally equivalent in the limit if teachers have rational expectations about their future match payoffs when choosing a permanent contract.

[19]Using an external source of data to separately identify schools' preferences using truthful rankings also allows to mitigate these concerns.

Table 3.2: Selected Preference Estimates

Panel A: Teachers' Preferences (in monthly USD)	
Amenities	-
Electricity	97.67 (34.71)
Sewage	16.76 (12.79)
Library	29.41 (15.78)
Internet	15.96 (19.86)
Spline Distance from Home Location	-
Slope $<$ 20km	-30.94 (1.96)
Slope $\in$ [20km,100km]	-7.11 (0.50)
Slope $\geq$ 100km	-1.38 (0.10)
Moving Costs	-
$\neq$ Province	-676.95 (51.43)
$\neq$ Region	-469.62 (57.21)
Panel B: Schools' Preferences	
Constant	-0.488 (0.082)
Experience	-
$<$ 3 years	-0.737 (0.041)
$>$ 10 years	0.057 (0.040)
Competency Score	-
Reading	0.669 (0.023)
Logic	0.571 (0.020)
Curricular Knowledge	1.397 (0.022)

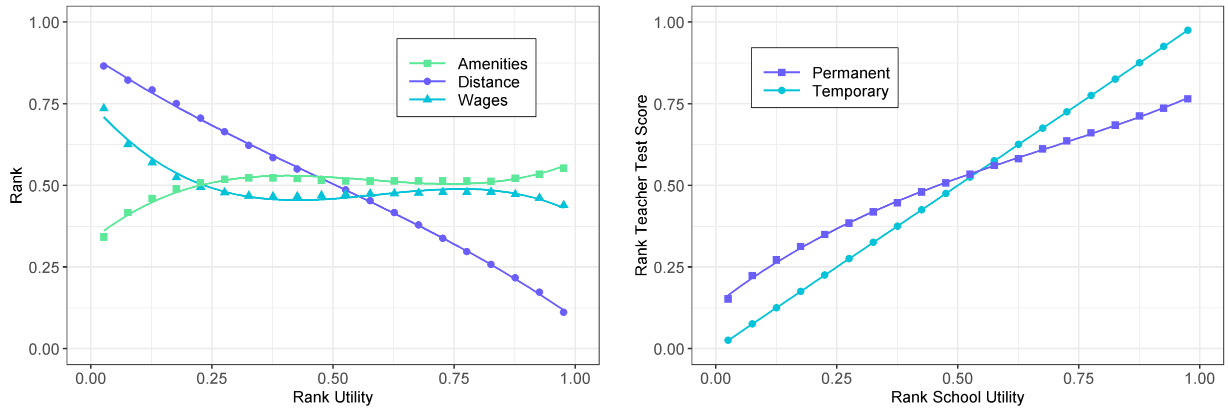
NOTES. This table shows selected estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ from the specification of teachers and schools preferences. θ_1 is normalized to one such that teachers' preference estimates are expressed in terms of monthly willingness to pay in USD.

away from home. This is quite substantial as this corresponds to 61 % of the base monthly teacher wage. Similarly, teachers' willingness to pay to avoid moving out of their current location is large. The cost of moving out of their current province is estimated at 677 USD while the cost of changing regions is estimated at 1,146 USD. In comparison, the willingness to pay for local amenities is quite small and ranges from 16 USD to 98 USD.[20]

To quantify how much these attributes explain the variation in teachers' preferences, I simulate teachers' lifetime utility by drawing random Gumbel shocks and by using the estimated parameters to compute their flow utility and continuation value for each job. I

[20]I report how the willingness to pay for different job attributes differ depending on whether the contract is permanent or temporary, under the assumption that agents are myopic, in Table 3.A.5. I strongly reject that $\boldsymbol{\theta}_{\text{perm}} = \boldsymbol{\theta}_{\text{temp}}$ which is equivalent to rejecting that agents are myopic, as permanent and temporary contracts do not differ in the first years of employment.

Figure 3.3: Spatial Job Ladder



a) Supply **b) Demand**

NOTES. Panel A plot the relationship between the rank of teachers' lifetime utility U_{ijt} estimated using the results displayed from Table 3.2 and the ranks of various job attributes such as: the distance between teachers' home location and the school's locality, the wage offered by the schools and the level of local amenities. Panel B performs the same exercise and plot the ranks of schools' estimated utility V_{ijt} against the ranks of teachers' test scores for both temporary and permanent positions.

then plot the ranking of each job according to its predicted lifetime utility against its ranking in terms of distance, amenities and wages. Panel A of Figure 3.3 shows that distance very strongly predicts how teachers rank jobs. The correlation between the ranking with respect to utility and the ranking with respect to distance is 0.68. On the other hand, I find that wages and amenities are poor predictors of how teachers rank jobs. This implies that labor markets are very local as labor supply is not mobile, which is consistent with the findings of Manning and Petrongolo (2017).

Panel B of Table 3.2 shows the results of the estimation of schools' preferences. I find that schools highly value observed measures of teacher quality such as test scores and experience. I investigate how much of the ranking of teachers with respect to schools' utility can be explained by their ranking with respect to test scores. Panel B of Figure 3.3 plots the relationship between the two for both permanent and temporary jobs. Mechanically, the relationship is one-to-one for temporary jobs, as test scores are used as priorities to allocate seats. The relationship is also very strong for permanent jobs as the correlation between the utility ranking and the test score ranking is 0.6.

Overall, the estimated preference parameters indicate that (i) geographical proximity is highly predictive of how teachers rank the available jobs and (ii) schools mostly value observed measures of teacher quality such as teachers' test scores. These two facts have strong implications for spatial sorting and inequalities.

I first show that the combination of fact (i) with the concentration of teachers' home location in cities and the dispersion of jobs across the country, documented in Section 3.3, implies the existence of a *spatial job ladder*. As teachers' home location is concentrated in cities, teachers' distaste for working far from their home location implies a strong distaste for schools located in remote areas. Additionally, as schools are geographically scattered, cities offer very few positions compared to the total number of applicants. Overall, this implies that remoteness becomes the main driver of how teachers rank the available jobs. This results in the existence of a spatial job ladder where jobs located in remote areas are at the bottom whereas jobs located in cities are at the top. As a result, schools in cities face excess supply and are free to hire the teachers they prefer from the set of new applicants or poach their preferred teachers from schools which are on a lower rung of the ladder.[21] This has direct implications on labor market dynamics, as teachers which start at the bottom of the ladder switch jobs at a higher rate to climb up toward urban areas. Consequently, schools located in remote areas face higher attrition rates than schools located in cities.

The extent to which the spatial job ladder translates into spatial inequalities in education provision depends on which teacher attributes schools value. If schools would select teachers at random, lucky teachers would be able to move to urban schools but this would not generate unequal sorting with respect to teaching quality.[22] Fact (ii) implies that urban schools ration excess supply using observed measures of teacher quality such as test scores and experience. Consequently, the spatial job ladder creates large spatial inequalities in teaching quality through two channels. First, among the set of new applicants, urban schools systematically select the highest scoring teachers while rural schools are left with the lowest scoring teachers. Second, urban schools poach teachers who have accumulated sufficient experience and human capital throughout their career from rural schools. The latter thus fail to retain skilled teachers and sustain disproportionately low levels of teaching experience and quality.

Reducing spatial inequalities in teaching quality thus requires shutting down the mechanisms through which the spatial job ladder operates or directly targeting the causes of the

[21]This contrasts with traditional models of the job ladder where productive firms offer higher wages to poach skilled workers from unproductive firms (Moscarini and Postel-Vinay, 2018). In this setting, where wage differentiation is very limited, the job ladder is instead determined by non-pecuniary factors such as geographical location.

[22]Still, attrition would be higher in rural schools which could have a disruptive effect on student learning and imply a net efficiency loss in teaching quality due to the loss of school specific experience. See Appendix 3.C for estimates of the net loss in teacher value added implied by a move from one school to another.

existence of the spatial job ladder. Next, I use the tools developed in Section 3.4 along with the estimated teachers' and schools' payoff functions to perform several counterfactual experiments aiming at achieving these goals. Before doing so, I assess the credibility of the equilibrium predictions generated by the model by testing its ability to predict patterns consistent with the data.

3.7.2 Model Fit

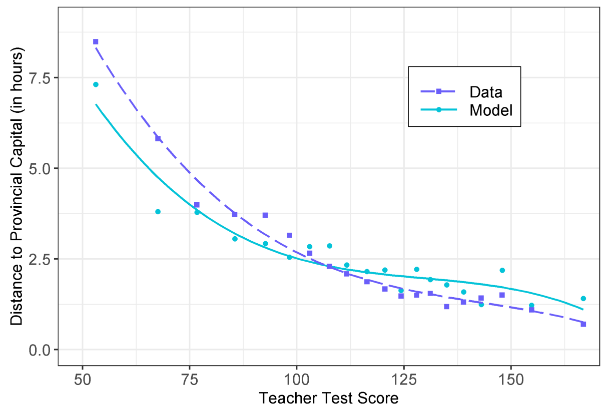
I perform several checks to assess how well the model predicts the patterns generated by the spatial job ladder. I first test whether the cross-sectional spatial sorting patterns predicted by the model match the ones observed in the data. I then simulate job-to-job flows and check whether the model can replicate the observed movements of teachers from rural to urban areas.

To simulate the status quo equilibrium matching, I first derive $U(\boldsymbol{x_{it}}, \boldsymbol{z_{jt}}, \hat{\boldsymbol{\theta}})$ and $V(\boldsymbol{x_{it}}, \boldsymbol{z_{jt}}, \hat{\boldsymbol{\gamma}})$ for all t. I then randomly draw Extreme Value Type I taste shocks ϵ_{ijt} and η_{ijt} for all (i, j, t) to construct the flow utilities. I solve the fixed point problem described in Equation 3.4 by fixing the aggregate distributions of observables in 2016 as the baseline m_1 and w_1 to obtain the equilibrium inclusive values for each agent. Given the inclusive values, I then compute teachers' continuation value from choosing a temporary contract. I then simulate forward by constructing the lifetime utilities U_{ij1} and V_{ij1}, deriving the teacher-optimal stable match using the teacher-proposing Deferred Acceptance algorithm and updating teachers' location, experience, age and test scores using the estimated transition process. I then iterate this procedure to simulate the entire non-stationary equilibrium path.

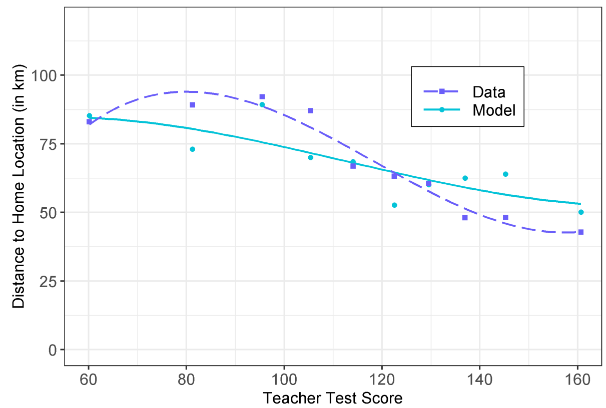
Panel A of Figure 3.4 shows a binned scatter plot of the relationship between teachers' test scores and the remoteness of their matched school. The model is able to replicate the rationing of excess supply through test scores and generate the strong negative relationship between remoteness and test scores. Panel B of Figure 3.4 also shows that the model is able to replicate sorting with respect to geographical proximity. Overall, this indicates that the main drivers of spatial sorting are well captured by the estimated preferences.

Panel C and D of Figure 3.4 compare the career paths of teachers depending on where they started on the spatial job ladder with their simulated counterparts. Specifically, I compare the job-to-job flows from rural to urban areas as teachers climb up the ladder. I find that the model captures the trend that teachers originally matched to rural schools climb the job

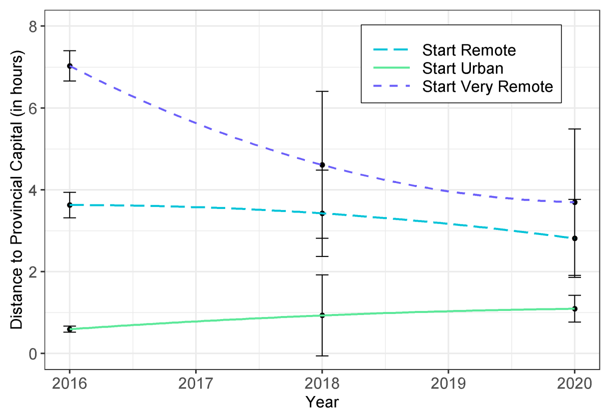
Figure 3.4: Model Fit



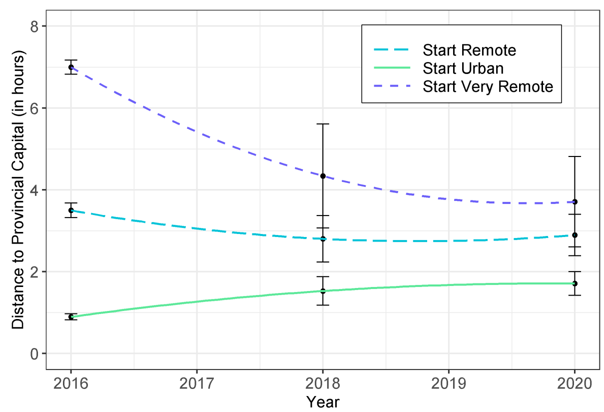
a) Sorting: Remoteness



b) Sorting: Distance



c) Job-to-Job Transitions: Data



d) Job-to-Job Transitions: Model

Notes. This figure uses the centralized assignment data from 2016 to 2020 and compares realized sorting patterns and job-to-job transitions with model predictions. Panel A plots averages of the remoteness of teachers' matched schools based on equally spaced bins of the distribution of teachers' test scores both in the actual data and in the simulated equilibrium. Panel B shows the result of a similar exercise using the distance between teachers' matched schools and their home location. Panel C plots the evolution over time of teachers' matched schools observed in the data depending on the remoteness of the school in which they started in 2016. Panel D plots the same trend using the job-to-job transitions simulated by the model.

ladder by moving toward urban areas. The model is thus able to replicate the important labor market dynamics that characterize the spatial job ladder.

3.7.3 Counterfactuals

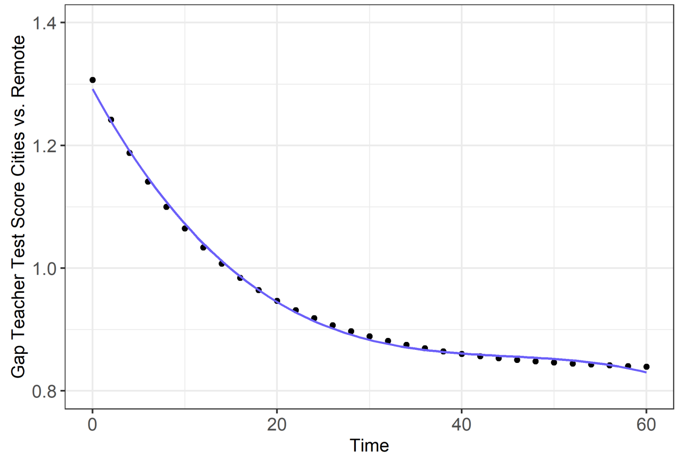
In this section, I first quantify the gains of shutting down labor mobility along the job ladder to isolate the role of labor market dynamics in explaining the observed urban-rural gap in teaching quality. I then explore the effectiveness of retention policies that would prevent teachers from climbing up the ladder. More specifically, I simulate the effect of imposing a minimum contract length, which is a commonly used retention policy in the public sector. Finally, I explore the equilibrium effects of tackling directly the root causes of the existence of the spatial job ladder. To do so, I simulate equilibrium sorting and mobility under the scenario where teachers' home location would be scattered across the country instead of being concentrated in cities.

Labor Market Dynamics and Spatial Inequalities

Teacher mobility likely contributes to a large extent to the urban-rural gap in teacher quality as teachers matched to rural areas leave toward urban areas once they have accumulated skills and experience. As shown in Section 3.3, movers are of significantly higher quality that those who replace them as the job ladder rewards teachers with higher test scores and more experience. To quantify how much of the urban-rural gap in teaching quality is explained by labor mobility, I start by simulating the equilibrium path under a counterfactual scenario where agents would have a very high preference for staying in their current job. Assuming that μ^*_{2016} is the equilibrium match under the status quo in 2016, I thus artificially increase $U_{i\mu^*_{2016}(i)t}$ for all teachers i and all subsequent years $t > 2016$ and simulate the long-run equilibrium paths. This counterfactual exercise shuts down voluntary moves away from rural areas such that rural schools no longer lose their most qualified and experienced teachers and can benefit from the accumulation of human capital on-the-job.

Figure 3.5 plots the evolution of the urban-rural gap in teacher quality from 2016 onward under this counterfactual. I find that shutting down labor mobility makes the urban-rural gap in teacher test score sharply drops from 1.3 to 0.8 standard deviation in the long run. This decline is stronger in the short run as the gap decreases by 0.1 standard deviation after

Figure 3.5: Labor Market Dynamics and Spatial Inequalities



NOTES. This figure shows the result of artificially increasing teachers utility for their matched school in 2016 in order to quantify the share of the urban rural gap in teacher quality explained by teacher mobility. It plots the difference between the average teacher score (in standard deviations) in schools located in cities and schools located more than six hours away from the provincial capital along the transition path triggered by this counterfactual exercise over 60 years.

four years only.[23]

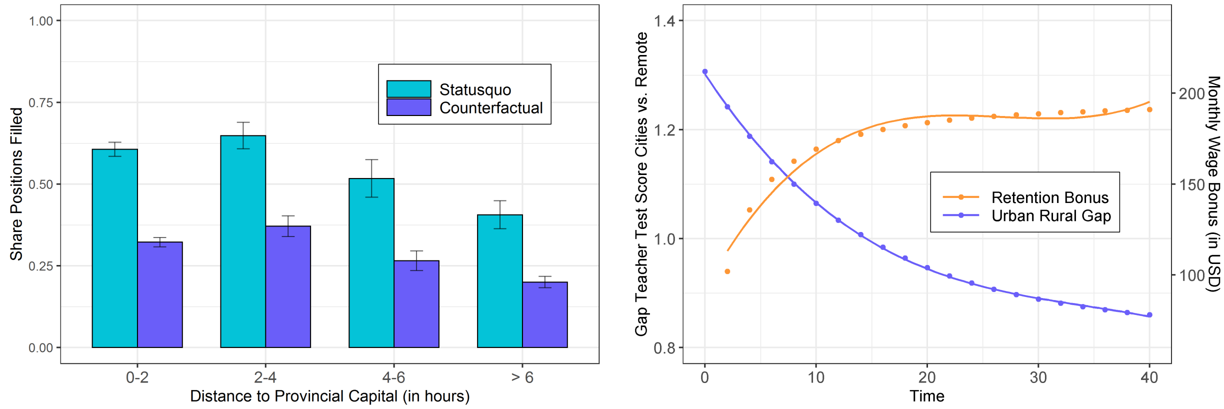
This exercise allows us to decompose the channels through which the spatial job ladder fuels spatial inequalities in teaching quality by shutting down labor market dynamics. Schools at the bottom of the ladder can now retain their highest skilled teachers while schools at the top of the ladder, on the contrary, can no longer poach skilled teachers from rural schools. Overall, this exercise shows that labor market dynamics explain 38% of the existing urban-rural gap in teacher quality. The remaining 62% are explained by initial unequal sorting in 2016 that cannot be offset by human capital accumulation on-the-job. This result highlights the importance of labor market dynamics in explaining spatial sorting and inequalities, even in a frictionless setting with rigid wages. It also suggests that there might be important benefits in investing in retaining existing teachers rather than aiming at recruiting higher quality teachers.

Evaluating Retention Policies

I then investigate the effectiveness of retention policies aiming at shutting down labor mobility and its adverse effects on spatial inequalities. Using the estimated model, I simulate the effects of removing teachers' option to rematch by enforcing a minimum contract length. If agents

[23]This is driven by the fact that teachers test scores evolve more rapidly when they start from lower initial values, as suggested by the estimates in Table 3.A.4.

Figure 3.6: Compulsory Service Policy



a) Adverse Sorting **b) Retention Bonuses**

NOTES. Panel A of this figure plots the share of filled vacancies for different bins of schools' remoteness in the status quo and under the counterfactual scenario where we would enforce a minimum contract length of four years. Panel B plots the effect of enforcing this policy on the urban rural gap in teacher test scores (in standard deviations) along with the monthly wage bonuses that would offset the adverse sorting effect shown in Panel A. The x-axis represents the minimum length of the contract.

were myopic, I find that this policy would reach the same results as described in Figure 3.5 and close the urban-rural gap in teaching quality by 38% in the long-run by stopping skilled teachers from leaving rural schools. However, as agents are forward looking, teachers react ex-ante to this policy and their labor market participation plunges creating large shortages. Panel A of Figure 3.6 shows the share of filled vacancies under the status quo and under the policy which would enforce a minimum contract length of four years. As this policy forces teachers to commit and does not allow them to rematch and climb the ladder, they prefer to wait until they get better matching opportunities in the future. This results in a sharp drop in the share of filled vacancies. This finding highlights a key trade off between recruitment and retention. In the presence of a job ladder, retention policies that make rematching more difficult imply a significant decrease in the continuation value of accepting a job and generate strong adverse sorting responses. To avoid the latter, such policies should compensate workers for preventing them to improve their matching outcomes through job switching.

I then compute the amount that should be given as compensation to avoid this adverse sorting effect. I define the status quo match as μ^* and compute the monthly wage bonuses b_i for each teacher i which solve the following equation:

$$U_{i\mu_t^*(i)t} = U(x_{it}, z_{jt}, \hat{\boldsymbol{\theta}}) + \hat{\theta}_1 b_i + \eta_{ijt} + 0.9\Big(\int\int U(x, z, \hat{\boldsymbol{\theta}}) w(x|x_{it}, z_{jt}) m(z|x_{it}, z_{jt}) + \hat{\theta}_1 b_i + \gamma$$

$$+ 0.9 \int \overline{U}_{t+2}(s) w(s|x, z) ds dx dz \Bigg)$$

The bonus b_i solving this equation makes teacher i indifferent between matching to a school for at least two years and matching to the same school for at least four years. Implementing this retention bonus scheme thus avoids the adverse sorting effect documented in Panel A of Figure 3.6. A similar equation can be formulated to compute the bonuses necessary to retain teachers for an additional τ years. I denote the solution to these equations for teacher i b_i^τ.

I compute b_i^τ for $\tau \in \{2, 4, 6, ..., 40\}$ and plot its average for each τ in Panel B of Figure 3.6. I also report the effect of this policy on the urban rural gap in teacher test scores measured in standard deviations on the same figure. Imposing a minimum contract length of four years would entail compensating teachers by a 100 USD bonus on their monthly salary on average. This number gradually increases as the minimum contract length increases before reaching a plateau of approximately 200 USD.[24]

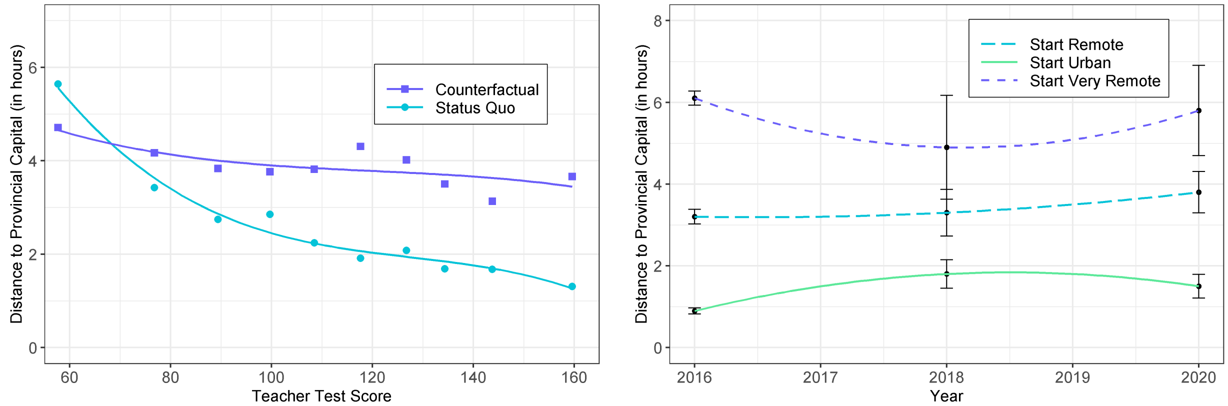
Overall, this result shows that retention policies have large potential benefits in the long run but come at a cost which should be benchmarked against other alternatives. I find that this policy would reduce the urban-rural gap in teacher quality by 38% for an average monthly cost of 200 USD per teacher, which corresponds to a 40% increase in their monthly salaries.

Shutting Down the Spatial Job Ladder

The existence of the spatial job ladder is mainly caused by three factors: (i) teachers' distaste for moving far from home, (ii) the concentration of teachers' home location in cities and (iii) the geographical dispersion of schools. As the spatial job ladder is responsible for the observed spatial inequalities, the most effective way of reducing inequalities would be to target its fundamental causes. In this section, I take (i) and (iii) as given and explore, as a thought experiment, what would be the consequences of shutting down (ii). To do so, I perform a counterfactual exercise that randomly changes teachers' home locations such that they are scattered across the country. I randomly draw teachers' new home location from the set of localities in which schools are situated. I then simulate equilibrium sorting and movements across locations.

[24]The concavity of the wage bonus scheme comes from the fact that agents discount the future at an increasing rate.

Figure 3.7: Random Home Location



a) Sorting **b) Job-to-Job Flows**

NOTES. This figure shows the results of a counterfactual experiment which would reallocate teachers' home location randomly across Peru. I randomly draw the location of each teacher from the list of localities in which school are situated and recompute the equilibrium. Panel A plots binned averages of teachers matched school's remoteness in the status quo and under this counterfactual. Panel B plots the counterfactual evolution of the remoteness of teachers matched schools from 2016 to 2020 starting from different initial levels of remoteness.

Panel A of Figure 3.7 plots teacher sorting with respect to test scores and schools' remoteness under this counterfactual exercise. I find that the spatial job ladder collapses. As labor supply is scattered across the country, competition for local jobs disappears. Teachers match overall close to home and no longer have a systematic distaste for remote schools. High skilled teachers are thus no longer disproportionately matched to schools located in cities. Labor market dynamics are also strongly affected. The rate at which teachers move throughout the period 2016-2020 drops by half as low quality teachers are no longer sent far from home. Panel B of Figure 3.7 shows that the direction of the flows also changes as teacher no longer leave rural schools to get closer to urban centers. As a result, urban-rural inequalities in teacher attrition disappear and rural schools can benefit from experience and skill accumulation on-the-job. These results shows that designing policies targeting the root causes of the existence of the spatial job ladder, such as investing in training local teachers, might be more effective than aiming at slowing down its symptoms through recruitment or retention policies.

3.8 Conclusion

This paper investigates the causes of teacher spatial sorting and mobility and their consequences on spatial inequalities in teaching quality. To this end, I develop an empirical framework of dynamic matching without transfers. I assume that agents make forward-looking matching decisions and that their payoff functions depend on a various set of job and teacher attributes. I provide a tractable way to map teachers' and schools' preferences into sorting and job-to-job flows. I then show that one can invert this mapping and identify agents' preferences from data on realized sorting.

Using this methodology, I then show the existence of a spatial job ladder. Teachers concentrate in cities while jobs are scattered geographically. As teachers have a strong distaste for moving, this creates excess supply in cities which is rationed using observed measures of teacher quality. As a consequence, teacher quality is highly unequally distributed and teachers working in remote areas leave toward urban areas as soon as they have accumulated enough experience. Overall I find that labor mobility magnifies inequalities in teaching quality by one third. Finally, I assess the effectiveness of retention policies aimed preventing teachers from rematching along the job ladder. I find that this triggers a massive flow out of the teaching profession such that the positive effects of retention are largely outweighed by the losses incurred through teacher shortages. This highlights a key trade off between recruitment and retention in the presence of a job ladder and shows that retention policies should compensate for the implied lack of flexibility.

Appendices

3.A Additional Tables and Figures

Table 3.A.1: Data Description

	2016	2017	2018	2019	2020	2021
Panel A: Teacher Assignment Data						
# Teachers	116,559	116,939	116,128	115,358	115,233	116,024
Permanent	94,162	89,604	91,683	90,889	89,106	87,507
Temporary	22,397	27,361	24,466	24,505	26,174	28,516
Panel B: Centralized Allocation Mechanism						
# Test Takers	77,594	-	78,758	68,301	71,586	-
in Permanent Position Alloc.	6,770	-	9,777	5,905	4,005	-
in Temporary Position Alloc.	60,853	-	66,280	-	60,294	-
in Both	3,436	-	4,195	-	2,517	-
in None	13,407	-	6,896	-	9,804	-
# Vacancies	18,493	-	36,113	9,818	17,858	-
in Permanent Position Alloc.	6,460	-	13,620	9,818	5,014	-
in Temporary Position Alloc.	15,372	-	30,645	-	16,481	-
in Both	3,339	-	8,152	-	3,637	-

NOTES. Panel A shows the total number of employed teachers in each year depending as well as the number of teachers holding a temporary or a permanent contract. Panel B displays the number of participants to the national competency test in each year it took place. It also shows the number of applicants and vacancy which participated in the allocation of temporary and permanent positions.

Table 3.A.2: Summary Statistics: Job Characteristics

	Mean	Std. Deviation	Min	25% Pctile	75% Pctile	Max
Job Characteristics						
Baseline Monthly Wage (USD)	537.286	50.234	507.614	507.614	532.995	799.492
Temporary	0.186	0.389	0.000	0.000	0.000	1.000
Multigrade	0.207	0.405	0.000	0.000	0.000	1.000
Single Teacher	0.0462	0.210	0.000	0.000	0.000	1.000
Bilingual	0.107	0.309	0.000	0.000	0.000	1.000
School Characteristics						
Distance Prov. Capital (hours)	1.406	4.073	0.000	0.0609	1.144	72.000
Population	1143.708	2593.554	0.001	0.382	350.766	7567.716
Altitude (meters)	1506.684	1501.124	1.000	120.000	3104.000	5002.000
Local Amenities						
Electricity	0.952	0.215	0.000	1.000	1.000	1.000
Water	0.853	0.354	0.000	1.000	1.000	1.000
Sewage	0.726	0.446	0.000	0.000	1.000	1.000
Medical Center	0.770	0.421	0.000	1.000	1.000	1.000
Internet	0.582	0.493	0.000	0.000	1.000	1.000
Bank	0.388	0.487	0.000	0.000	1.000	1.000
Library	0.303	0.460	0.000	0.000	1.000	1.000

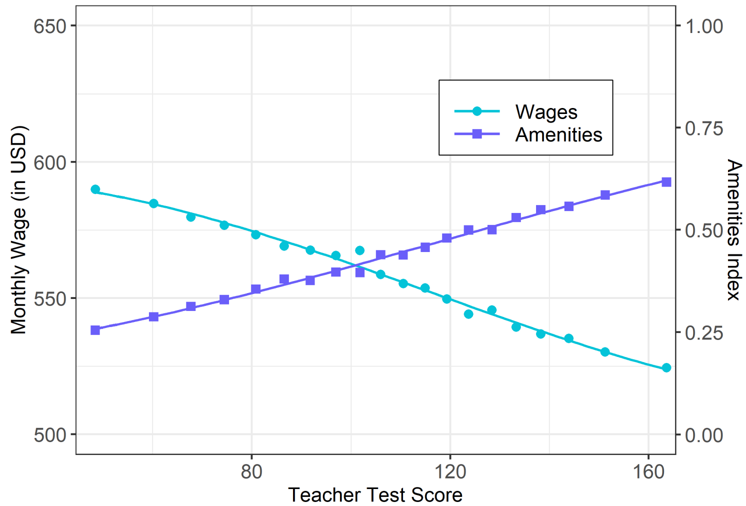
Notes. This table uses the teacher assignment data to show summary statistics on the characteristics of the jobs filled in 2016. The baseline monthly wage does not contain experience bonuses.

Table 3.A.3: Summary Statistics: Teacher Characteristics

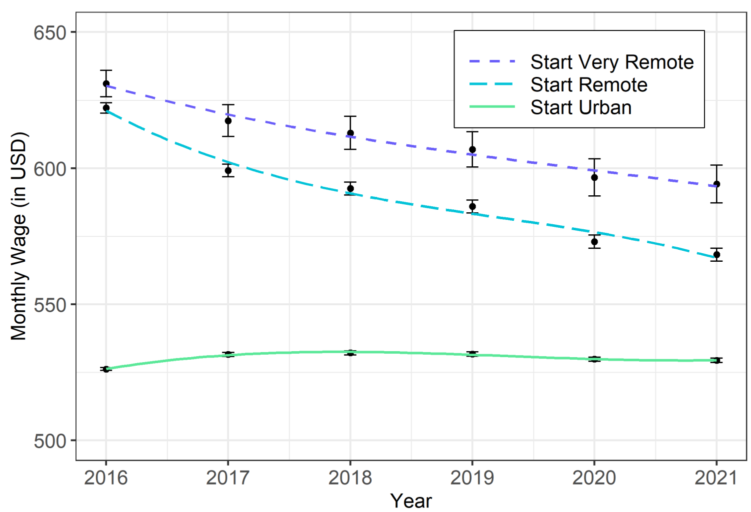
	Mean	Std. Deviation	Min	25% Pctile	75% Pctile	Max
Age	33.676	7.458	19.000	32.000	42.000	78.000
Female	0.707	0.455	0.000	0.000	1.000	1.000
Lives in Provincial Capital	0.819	0.385	0.000	1.000	1.000	1.000
Married	0.471	0.499	0.000	0.000	1.000	1.000
Total Test Score	97.672	29.895	0.000	74.500	119.500	191.500
Score 1: Reading	29.626	9.710	0.000	22.000	38.000	50.000
Score 2: Logical Reasoning	21.963	9.323	0.000	14.000	28.000	50.000
Score 3: Curricular Knowledge	46.083	15.881	0.000	35.000	57.500	97.500
Experience $<$ 3 years	0.237	0.426	0.000	0.000	0.000	1.000
Experience $>$ 10 years	0.130	0.337	0.000	0.000	0.000	1.000

Notes. This table shows summary statistics on the characteristics of the applicants to the centralized assignment platform in 2017.

Figure 3.A.1: Sorting and Movements Across Locations: Wages



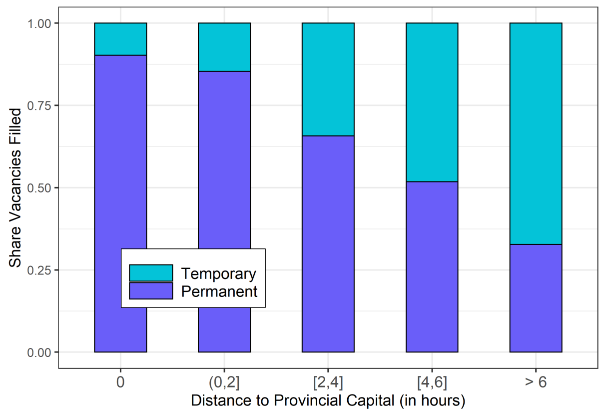
a) Sorting



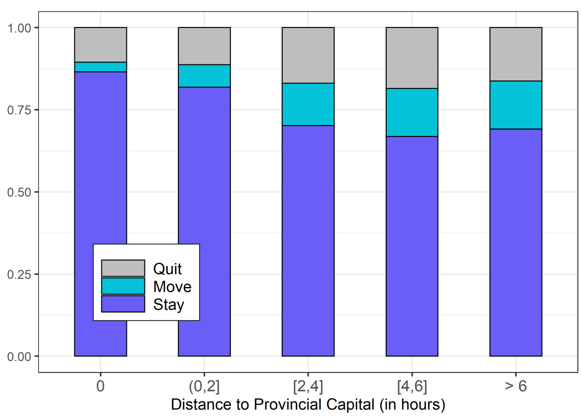
b) Job-to-Job Transitions

NOTES. This figure uses the teacher assignment data to document teacher sorting and movements along other dimensions such as wages and amenities. Panel A plots binned averages of the monthly wage teachers receive as well as the level of amenities in the locality of their matched school based on their test scores. Bins are equally spaced based on vigintiles of the test score distribution. Panel B plots the evolution of the wage received by teachers over the period 2016-2021 depending on where they start in 2016. The purple line corresponds to teachers which start in schools located between 6 and 8 hours away from the provincial capital. The blue line corresponds to teachers which start in schools located between 4 and 6 hours. The green line corresponds to teachers which start in schools located between 2 and 4 hours.

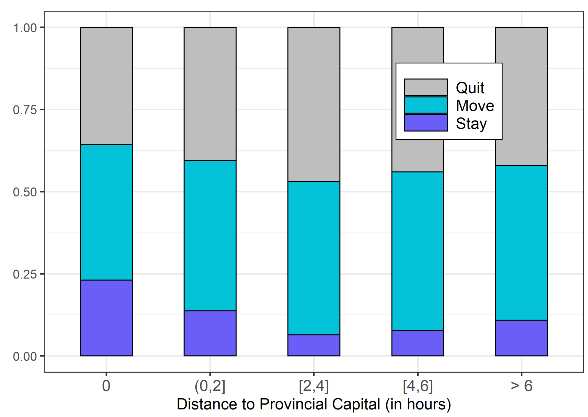
Figure 3.A.2: Temporary vs. Permanent Contracts



a) Sorting Temporary vs. Permanent



b) Transition 2016-2018: Permanent



c) Transition 2016-2018: Temporary

NOTES. Panel A of this figure uses the teacher assignment data to document how teachers sort across types of contract depending on the distance of their matched school to the provincial capital. Panel B and C show, for both permanent and temporary contracts, the share of teachers that stayed in the same school, moved to another school or quit teaching in the public sector for several bins of the schools' distance to the provincial capital.

Table 3.A.4: Transition Process Test Scores

	Estimate	Std. Error
Constant	0.104	0.104
Teacher Characteristics		
Test Score t	0.868	0.007
Female	-0.008	0.011
Age < 30	0.102	0.011
Age > 50	-0.040	0.023
Experience < 3	-0.037	0.016
Exerience > 10	0.016	0.014
Married with kids	-0.030	0.011
School Characteristics		
Wage	-0.040	0.042
Frontier	-0.017	0.020
Bilingual	0.008	0.015
VRAEM	-0.017	0.021
log(Population)	-0.000	0.015
log(Population)^2	0.001	0.002
log(Population)^3	-0.000	0.000
Distance to Capital	-0.003	0.008
Distance to Capital^2	0.000	0.000
Distance to Capital^3	-0.000	0.000
log(pop) $\times$ Distance	0.000	0.001

Notes. This table displays the estimates of the coefficients of the following linear regression: $s_{it+1} = \boldsymbol{x}'_{it}\boldsymbol{\beta_x} + \boldsymbol{z}'_{\mu_{wt}(i)t}\boldsymbol{\beta_z} + \epsilon_{it}$. The estimation sample is composed of the set of teachers who took the test both in 2015 and 2017 as well as the set of teachers who took the test both in 2017 and 2019. Test scores are standardized to have mean 0 and standard deviation 1.

Table 3.A.5: Teachers Preferences: Temporary vs. Permanent

	Temporary	$\times$ Permanent
Constant	356.22 (94.29)	4188.93 (504.84)
School/Locality Characteristics		
Amenities	23.40 (7.02)	50.24 (35.34)
Bilingual	-60.32 (10.51)	-721.86 (55.22)
Frontera	20.01 (11.36)	-15.71 (75.54)
VRAEM	-91.25 (14.99)	144.13 (101.53)
Preference for Home		
Dist ($<$ 20km)	-18.18 (0.50)	-13.75 (2.50)
20km $\leq$ Dist $<$ 100km	-6.02 (0.14)	-0.49 (0.62)
Dist $\geq$ 100km	-0.63 (0.02)	-0.56 (0.10)
Moving Costs		
$\neq$ Province	-520.44 (14.39)	425.44 (50.34)
$\neq$ Region	-121.98 (14.34)	-1006.74 (52.66)
Other Wage Determinants		
log(Pop)	83.30 (20.42)	-161.06 (125.39)
log(Pop)^2	-14.48 (2.12)	19.70 (13.78)
log(Pop)^3	0.54 (0.07)	-0.81 (0.47)
Distance to Capital	90.43 (12.89)	-277.86 (62.11)
Dist^2	-0.66 (1.76)	0.42 (5.44)
Dist^3	0.01 (0.08)	0.01 (0.24)
Dist $\times$ log(Pop)	-9.40 (1.46)	25.52 (6.66)
Teacher Characteristics		
Female	-58.77 (5.75)	-173.92 (40.28)
Urban	-62.59 (7.83)	-194.23 (68.31)
Married with kids	-11.96 (5.36)	99.66 (34.66)
Age $<$ 30	18.32 (5.65)	396.84 (35.24)
Age $>$ 50	-47.80 (15.64)	-141.33 (101.07)
Exp. $<$ 3	14.68 (7.65)	20.88 (56.92)
Exp. $>$ 10	-101.81 (8.55)	106.56 (52.92)

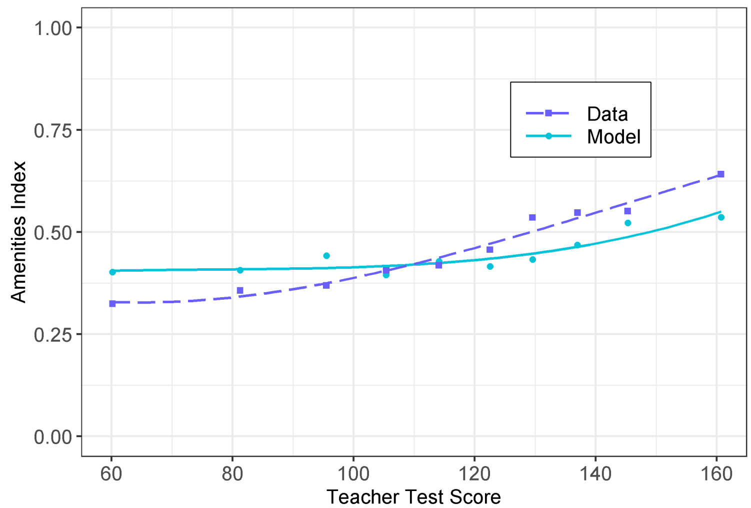
NOTES. This table displays the estimates of $\boldsymbol{\theta}_{\text{temp}}$ and $\boldsymbol{\theta}_{\text{perm}} - \boldsymbol{\theta}_{\text{temp}}$ assuming that $\beta = 0$, meaning that agents are myopic. The wage coefficient is normalized to 1 such that estimates are expressed in monthly willingness to pay in USD. Standard errors are in parenthesis.

Table 3.A.6: Preference Estimates: Schools

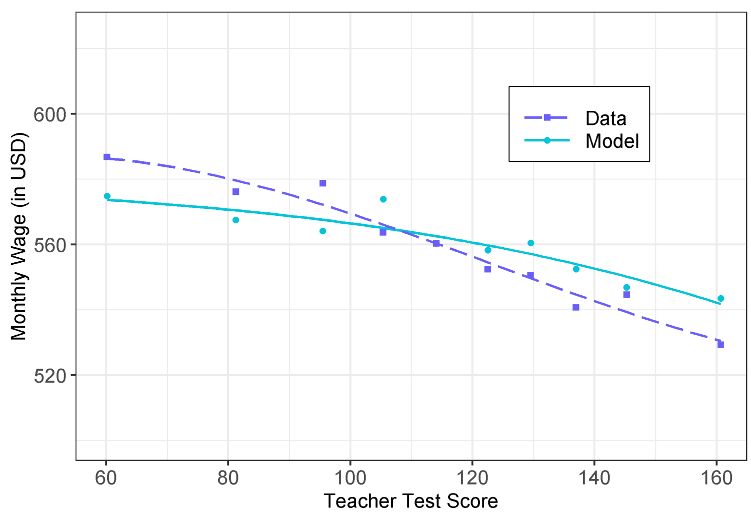
	(1)
Constant	-0.488 (0.082)
Female	-0.349 (0.022)
Married with kids	-0.038 (0.022)
Age < 30	0.128 (0.023)
Age > 50	-0.067 (0.089)
Experience < 3	-0.737 (0.041)
Experience > 10	0.057 (0.040)
Score 1: Reading	0.669 (0.023)
Score 2: Logic	0.571 (0.020)
Score 3: Curricular Knowledge	1.397 (0.022)

NOTES. This table displays the estimates of $\boldsymbol{\gamma}$ which are schools' preference parameters defined in Section 3.6. Standard errors are in parentheses.

Figure 3.A.3: Model Fit Spatial Sorting: Additional Figures



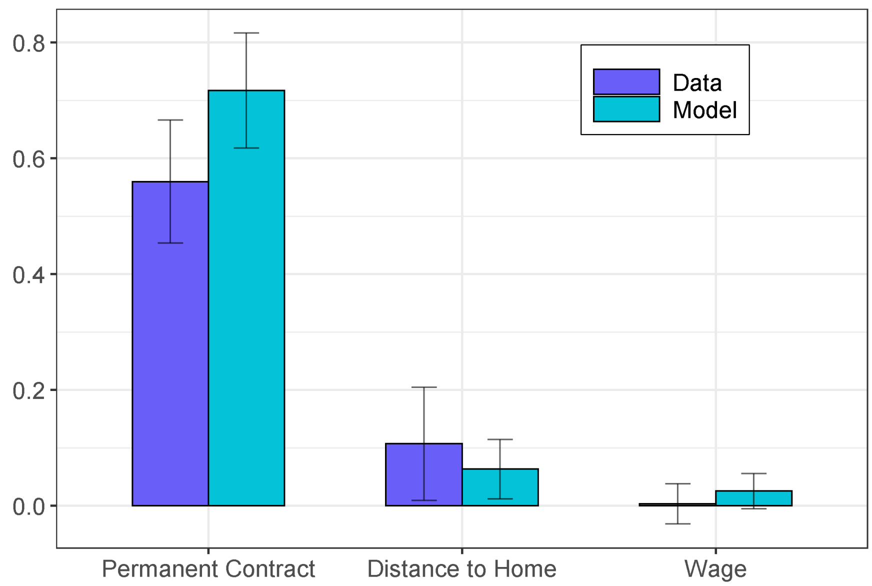
a) Amenities



b) Wages

Notes. This figure uses the centralized assignment data in 2018 and compares realized sorting patterns with model predictions. Panel A plots averages of level of amenities of teachers' matched schools based on equally spaced bins of the distribution of teachers' test scores both in the actual data and in the simulated equilibrium. Panel B shows the result of a similar exercise with wages instead.

Figure 3.A.4: Validation RDD: Eligibility Cutoff



Notes. This figure displays the effect of crossing the test score threshold determining eligibility to permanent contracts on the probability to choose a permanent contract, the distance between teachers' matched schools and their home location, and the wage received from their matched schools. It computes these threshold crossing effects both in the actual data and in the equilibrium match simulated by the model and compares them.

3.B Context & Data: Details

3.B.1 Additional Institutional Details

Contracts and Wages

Public teachers in Peru can be hired under two types of contract. *Temporary* contracts last at least one year and can be renewed up to a second year, if both the school and the teacher agree. After two years, the position is either destroyed, if the allocated budget was fixed, or proposed again on the labor market. The same teacher could eventually teach in the same position but would have to apply again to get hired. *Permanent* contracts can last indefinitely. The coexistence of these two types of contracts is a common feature of civil servants' labor markets around the world. Permanent contracts are akin to usual civil servants contracts which make the profession attractive by insuring teacherss against unemployment. Temporary contracts are more precarious and are usually meant for schools to get a flexible access to a larger pool of applicants and react to unexpected transfers and/or the creation of new classrooms.

Wages are set by the government at the country level and vary along several dimensions. Temporary contracts are paid a fixed rate which does not vary with experience. To make the profession more attractive and keep up with inflation, the base monthly wage increased gradually from S/1,396 (363.19$) in 2016 to S/2,000 (520.33$) in 2017 and S/2,200 (572.36$) in 2019 to finally reach S/2,400 (624.39$) in 2021. Regarding permanent contracts, the pay scale is divided in six categories and teachers can apply once a year for a promotion through a centralized platform.[25] At the highest scale, the wage is 75% higher than the starting wage. Note that, at the exception of 2016 where it was S/1,550 (403.25$), the starting wage is exactly similar to the base wage for temporary contracts and followed the same time trend.

A wage bonus scheme was implemented by the Ministry of Education in order to make schools located in distressed areas or with worse teaching conditions more attractive. Teachers handling several grades receive a monthly wage bonus of S/140, schools with a single teacher provide a bonus of S/200, schools located in guerilla zones (VRAEM) provide a bonus of S/300, schools located close to the country borders provide a bonus of S/100 and schools which teach in several languages provide a bonus of S/50. Finally a set of wage bonuses

[25] Promotions are awarded through a national standardized evaluation and a decentralized evaluation made by a committee which evaluates teachers' performance and professional career.

ranging from S/70 to S/500 based on arbitrary cutoff rules compensates teachers based on the remoteness of the school's locality. Bobba et al. (2022) use these threshold in a regression discontinuity design to estimate the causal impact of increasing wages on recruitment and student achievement.

Allocation Mechanism

To make the allocation process of teaching positions more transparent, the Ministry of Education switched from a decentralized to a centralized application system in 2015. The use of centralized clearinghouses to allocate public sector jobs is becoming increasingly common (Roth, 2018) as they allow to reduce search frictions by regrouping all offers and applicants on the same market. The allocation of both temporary and permanent contracts is organized sequentially between November and March. Note that once teachers get awarded a permanent contract, they need to go through a separate procedure in order to be transferred to another school.[26]

National Competency Test: Before teaching positions are allocated, all applicants take a test evaluating their teaching competency. They get graded on three skills: (i) reading comprehension, (ii) logic reasoning and (iii) curricular knowledge. To be eligible for a permanent position, a teacher should get a score of at least 30/50 in part (i) and (ii) of the test and a score of a least 60/100 in part (iii) of the test. These are stringent requirements since only 9% of applicants end up being eligible (see Table 3.A.1).

Allocation of permanent positions: The Ministry first publishes the list of available positions. Teachers eligible for a permanent position then form a list of choices within the same province.[27] Applicants are then assigned for interviews to their preferred three schools, with a total of 10 available slots per school.[28] For schools that are oversubscribed, test scores are used as priorities. Schools then interview and rank each applicant. Finally, they make offers sequentially to their preferred applicants. All unassigned applicants can then participate to an exceptional stage that allocates the remaining unfilled slots. At the end of this round, unassigned teachers can decide to participate in the allocation of temporary positions which takes place shortly after.

[26]Transfers are handled every year in a decentralized way. Priority in the transfer application system depends on seniority and other criterias which are not made public by the Ministry of Education.

[27]The maximum length of the list went from five in 2015 to being unrestricted from 2017 onwards.

[28]In 2015, applicants were assigned to two schools maximum and there were 20 slots per school.

Allocation of temporary positions: All ineligible applicants along with eligible applicants which did not choose a permanent position participate in the allocation of temporary contracts. Teachers choose first a province. Within each province, serial dictatorship is used to assign teachers to schools using test scores as priorities. Schools do not have any role in the allocation process and cannot express their preferences over applicants. As in the allocation of permanent positions, unfilled vacancies are proposed to unassigned teachers from a different province in an exceptional stage.

This mechanism took place every year from 2015 to 2021 except in 2016. Note that in 2018 and 2020, only permanent positions were proposed.

3.B.2 Data Construction

I combine several sources of data provided by the Ministry of Education in Peru to construct the teacher assignment data and centralized assignment data described in Section 3.2.

Teacher occupation and payroll system (NEXUS): This dataset records annually each teacher and its matched position over the period 2012-2021. I restrict the data to primary school teachers which hold either a permanent or temporary contract. I exclude teachers working in several jobs by acting as a temporary replacement for other teachers on leave. Each teacher and position are identified by a unique ID which can be linked to other data sources. Each position is linked to the corresponding school which is also identified by a unique ID.

School census: This dataset contains information on a wide range of schools' and localities' characteristics. I observe detailed information on access to a wide range of services at the locality level such as electricity, water, sewage, medical centers, libraries or internet. I also observe the travel time between the locality and the closest provincial capital. I observe the number of inhabitants in the locality. I know whether the school has a second language of instruction, whether it has a single classroom. I also have access to the precise geocoordinates of the locality.

Household Targeting System (SISFOH): This dataset comes from Bobba et al. (2022) and contains information on the socio-economic status of the population of Peru in order to better target social benefits. It regroups individuals into households and records their home location, highest level of education, gender and their poverty status. I also observe their role with respect to the head of the household meaning that I can identify if individuals have children, are married or live with their parents.

Survey Centralized Allocation: The Ministry of Education surveys all the applicants that participate in the centralized allocation mechanism. I have thus additional information about applicants' level of experience in the public and private sector. I know which languages they speak and in which university or institute they went.

Centralized Allocation Mechanism: This dataset contains all the details of each step of the centralized allocation mechanism over the period 2015-2019. I observe the results of the national competency test for each applicant. I observe the set of applicants and positions participating in the allocation of permanent positions. I know where teachers apply, which schools interview them, how schools rank them and the final match. Finally, I observed the set of applicants and positions participating in the allocation of temporary positions. I do not observed teachers' final decision but I infer their match using the teacher assignment data.

The teacher assignment data combines the NEXUS with the school census and the SISFOH. The centralized assignment data combines the centralized allocation mechanism with the survey, the SISFOH and the school census.

3.C Value Added Model

I use data on the national evaluation of students in 2nd and 4th grade. I observe standardized test scores in math and in Spanish and I can match each classroom to its corresponding teacher. I can also match students to the SISFOH data to recover parental characteristics such as their education level or their poverty status.

Following closely Chetty et al. (2014a), I assume that each student i in year t is assigned to classroom $c = c(i,t)$ and that each teacher $j(c)$ is assigned to a classroom c. I restrict the analysis to primary schools meaning that teachers only teach one class per year. I denote μ_{jt} the value added of teacher j in year t normalized to have mean 0 and measured in student test scores standard deviations. I allow value added to drift over time. Finally, I assume that student i's test score in year t A^*_{it} relates to value added in the following way:

$$A^*_{it} = \boldsymbol{X}'_{it}\boldsymbol{\beta} + \mu_{jt} + \theta_c + \epsilon_{it}$$

where $\boldsymbol{X}_{it}$ includes a set of student, classroom and school characteristics, θ_c is an exogenous shock at the classroom level and ϵ_{it} is an idiosyncratic shock at the student-year level. I assume that the stochastic processes μ_{jt} and ϵ_{it} are stationary meaning that $\mathbb{E}[\mu_{jt}|t] = \mathbb{E}[\epsilon_{it}|t] = 0$, $\text{Cov}(\mu_{jt}, \mu_{jt+s}) = \sigma_{\mu s}$, $\text{Cov}(\epsilon_{it}, \epsilon_{it+s}) = \sigma_{\epsilon s}$ and $\text{Var}(\mu_{jt}) = \sigma^2_\mu$ for all t.

I estimate μ_{jt} using the following procedure. First, I estimate $\boldsymbol{\beta}$ by regressing test scores A^*_{it} on $\boldsymbol{X}_{it}$ and teacher fixed effects α_j. Estimating $\boldsymbol{\beta}$ using within-teacher variation avoids attributing the teacher effect to variation in $\boldsymbol{X}_{it}$.[29] I then construct the following residualized test scores:

$$A_{it} = A^*_{it} - \boldsymbol{X}'_{it}\hat{\boldsymbol{\beta}}$$

and average them at the teacher-year level to construct $\overline{A}_{jt} = \frac{1}{n}\sum_{i\in\{i:j=j(c(i,t))\}} A_{it}$ for all j,t. Finally, I shrink these estimates by projecting $\overline{A}_{jt}$ on past residualized test scores $\boldsymbol{A}^{-t}_j = (\overline{A}_{j1}, ..., \overline{A}_{jt-1})$. The estimator of VA $\hat{\mu}_{jt}$ can thus be written as:

$$\hat{\mu}_{jt} = \sum_{s=1}^{t-1} \psi_s \overline{A}_{js}$$

where $\boldsymbol{\psi} = (\psi_1, ..., \psi_{t-1})$ are the coefficients of the OLS regression of $\overline{A}_{jt}$ on $\boldsymbol{A}^{-t}_j$. Note that

[29] Table 3.C.1 shows that not including teacher FE introduces severe bias in the coefficients associated with classroom and school characteristics.

I only use $t-1$ to predict VA such that $\psi = \frac{\sigma_{A,1}}{\sigma_A^2} = \frac{\text{Cov}(\overline{A}_{jt}, \overline{A}_{jt-1})}{\text{Var}(\overline{A}_{jt})}$.

The results of the estimation of $\hat{\mu}_{jt}$ are displayed in Table 3.C.2. The auto-correlation ψ is estimated at 0.466. To get a proper estimate for the standard deviation of value added $\sigma_\mu = \sigma_{A,0}$ in elementary schools, Chetty et al. (2014a) perform a non-linear extrapolation from their estimates of $\sigma_{A,s}$ for $1 \leq s \leq 7$. However, I do not have access to test score data prior to 2016 making the replication of this exercise impossible. As pointed out in Chetty et al. (2014a), $\sigma_{A,0} \geq \sigma_{A,1}$ making $\sqrt{\hat{\sigma}_{A,1}}$ an estimator of a lower bound on σ_μ. I estimate this lower bound to be 0.3 which is substantially larger than previous estimates.[30]

I then perform the usual checks for forecast unbiasedness of $\hat{\mu}_{jt}$ following Chetty et al. (2014a). I first regress A_{it} on $\hat{\mu}_{jt}$ and find a coefficient of 1.030 with 95% confidence interval $[0.944, 1.116]$. Standard errors are clustered at the school level. This regression should give us a coefficient of 1 which is not rejected by the data. I then project A_{it} on parental characteristics that are excluded from X_{it} such as socio-economic status and regress μ_{jt} on this projection. I find a coefficient of 0.008 with a tight 95% confidence interval $[0.002, 0.014]$ meaning that we can rule out any substantial sorting of students across teachers based on parental characteristics.[31]

To estimate the cost of attrition, I use teacher switching as a quasi-experiment as in Chetty et al. (2014a) to test two hypotheses: (i) skills are not perfectly transferable across schools and (ii) attrition implies a net loss for the origin school. If skills are perfectly transferable across schools, switching to a different school after a long employment spell should have no effect on value added. I assume that switching decisions are independent of unobserved factors that could affect drift in value added. This rules out scenarios where teachers decide to switch to a different school because they anticipate that they will have a higher value added there. I then compare the difference in value added between 2016 and 2018 for teachers that stayed in the same school with the same difference for teachers that moved to a different school. To do so, I estimate β in the following two-way fixed effects regression:

$$\mu_{gt} = \alpha_g + \delta_t + \beta D_{gt} + \epsilon_{gt} \tag{3.5}$$

[30] Chetty et al. (2014a) estimate $\sigma_\mu = 0.163$ and Bates et al. (2022) estimate $\sigma_\mu = 0.249$. This could be explained by the fact that most other studies use data on urban districts while the data used in this paper covers the universe of teachers in Peru. If high value added teachers are concentrated in cities, estimating σ_μ in urban districts could understate its population value.

[31] I explore the relationship between $\hat{\mu}_{jt}$ and A_{it} as well as between $\hat{\mu}_{jt}$ and predicted scores using parental SES non parametrically in Figure 3.C.1. To do this, I construct averages for 20 equal sized bins of value added to get an approximation of the conditional expectation function.

where $\mu_{gt} = \frac{1}{N_{gt}} \sum_{j \in g} \hat{\mu}_{jt}$, α_g is a group fixed effect, δ_t is a time fixed effect and D_{gt} is group g's treatment status in period t. In this simple setting $g \in \{\text{Movers}, \text{Stayers}\}$ and $t \in \{2016, 2018\}$ and I assume that $D_{\text{Stayers},t} = 0$ for all t and $D_{\text{Movers},2016} = 0$ and $D_{\text{Movers},2018} = 1$. In this setting, β corresponds to the ATT.

Table 3.C.3 shows the results of the estimation of β with standard errors clustered at the teacher level. In Panel A, I estimate β conditional on movers having more than one year of experience in the school they taught in before switching. I find that moving implies a net loss of value added of 0.056σ corresponding to 26% of a standard deviation of teacher value added. This is consistent with the hypothesis that skills are not perfectly transferable across schools. As a placebo test, I consider movers with no prior experience in the schools they were before switching in Panel B. They should not have accumulated school specific skills prior to moving which is consistent with the zero effect found in Table 3.C.3. These results show that job-to-job transitions imply a sizeable aggregate loss in value added.

I then perform a second exercise quantifying the loss in productivity following a move at the school level. I find that leavers are substantially of higher quality than the teachers who replace them. Using value added prior to moving I find a difference of 0.10 standard deviation which corresponds to around 50% of a standard deviation in value added.

Table 3.C.1: Value Added: Estimation of $\boldsymbol{\beta}$

	(1)	(2)
Constant	0.781 (0.047)	0.559 (0.133)
$t = 2018$	0.020 (0.005)	0.023 (0.005)
Student Level Controls		
Lagged Math Score	0.431 (0.006)	0.444 (0.005)
Lagged Math Score^2	0.010 (0.002)	0.012 (0.002)
Lagged Math Score^3	-0.027 (0.002)	-0.029 (0.001)
Lagged Spanish Score	0.236 (0.005)	0.444 (0.005)
Lagged Spanish Score^2	0.010 (0.002)	0.012 (0.002)
Lagged Spanish Score^3	-0.012 (0.001)	-0.029 (0.001)
Female	-0.112 (0.005)	-0.111 (0.005)
Age	-0.071 (0.005)	-0.052 (0.005)
Ethnicity: Quechua	0.045 (0.027)	0.058 (0.025)
Ethnicity: Native	-0.022 (0.026)	-0.036 (0.023)
Classroom Level Controls		
Ethnicity: Quechua	1.657 (0.094)	0.379 (0.160)
Ethnicity: Native	-1.303 (0.093)	-0.478 (0.151)
Size	0.003 (0.000)	-0.001 (0.001)
School Level Controls		
Lagged Math Score	-0.052 (0.018)	-0.384 (0.073)
Lagged Math Score^2	0.004 (0.016)	-0.213 (0.058)
Lagged Math Score^3	0.074 (0.016)	0.031 (0.045)
Lagged Spanish Score	0.245 (0.019)	0.483 (0.073)
Lagged Spanish Score^2	-0.047 (0.015)	0.140 (0.057)
Lagged Spanish Score^3	-0.001 (0.012)	0.038 (0.037)
Teacher FE	✗	✓

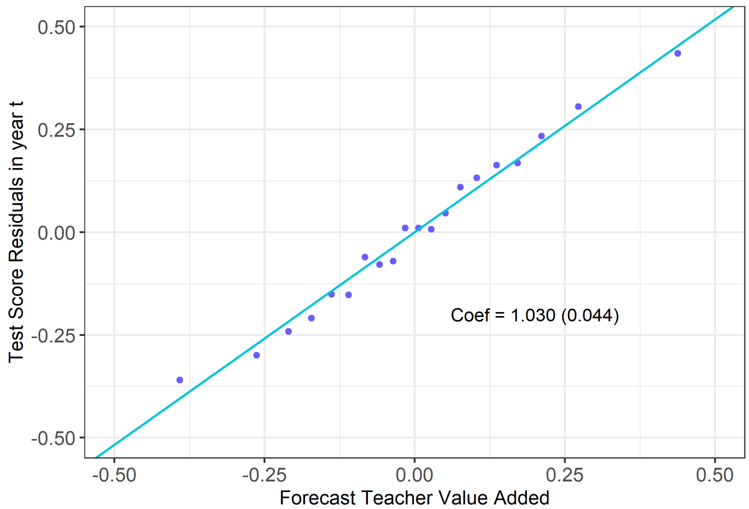
NOTES. This table displays the estimates of $\boldsymbol{\beta}$ from the linear regression of student test scores on student, classroom and school characteristics described in Section 3.D. Column 1 shows the results of this regression without teacher fixed effects. Column 2 includes teacher fixed effects. Standard errors are in parentheses

Table 3.C.2: Value Added: Structural Parameters

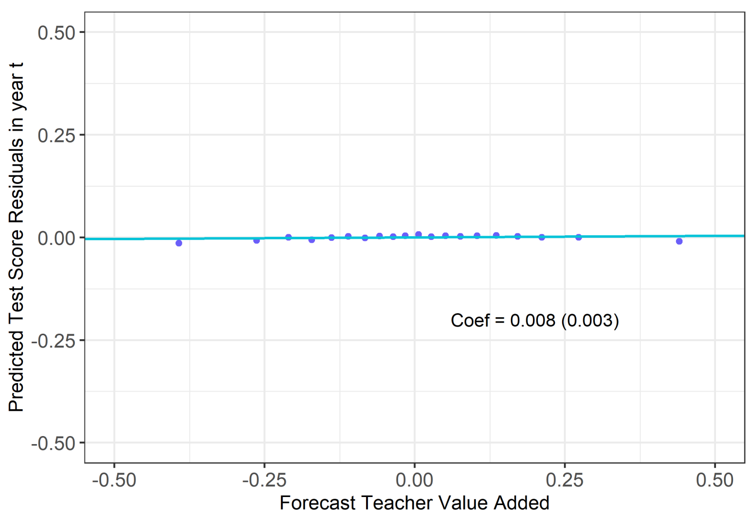
Parameter	Estimate	Std. Error	95% CI
$\sigma_{A,1}$	0.089	0.005	$[0.079, 0.100]$
σ_A	0.192	0.007	$[0.179, 0.205]$
ψ	0.466	0.019	$[0.446, 0.485]$
Lower Bound σ_μ	0.300	0.009	$[0.282, 0.316]$

NOTES. This Table displays the estimates of the structural parameters of the teacher value added model described in Section 3.D.

Figure 3.C.1: Value Added: Robustness Checks



a) Actual Score



b) Predicted Score using Parental SES

Notes. Panel A of this figure plot averages of the test score residuals A_{it} for 20 equally spaced bins of the forecasted teacher value added. Panel B of plot averages of the test score residuals A_{it} projected onto parental socio-economic status for 20 equally spaced bins of the forecasted teacher value added. The reported coefficients correspond to the slope of the blue line. Standard errors are in parentheses.

Table 3.C.3: Imperfectly Transferable Skills

	Estimate	Std. Errors	95% CI
Panel A: Past Tenure			
ATE Movers: β	-0.056	0.023	$[-0.102, -0.011]$
α_{Stayers}	-0.005	0.004	$[-0.014, 0.003]$
α_{Movers}	-0.026	0.020	$[-0.066, 0.024]$
δ_{2018}	0.004	0.005	$[-0.005, 0.013]$
Panel B: No Past Tenure			
ATE Movers: β	0.004	0.018	$[-0.031, 0.038]$
α_{Stayers}	-0.004	0.004	$[-0.012, 0.005]$
α_{Movers}	-0.021	0.014	$[-0.048, 0.007]$
δ_{2018}	0.004	0.005	$[-0.006, 0.013]$

NOTES. This table displays the results of the estimation of Equation (3.5). Panel A restricts the sample to teachers which have been in the same school prior to 2016 for more than three years. Panel B restricts the sample to teachers which have been in the same school prior to 2016 for less than three years. Standard errors are clustered at the teacher level.

3.D Proofs

3.D.1 Proof of Proposition 1

I first show that part (i) of Proposition 1 is a direct implication of Assumption 4 (i) and (ii), i.e that the match is stable in period t.

Consider a match μ_t and suppose first that either Assumption 4 (i) or (ii) is violated such that μ_t is not stable. First, suppose that (i) does not hold meaning that there exists a teacher-school pair (i, j) such that $U_{ijt} > U_{i\mu_{wt}(i)t}$ and $V_{ijt} > V_{\mu_{mt}(j)jt}$. This would mean that $j \in M_{it}(\mu_t)$ and $U_{ijt} > U_{i\mu_{wt}(i)t}$ which contradicts that $U_{i\mu_{wt}(i)t} = \max_{k \in M_{it}(\mu_t) \cup \{0\}} U_{ikt}$. Now, suppose that (ii) does not hold meaning that $U_{i0t} > U_{i\mu_{wt}(i)t}$ or $V_{0jt} > V_{\mu_{mt}(j)jt}$. In both cases, this would contradict that $U_{i\mu_{wt}(i)t} = \max_{k \in M_{it}(\mu_t) \cup \{0\}} U_{ik}$ or $V_{\mu_{mt}(j)jt} = \max_{l \in W_{jt}(\mu_t) \cup \{0\}} V_{ljt}$.

Now, suppose that for a given i, $U_{i\mu_{wt}(i)t} < \max_{k \in M_{it}(\mu_t) \cup \{0\}} U_{ikt}$. This means that there exists a school $k' \in M_{it}(\mu) \cup \{0\}$ such that $U_{ik't} > U_{i\mu_{wt}(i)t}$. If $k' = 0$ this immediately contradicts stability. If $k' \in M_{it}(\mu)$ this implies that $V_{ik't} \geq V_{\mu_{mt}(k')k't}$ and $U_{ik't} > U_{i\mu_{wt}(i)t}$. If $V_{ik't} = V_{\mu_{mt}(k')k't}$ this implies that $k' = \mu_w(i)$ and we reach a contradiction. Otherwise we have that $U_{ik't} > U_{i\mu_{wt}(i)t}$ and $V_{ik't} > V_{\mu_m(k')k't}$ which contradicts stability. The argument is symmetric for the school's side.

Part (ii) of Proposition 1 is a direct consequence of part (i) and Assumption 2.

3.D.2 Proof of Proposition 2

As $\overline{U}_{t+1}$ is independent of η_{ijt} under Assumption 1 and with exogenous choice sets, I treat it as fixed and rewrite $U_{ijt} = u_{ijt} + \sigma\eta_{ijt}$ for simplicity. The proof of part (i) of Proposition 2 is then identical to the proof of Lemma 3.1 in Menzel (2015).

$$
\begin{aligned}
\mathbb{P}(U_{ijt} \geq \max_{k=0,1,\dots,J} U_{ikt} | (u_{ikt})_{k=1}^{J}) &= \int \mathbb{P}(U_{ijt} \geq U_{ikt}, k \in \mathcal{I} - \{j\} | (u_{ikt})_{k=1}^{J}, \eta_{ijt} = s) g(s) ds \\
&= \int \prod_{k \in \mathcal{I} - \{j\}} G(\sigma^{-1}(u_{ijt} - u_{ikt}) + s) g(s) ds \\
&= \int \prod_{k=1}^{2J} G(\sigma^{-1}(u_{ijt} - u_{ikt}) + s) \frac{g(s)}{G(s)} ds
\end{aligned}
$$

As in Menzel (2015), I then do the change of variables $s = a_J h + b_J$ where $a_J = a(b_J)$ and $b_J = G^{-1}(1 - J^{-1/2})$ and multiply by J on both sides:

$$J\mathbb{P}(U_{ijt} \geq \max_{k=0,1,\ldots,J} U_{ikt} | (u_{ikt})_{k=1}^{J}) = \int \exp\left(\frac{1}{J}\sum_{k=1}^{2J} J \log G(a_J(u_{ijt} - u_{ikt} + h) + b_J)\right) \frac{J a_J g(a_J h + b_J)}{G(a_J h + b_J)} dh$$

Following Resnick (1987) and under Assumption 1 we can show that:

$$J \log G(a_J(u_{ijt} - u_{ilt} + h) + b_J) \rightarrow -e^{-(u_{ijt} - u_{ikt} + h)}$$

$$\frac{J a_J g(a_J h + b_J)}{G(a_J h + b_J)} \rightarrow e^{-h}$$

We thus have under Assumption 1:

$$\begin{aligned} J\mathbb{P}(U_{ijt} \geq \max_{k=0,1,\ldots,J} U_{ikt} | (u_{ikt})_{k=1}^{J}) &= \int \exp\left(-\frac{1}{J}\sum_{k=1}^{2J} e^{-(u_{ijt} - u_{ikt} + h)}\right) e^{-h} dh + o(1) \\ &= \int \exp\left(-\frac{1}{J}\sum_{k=1}^{2J} e^{-h} e^{(u_{ikt} - u_{ijt})}\right) e^{-h} e^{-h} dh + o(1) \end{aligned}$$

I then do a final change of variable $s = e^{-h}$ such that we get:

$$\begin{aligned} J\mathbb{P}(U_{ijt} \geq \max_{k=0,1,\ldots,J} U_{ikt} | (u_{ikt})_{k=1}^{J}) &= \int_0^{+\infty} \exp\left(-\frac{1}{J}\sum_{k=1}^{2J} s e^{(u_{ikt} - u_{ijt})}\right) s ds + o(1) \\ &= \frac{\exp(u_{ijt})}{\frac{1}{J}\sum_{k=1}^{2J} \exp(u_{ikt})} + o(1) \end{aligned}$$

From this we can finally show that:

$$J\mathbb{P}(U_{ijt} \geq \max_{k=0,1,\ldots,J} U_{ikt} | \boldsymbol{x}_{it}, (\boldsymbol{z}_{kt})_{k=1}^{J}) =$$

$$\frac{\exp\left\{U_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{jt}) + \beta_w \int \overline{U}_{t+1}(s) w(s|\boldsymbol{x}_{it}, \boldsymbol{z}_{jt}) ds\right\}}{\exp\left\{\beta_w \int \overline{U}_{t+1}(s) w_0(s|\boldsymbol{x}_{it}) ds\right\} + \frac{1}{J}\sum_{k=1}^{J} \exp\left\{U_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) + \beta_w \int \overline{U}_{t+1}(s) w(s|\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) ds\right\}} + o(1)$$

which implies that:

$$J\mathbb{P}(U_{ijt} \geq \max_{k=0,1,\ldots,J} U_{ikt} | \boldsymbol{x}_{it}, \boldsymbol{z}_{jt}) \longrightarrow$$

$$\frac{\exp\left\{U_t(\boldsymbol{x}_{it},\boldsymbol{z}_{jt})+\beta_w\int\overline{U}_{t+1}(s)w(s|\boldsymbol{x}_{it},\boldsymbol{z}_{jt})ds\right\}}{\exp\left\{\beta_w\int\overline{U}_{t+1}(s)w_0(s|\boldsymbol{x}_{it})ds\right\}+\int\exp\left\{U_t(\boldsymbol{x}_{it},h)+\beta_w\int\overline{U}_{t+1}(s)w(s|\boldsymbol{x}_{it},h)ds\right\}m_t(h)dh}$$

which finishes the proof of part (i) of Proposition 2.

Using similar steps as in McFadden et al. (1973), we can then show that:

$$\mathbb{E}(\max_{k=0,1,\ldots,J}U_{ikt}|\boldsymbol{x}_{it},(\boldsymbol{z}_{kt})_{k=1}^J)=\log\left(\exp\left\{\beta_w\int\overline{U}_{t+1}(s)w_0(s|\boldsymbol{x}_{it})ds\right\}\right.$$
$$\left.+\frac{1}{J}\sum_{k=1}^J\exp\left\{U_t(\boldsymbol{x}_{it},\boldsymbol{z}_{kt})+\beta_w\int\overline{U}_{t+1}(s)w(s|\boldsymbol{x}_{it},\boldsymbol{z}_{kt})ds\right\}\right)+\log(J)+\gamma+o(1)$$

where γ is Euler's constant. Under Assumption 1, we can finally apply the law of large numbers to show that:

$$\mathbb{E}(\max_{k=0,1,\ldots,J}U_{ikt}|\boldsymbol{x}_{it},(\boldsymbol{z}_{kt})_{k=1}^J)=\log\left(\exp\left\{\beta_w\int\overline{U}_{t+1}(s)w_0(s|\boldsymbol{x}_{it})ds\right\}\right.$$
$$\left.+\int\exp\left\{U_t(\boldsymbol{x}_{it},h)+\beta_w\int\overline{U}_{t+1}(s)w(s|\boldsymbol{x}_{it},h)ds\right\}m_t(h)dh\right)+\log(J)+\gamma+o(1)$$

which concludes the proof of part (ii) of Proposition 2.

3.D.3 Definition Ψ_{wt} and Ψ_{mt}

$$\Psi_{wt}[\mathbf{\Gamma}](x)=\int\frac{\exp\left\{U_t(x,h)+V_t(x,h)+\beta\int\overline{U}_{t+1}[\mathbf{\Gamma}](s)w(s|x,h)ds+\beta\int\overline{V}_{t+1}[\mathbf{\Gamma}](s)m(s|x,h)ds\right\}}{\exp\left\{\beta\int\overline{V}_{t+1}[\mathbf{\Gamma}](s)m_0(s|h)ds\right\}+\Gamma_{mt}(h)}m_t[\mathbf{\Gamma}](h)dh$$

$$\Psi_{mt}[\mathbf{\Gamma}](z)=\int\frac{\exp\left\{U_t(h,z)+V_t(h,z)+\beta\int\overline{U}_{t+1}[\mathbf{\Gamma}](s)w(s|h,z)ds+\beta\int\overline{V}_{t+1}[\mathbf{\Gamma}](s)m(s|h,z)ds\right\}}{\exp\left\{\beta\int\overline{U}_{t+1}[\mathbf{\Gamma}](s)w_0(s|h)ds\right\}+\Gamma_{wt}(h)}w_t[\mathbf{\Gamma}](h)dh$$

$$\overline{U}_{t+1}[\mathbf{\Gamma}](x)=\log\left(\exp\left\{\beta\int\overline{U}_{t+2}[\mathbf{\Gamma}](s)w_0(s|x)ds\right\}+\Gamma_{wt+1}(x)\right)$$

$$\overline{V}_{t+1}[\mathbf{\Gamma}](z)=\log\left(\exp\left\{\beta\int\overline{V}_{t+2}[\mathbf{\Gamma}](s)m_0(s|z)ds\right\}+\Gamma_{mt+1}(z)\right)$$

$$w_t[\mathbf{\Gamma}](x)=\int_{\mathcal{X}_t}\int_{\mathcal{Z}_t}w(x|s,h)f_{t-1}[\mathbf{\Gamma}](s,h)dhds+\int_{\mathcal{X}_t}w_0(x|s)f_{t-1}[\mathbf{\Gamma}](s,*)ds$$

$$m_t[\mathbf{\Gamma}](z) = \int_{\mathcal{X}_t}\int_{\mathcal{Z}_t} m(z|s,h) f_{t-1}[\mathbf{\Gamma}](s,h) dh ds + \int_{\mathcal{Z}_t} m_0(z|s) f_{t-1}[\mathbf{\Gamma}](*,h) dh$$

3.D.4 Proof of Theorem 1

I start by proving part (i) and (ii) of Theorem 1. Throughout the rest of the proof I set WLOG $\gamma_w = \gamma_m = 0$ and $\beta_w = \beta_m = \beta$ for simplicity. I first restrict the space of functions to which the solutions to the fixed point problem described in Equation (3.4) can belong. Namely, I show that we can restrict ourselves to a Banach space of continuous functions. Assume that there exists a set of $2 \times T$ functions Γ^*_{wt} and Γ^*_{mt} for all $t = 1, ..., T$ that solve the fixed point problem. I start by showing that these solutions are bounded from above. By definition of Ψ_{wt} and $\overline{U}_t$ and using that $\Gamma^*_{mt} \geq 0$ for all t, we can proceed by backward induction and show:

$$\begin{aligned}
\Gamma^*_{wT}(x) = \Psi_{wT}[\mathbf{\Gamma}^*](x) &= \int \frac{\exp\{U_T(x,s) + V_T(x,s)\}}{1 + \Gamma^*_{mT}(s)} m_T(s) ds \\
&\leq \int \exp\{U_T(x,s) + V_T(x,s)\} m_T(s) ds \\
&\leq \exp\{\bar{U} + \bar{V}\}
\end{aligned}$$

where $\bar{U}$ and $\bar{V}$ are the upper bounds of the functions U_t and V_t for all t, respectively. From there we can bound $\overline{U}^*_T$ as follows:

$$\begin{aligned}
\overline{U}^*_T(x) &= \log\left(1 + \Gamma^*_{wT}(x)\right) \\
&\leq \log\left(1 + \exp\{\bar{U} + \bar{V}\}\right)
\end{aligned}$$

Similar bounds can be derived on the school side. We can then iterate this procedure and bound Γ_{wT-1} and $\overline{U}_{T-1}$:

$$\begin{aligned}
\Gamma^*_{wT-1}(x) &= \int \frac{\exp\{U_{T-1}(x,h) + V_{T-1}(x,h) + \beta \int \overline{U}_T(s) w(s|x,h) ds + \beta \int \overline{V}_T(s) m(s|x,h) ds\}}{1 + \Gamma^*_{mT-1}(h)} m_{T-1}(h) dh \\
&\leq \exp\left\{\overline{U} + \overline{V} + 2\beta \log(1 + \exp\{\overline{U} + \overline{V}\})\right\}
\end{aligned}$$

$$\overline{U}^*_{T-1}(x) = \log\left(\exp\left\{\beta\int \overline{U}^*_T(s)w_0(s|x)ds\right\} + \Gamma^*_{wT-1}(x)\right)$$
$$\leq \log\left(\exp\left\{\beta\log\left(1+\exp\{\overline{U}+\overline{V}\}\right)\right\} + \exp\left\{\overline{U}+\overline{V}+2\beta\log\left(1+\exp\{\overline{U}+\overline{V}\}\right)\right\}\right)$$

Boundedness of Γ^*_{wt} is thus implied by boundedness of $\overline{U}_{t+1}$ which is in itself implied by boundedness of Γ^*_{wt+1} and $\overline{U}_{t+2}$. By induction we can thus show that boundedness of Γ^*_{wT} implies boundedness of Γ^*_{wt} for all $t = 1, ..., T$. The same argument applies to Γ^*_{mt}. Continuity of Γ^*_{wt} and Γ^*_{mt} follows from continuity of U and V and that the integrals are nonnegative. Differentiability of Γ^*_{wt} and Γ^*_{mt} also follows from differentiability of U and V which is stated in Assumption 1. We can thus restrict the spaces in which Γ^*_{wt} and Γ^*_{mt} belong to a Banach space of nonnegative bounded continuous functions which I call $\mathcal{C}$.

I now turn to the proof that the mapping $(\log \boldsymbol{\Gamma_w}, \log \boldsymbol{\Gamma_m}) \mapsto (\log \boldsymbol{\Psi_m}[\boldsymbol{\Gamma}], \log \boldsymbol{\Psi_w}[\boldsymbol{\Gamma}])$ is a contraction. Consider two sets of functions $\boldsymbol{\Gamma} = (\Gamma_{mt}, \Gamma_{wt})_{t=1}^T$ and $\tilde{\boldsymbol{\Gamma}} = (\tilde{\Gamma}_{mt}, \tilde{\Gamma}_{wt})_{t=1}^T$ belonging to $\mathcal{C}^{2T}$. I show that there always exists a constant $\lambda < 1$ such that:

$$\Big|\Big|\log \boldsymbol{\Psi_w}[\boldsymbol{\Gamma_m}] - \log \boldsymbol{\Psi_w}[\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}}]\Big|\Big|_\infty \leq \lambda \Big|\Big|\log \boldsymbol{\Gamma_m} - \log \tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}}\Big|\Big|_\infty$$

The mean value inequality for vector valued functions defined on Banach spaces implies that:

$$\Big|\Big|\log \boldsymbol{\Psi_w}[\boldsymbol{\Gamma_m}](x) - \log \boldsymbol{\Psi_w}[\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}}](x)\Big|\Big|_\infty \leq$$
$$\sup_{a\in[0,1]}\Big|\Big|D\log \boldsymbol{\Psi_w}[a\log \boldsymbol{\Gamma_m} + (1-a)\log \tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}}](x)\Big|\Big|_\infty \Big|\Big|\log \boldsymbol{\Gamma_m}(x) - \log \tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}}(x)\Big|\Big|_\infty$$

where $D\log \boldsymbol{\Psi_w}$ are the Gateaux derivatives of $\log \boldsymbol{\Psi_w}$. The rest of the proof consists in showing that these derivatives are strictly bounded below 1.

Starting with $t = 1$, I rewrite $\log \Psi_{w1}$ such that:

$$\log \Psi_{w1}[\log \boldsymbol{\Gamma}](x) =$$
$$\log \int \frac{\exp\left\{U_1(x,h) + V_1(x,h) + \beta\int \overline{U}_2[\log \boldsymbol{\Gamma}](s)w(s|x,h)ds + \beta\int \overline{V}_2[\log \boldsymbol{\Gamma}](s)m(s|x,h)ds\right\}}{\exp\{\beta\int \overline{V}_2[\log \boldsymbol{\Gamma}](s)m_0(s|h)ds\} + \exp\{\log \Gamma^*_{m1}(h)\}} m_1(h)dh$$

where $\overline{U}_2$ and $\overline{V}_2$ are defined as:

$$\overline{U}_2[\log \mathbf{\Gamma}](x) = \log\left(\exp\left\{\beta \int \overline{U}_3[\log \mathbf{\Gamma}](s) w_0(s|x) ds\right\} + \exp\{\log \Gamma_{w2}(x)\}\right)$$

$$\overline{V}_2[\log \mathbf{\Gamma}](z) = \log\left(\exp\left\{\beta \int \overline{V}_3[\log \mathbf{\Gamma}](s) m_0(s|z) ds\right\} + \exp\{\log \Gamma_{m2}(z)\}\right)$$

The Gateaux derivative of $\log \Psi_{w1}$ with respect to $\log \Gamma_{m1}$ can be bounded in absolute value as:

$$\begin{aligned}
&\Bigg| -\frac{1}{\Psi_{w1}[\mathbf{\Gamma}](x)} \int \frac{\Gamma_{m1}(h)}{\exp\{\beta \int \overline{V}_2[\mathbf{\Gamma}](s) m_0(s|h) ds\} + \Gamma_{m1}(h)} \times \\
&\quad \frac{\exp\left\{U_1(x,h) + V_1(x,h) + \beta \int \overline{U}_2[\mathbf{\Gamma}](s) w(s|x,h) ds + \beta \int \overline{V}_2[\mathbf{\Gamma}](s) m(s|x,h) ds\right\}}{\exp\{\beta \int \overline{V}_2[\mathbf{\Gamma}](s) m_0(s|h) ds\} + \Gamma_{m1}(h)} m_1(h) dh \Bigg| \\
\leq \quad & \frac{\lambda_1}{\Psi_{w1}[\mathbf{\Gamma}](x)} \int \frac{\exp\left\{U_1(x,h) + V_1(x,h) + \beta \int \overline{U}_2(s) w(s|x,h) ds + \beta \int \overline{V}_2(s) m(s|x,h) ds\right\}}{\exp\{\beta \int \overline{V}_2(s) m_0(s|h) ds\} + \Gamma_{m1}(h)} m_1(h) dh \\
= \quad & \lambda_1
\end{aligned}$$

where λ_1 is an upper bound of the ratio

$$\frac{\Gamma^*_{m1}(h)}{\exp\{\beta \int \overline{V}_2[\log \mathbf{\Gamma}](s) m_0(s|h) ds\} + \Gamma^*_{m1}(h)} \leq \frac{\Gamma^U}{\exp\{\beta \overline{U}^U\} + \Gamma^U} = \lambda_1 < 1$$

A similar bound can be computed for the Gateaux derivative of $\log \Psi_{m1}$ with respect to $\log \Gamma_{w1}$.

I use a similar argument to show that the Gateaux derivative of $\log \Psi_{w1}$ with respect to $\log \Gamma_{mt}$ for $t > 1$ can be bounded in absolute value by the upper bound of the following expression:

$$\beta \int D_{mt}\overline{V_2}[\log \mathbf{\Gamma}](s) m(s|x,h) ds - \beta \int D_{mt}\overline{V_2}[\log \mathbf{\Gamma}](s) m_0(s|h) ds \frac{\exp\{\beta \int \overline{V}_2(s) m_0(s|h) ds\}}{\exp\{\beta \int \overline{V}_2(s) m_0(s|h) ds\} + \Gamma^*_{m1}(h)}$$

where I define $D_{mt}\overline{V_2}$ as the Gateaux derivative of $\overline{V_2}$ with respect to $\log \Gamma_{mt}$. From there, we can show that for all $1 < t < T$:

$$D_{mt}\overline{V_t}[\log \mathbf{\Gamma}](z) = \frac{\Gamma^*_{mt}(z)}{\exp\{\beta \int \overline{V}_{t+1}(s) m_0(s|z) ds\} + \Gamma^*_{mt}(z)} \leq \lambda_1 \tag{3.6}$$

From this result, we proceed by induction and show that for all $1 < t < T$:

$$\begin{aligned} D_{mt+1}\overline{V_t}[\log \mathbf{\Gamma}](z) = & \quad \beta \int D_{mt+1}\overline{V}_{t+1}[\log \mathbf{\Gamma}](s) m_0(s|z) ds \frac{\exp\{\beta \int \overline{V}_{t+1}(s) m_0(s|z) ds\}}{\exp\{\beta \int \overline{V}_{t+1}(s) m_0(s|z) ds\} + \Gamma^*_{mt}(z)} \\ \leq & \quad \beta\lambda_1 \frac{\exp\{\beta \overline{U}^U\}}{\exp\{\beta \overline{U}^U\} + \Gamma^U} = \beta\lambda_1\lambda_2 \end{aligned}$$

We can iterate this procedure to show that for all $1 < t \leq t' < T$:

$$D_{mt'}\overline{V_t}[\log \mathbf{\Gamma}](x) \leq \beta^{t'-t}\lambda_1\lambda_2^{t'-t} < 1 \tag{3.7}$$

For $t' = T$ we can easily verify that:

$$D_{mT}\overline{V_t}[\log \mathbf{\Gamma}](x) \leq \beta^{T-t}\lambda_1^2\lambda_2^{T-t-1} < 1$$

This implies that we can bound from above the first term of the derivative of $\log \Psi_{w1}$ with respect to $\log \Gamma_{mt}$ for all $1 < t < T$ by:

$$\beta^{t-1}\lambda_1\lambda_2^{t-2} < 1$$

while the second term can be bounded by:

$$\beta^{t-1}\lambda_1\lambda_2^{t-1} < 1$$

This implies that the difference between the two is strictly below 1. Similarly for $t = T$, we can bound the first term from above by

$$\beta^{T-1}\lambda_1^2\lambda_2^{T-3} < 1$$

while the second term can be bounded by:

$$\beta^{T-1}\lambda_1^2\lambda_2^{T-2} < 1$$

Again, this holds symetrically for Ψ_{m1}. This finishes to show that the Gateaux derivatives of $\log \Psi_{w1}$ and $\log \Psi_{m1}$ are strictly bounded below 1.

I now consider $\log \Psi_{wt}$ such that $1 < t < T$. I rewrite $\log \Psi_{wt}$ such that:

$$\log \Psi_{wt}[\log \mathbf{\Gamma}](x) =$$
$$\log \int \frac{\exp\left\{U_t(x,h) + V_t(x,h) + \beta \int \overline{U}_{t+1}[\log \mathbf{\Gamma}](s) w(s|x,h) ds + \beta \int \overline{V}_{t+1}[\log \mathbf{\Gamma}](s) m(s|x,h) ds\right\}}{\exp\{\beta \int \overline{V}_{t+1}[\log \mathbf{\Gamma}](s) m_0(s|h) ds\} + \exp\{\log \Gamma_{mt}(h)\}} m_t[\log \mathbf{\Gamma}](h) dh$$

where $\overline{U}_{t+1}$, $\overline{V}_{t+1}$ and m_t are defined as:

$$\overline{U}_{t+1}[\log \mathbf{\Gamma}](x) = \log \left(\exp\left\{ \beta \int \overline{U}_{t+2}[\log \mathbf{\Gamma}](s) w_0(s|x) ds \right\} + \exp\{\log \Gamma_{wt+1}(x)\} \right)$$

$$\overline{V}_{t+1}[\log \mathbf{\Gamma}](z) = \log \left(\exp\left\{ \beta \int \overline{V}_{t+2}[\log \mathbf{\Gamma}](s) m_0(s|z) ds \right\} + \exp\{\log \Gamma_{mt+1}(z)\} \right)$$

$$m_t[\log \mathbf{\Gamma}](z) = \int_{\mathcal{X}_{t-1}} \int_{\mathcal{Z}_{t-1}} m(z|s,h) f_{t-1}[\log \mathbf{\Gamma}](s,h) dh ds + \int_{\mathcal{Z}_{t-1}} m_0(z|s) f_{t-1}[\log \mathbf{\Gamma}](*,h) dh$$

and f_{t-1} can be expressed as follows:

$$f_{t-1}(x,z) = \frac{\exp\left\{ U_{t-1}(x,z) + V_{t-1}(x,z) + \beta \int \overline{U}_t(s) w(s|x,z) ds + \beta \int \overline{V}_t(s) m(s|x,z) ds \right\} w_{t-1}[\log \mathbf{\Gamma}](x) m_{t-1}[\log \mathbf{\Gamma}](z)}{\left(\exp\left\{ \beta \int \overline{U}_t(s) w_0(s|x) ds \right\} + \exp\{\log \Gamma_{wt-1}(x)\} \right) \left(\exp\left\{ \beta \int \overline{V}_t(s) m_0(s|z) ds \right\} + \exp\{\log \Gamma_{mt-1}(z)\} \right)}$$

$$f_{t-1}(*,z) = \frac{\exp\left\{ \beta \int \overline{V}_t(s) m_0(s|z) ds \right\}}{\left(\exp\left\{ \beta \int \overline{V}_t(s) m_0(s|z) ds \right\} + \exp\{\log \Gamma_{mt-1}(z)\} \right)} m_{t-1}[\log \mathbf{\Gamma}](z)$$

I first consider the derivative of $\log \Psi_{wt}$ with respect to $\log \Gamma_{mt-1}$ and write is as:

$$- \frac{1}{\Psi_{wt}[\mathbf{\Gamma}](x)} \int \frac{\exp\left\{U_t(x,h) + V_t(x,h) + \beta \int \overline{U}_{t+1}[\mathbf{\Gamma}](s) w(s|x,h) ds + \beta \int \overline{V}_{t+1}[\mathbf{\Gamma}](s) m(s|x,h) ds\right\}}{\exp\{\beta \int \overline{V}_{t+1}[\mathbf{\Gamma}](s) m_0(s|h) ds\} + \Gamma_{mt}(h)}$$
$$\times \frac{D_{mt-1} m_t[\mathbf{\Gamma}](h)}{m_t[\mathbf{\Gamma}](h)} m_t[\mathbf{\Gamma}](h) dh$$

where I define $D_{mt-1} m_t$ as the derivative of m_t with respect to $\log \Gamma_{mt-1}$ which can be written as:

$$D_{mt-1} m_t[\mathbf{\Gamma}](z) = \int_{\mathcal{X}_{t-1}} \int_{\mathcal{Z}_{t-1}} m(z|s,h) f_{t-1}(s,h) \left[- \frac{\Gamma_{mt-1}(h)}{\exp\{\beta \int \overline{V}_t(s) m_0(t|h) dt\} + \Gamma_{mt-1}(h)} \right.$$
$$\left. + \frac{D_{mt-1} m_{t-1}[\mathbf{\Gamma}](h)}{m_{t-1}(h)} \right] dh ds$$

$$
\begin{aligned}
&+\int_{\mathcal{Z}_1} m_0(z|h)f_1(*,h)\left[-\frac{\Gamma_{mt-1}(h)}{\exp\{\beta\int\overline{V}_t(s)m_0(t|h)dt\}+\Gamma_{mt-1}(h)}\right.\\
&\left.+\frac{D_{mt-1}m_{t-1}[\mathbf{\Gamma}](h)}{m_{t-1}(h)}\right]dh\\
&\leq m_t(z)\left[-\lambda_1+\frac{D_{mt-1}m_{t-1}[\mathbf{\Gamma}](h)}{m_{t-1}(h)}\right]
\end{aligned}
$$

Similarly, we can iterate once more and write using Equation 3.6

$$
\begin{aligned}
D_{mt-1}m_{t-1}[\mathbf{\Gamma}](z) =& \int_{\mathcal{X}_{t-2}}\int_{\mathcal{Z}_{t-2}} m(z|s,h)f_{t-2}(s,h)\left[\beta\int D_{mt-1}\overline{V}_{t-1}(s)m(t|s,h)dt\right.\\
&-\beta\int D_{mt-1}\overline{V}_{t-1}(s)m_0(t|h)dt\frac{\exp\{\beta\int\overline{V}_{t-1}(s)m_0(t|h)dt\}}{\exp\{\beta\int\overline{V}_{t-1}(s)m_0(t|h)dt\}+\Gamma^*_{mt-2}(h)}\\
&\left.+\frac{D_{mt-1}m_{t-2}[\mathbf{\Gamma}](h)}{m_{t-2}(h)}\right]dhds\\
&+\int_{\mathcal{Z}_1} m_0(z|h)f_1(*,h)\left[\beta\int D_{mt-1}\overline{V}_{t-1}(s)m_0(t|s,h)dt\right.\\
&-\beta\int D_{mt-1}\overline{V}_{t-1}(s)m_0(t|h)dt\frac{\exp\{\beta\int\overline{V}_{t-1}(s)m_0(t|h)dt\}}{\exp\{\beta\int\overline{V}_{t-1}(s)m_0(t|h)dt\}+\Gamma^*_{mt-2}(h)}\\
&\left.+\frac{D_{mt-1}m_{t-2}[\mathbf{\Gamma}](h)}{m_{t-2}(h)}\right]dh\\
&\leq m_{t-1}(z)\left[\beta\lambda_1-\beta\lambda_1\lambda_2+\frac{D_{mt-1}m_{t-2}[\mathbf{\Gamma}](h)}{m_{t-2}(h)}\right]
\end{aligned}
$$

Using Equation 3.7, we then iterate further:

$$
\begin{aligned}
D_{mt-1}m_{t-1}[\mathbf{\Gamma}](z) &\leq m_{t-1}(z)\left[\beta\lambda_1-\beta\lambda_1\lambda_2+\beta^2\lambda_1\lambda_2-\beta^2\lambda_1\lambda_2^2+\frac{D_{mt-1}m_{t-3}[\mathbf{\Gamma}](h)}{m_{t-3}(h)}\right]\\
&\leq m_{t-1}(z)\left[\beta\lambda_1-\beta^2\lambda_1\lambda_2^2+\frac{D_{mt-1}m_{t-3}[\mathbf{\Gamma}](h)}{m_{t-3}(h)}\right]
\end{aligned}
$$

Given that $D_{mt-1}m_1 = 0$ by definition and that $\lambda_1 < 1$,λ_2 and $\beta < 1$, we can thus conclude by induction that:

$$\frac{D_{mt-1}m_{t-1}[\mathbf{\Gamma}](h)}{m_{t-1}(h)} < 1$$

which directly implies that:

$$\frac{D_{mt-1}m_t[\mathbf{\Gamma}](h)}{m_t(h)} < 1$$

and that the derivative of $\log\Psi_{wt}$ with respect to $\log\Gamma_{mt-1}$ is strictly bounded from above in absolute value by 1. Similar steps can be used to show the same result for the Gateaux derivative of $\log\Psi_{wt}$ with respect to any $\log\Gamma_{mt'}$ or $\log\Gamma_{wt'}$ with $t \neq t'$. Symmetrical results apply for $\log \Psi_{mt}$.

Overall this implies that:

$$\sup_{a\in[0,1]} \left|\left| D \log \boldsymbol{\Psi_w}[a \log \boldsymbol{\Gamma_m} + (1-a)\log \tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}}](x) \right|\right|_\infty < 1$$

which finishes to prove that there exists a constant $\lambda < 1$ such that:

$$\left|\left| \log \boldsymbol{\Psi_w}[\boldsymbol{\Gamma_m}] - \log \boldsymbol{\Psi_w}[\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}}] \right|\right|_\infty \leq \lambda \left|\left| \log \boldsymbol{\Gamma_m} - \log \tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}} \right|\right|_\infty$$

I thus conclude that the mapping $(\log \boldsymbol{\Gamma_w}, \log \boldsymbol{\Gamma_m}) \mapsto (\log \Psi_w[\boldsymbol{\Gamma}], \log \Psi_m[\boldsymbol{\Gamma}])$ is a contraction which proves claim (i) of Theorem 1. The proof of part (ii) is a direct implication of the Banach fixed point theorem.

Before proving part (iii) of Theorem 1, intermediary steps are needed. In what follows, I follow Menzel (2015) and first prove that the size of opportunity sets grow at a rate $\sqrt{n}$. From this, I then show that the dependence between opportunity sets and taste shocks under the extremal matchings vanishes as n grows to infinity. I then use this result to show that we can approximate inclusive values arising from any stable match by inclusive value functions which have an approximate fixed point representation. I then finally prove that the solution to the finite sample fixed point problem converges to the unique solution of the population fixed point problem which concludes the proof of Theorem 1.(iii).

Rate of Size of Feasible Choice Sets

Define, for a given stable matching μ_t^*, the number of schools feasible to teacher i and the number of teachers feasible to school j in period t as:

$$J_{wit}^* = \sum_{j=1}^{n_m} \mathbb{1}\{V_{ijt} \geq \max_{l \in W_{jt}(\mu_t^*)\cup\{0\}} V_{ljt}\} \quad \text{and} \quad J_{mjt}^* = \sum_{i=1}^{n_w} \mathbb{1}\{U_{ijt} \geq \max_{k \in M_{it}(\mu_t^*)\cup\{0\}} U_{ikt}\}$$

Similarly, define the number of school that teacher i would accept and the number of teachers that school j would accept:

$$L_{wit}^* = \sum_{j=1}^{n_m} \mathbb{1}\{U_{ijt} \geq \max_{k \in M_{it}(\mu_t^*)\cup\{0\}} U_{ikt}\} \quad \text{and} \quad L_{mjt}^* = \sum_{i=1}^{n_w} \mathbb{1}\{V_{ijt} \geq \max_{l \in W_{jt}(\mu_t^*)\cup\{0\}} V_{ljt}\}$$

I now state the following result:

Lemma 1 *Under Assumptions 1-3 and for any stable matching μ_t^*, we have:*

$$n^{1/2}\frac{\exp(-\bar{V}-\beta\overline{V}^U+\gamma_m)}{1+\exp(\bar{U}+\bar{V}+\beta\overline{U}^U+\beta\overline{V}^U+\gamma_w)} \leq J_{wi}^* \leq n^{1/2}\exp(\bar{V}+\beta\overline{V}^U+\gamma_m)$$
$$n^{1/2}\frac{\exp(-\bar{U}-\beta\overline{U}^U+\gamma_w)}{1+\exp(\bar{U}+\bar{V}+\beta\overline{U}^U+\beta\overline{V}^U+\gamma_m)} \leq J_{mj}^* \leq n^{1/2}\exp(\bar{U}+\beta\overline{U}^U+\gamma_w)$$
$$n^{1/2}\frac{\exp(-\bar{U}-\beta\overline{U}^U+\gamma_m)}{1+\exp(\bar{U}+\bar{V}+\beta\overline{U}^U+\beta\overline{V}^U+\gamma_m)} \leq L_{wi}^* \leq n^{1/2}\exp(\bar{U}+\beta\overline{U}^U+\gamma_m)$$
$$n^{1/2}\frac{\exp(-\bar{V}-\beta\overline{V}^U+\gamma_w)}{1+\exp(\bar{U}+\bar{V}+\beta\overline{U}^U+\beta\overline{V}^U+\gamma_w)} \leq L_{mj}^* \leq n^{1/2}\exp(\bar{V}+\beta\overline{V}^U+\gamma_w)$$

for each $i = 1, ..., n_w$ and $j = 1, ..., n_m$ with probability approaching 1 as $n \to \infty$.

Proof: As in Menzel (2015), we can define exogenous sets $\overline{W}_{jt} = \{i : U_{ijt} \geq U_{i0t}\}$ and $\overline{M}_{it} = \{j : V_{ijt} \geq V_{0jt}\}$ such that $W_{jt}(\mu_t^*) \subset \overline{W}_{jt}$ and $M_{it}(\mu_t^*) \subset \overline{M}_{it}$ as well as $W_{jt}^\circ = \{i : U_{ijt} \geq \max_{k\in\overline{M}_{it}(\mu_t^*)\cup\{0\}} U_{ikt}\}$ and $M_{it}^\circ = \{j : V_{ijt} \geq \max_{l\in\overline{W}_{jt}(\mu_t^*)\cup\{0\}} V_{ljt}\}$ such that $W_{jt}^\circ \subset W_{jt}(\mu_t^*)$ and $M_{it}^\circ \subset M_{it}(\mu_t^*)$.

From this, I construct the following bounds on J_{wi}^*:

$$J_{wit}^\circ = \sum_{j=1}^{n_m} \mathbb{1}\{j \in M_{it}^\circ\} \leq \sum_{j=1}^{n_m} \mathbb{1}\{j \in M_{it}(\mu^*)\} \leq \sum_{j=1}^{n_m} \mathbb{1}\{j \in \overline{M}_{it}\} = \overline{J}_{wit}$$

from there, using Proposition 2, we can show that:

$$\begin{aligned}\mathbb{E}[\overline{J}_{wit}|x_{it}, z_{1t}, ..., z_{n_m t}] &= \frac{1}{J}\sum_{j=1}^{n_m}\frac{\exp\{V(x_{it}, z_{jt}) + \beta\int\overline{V}_{jt+1}(s)m(s|x_{it}, z_{jt})ds\}}{1+\frac{1}{J}\exp\{V(x_{it}, z_{jt}) + \beta\int\overline{V}_{jt+1}(s)m(s|x_{it}, z_{jt})ds\}} + o(1)\\ &\leq \frac{n_m}{J}\exp\{\bar{V}+\beta\overline{V}^U\} + o(1)\end{aligned}$$

which implies under Assumption 3 that:

$$\mathbb{E}[\overline{J}_{wit}] \leq n^{1/2}\exp\{\bar{V}+\beta\overline{V}^U+\gamma_m\} + o(1)$$

Following the same steps as Menzel (2015) we can then show that the variance of $\overline{J}_{wit}$

converges to zero which implies that:

$$n^{-1/2}(\overline{J}_{wit} - \mathbb{E}[\overline{J}_{wit}]) \to 0$$

We have thus established that $J^*_{wit} \leq n^{1/2}\exp\{\overline{V} + \beta\overline{V}^U + \gamma_m\}$ with probability approaching 1 as $n \to \infty$. Following the same steps, we can show symmetrically that:

$$J^*_{mjt} \leq n^{1/2}\exp\{\bar{U} + \beta\overline{U}^U + \gamma_w\}$$

$$L^*_{wit} \leq n^{1/2}\exp\{\bar{V} + \beta\overline{V}^U + \gamma_m\}$$

$$L^*_{mjt} \leq n^{1/2}\exp\{\bar{U} + \beta\overline{U}^U + \gamma_w\}$$

with probability approaching 1 as $n \to \infty$. We now consider the lower bound J°_{wit}. We can again use Proposition 2 to show that:

$$\begin{aligned}\mathbb{E}[J^\circ_{wit}|(x_{lt})_{l\in\overline{W}_{jt}}, (z_{kt})_{k=1}^{n_m}] &= \frac{1}{J}\sum_{j=1}^{n_m}\frac{\exp\{V_t(x_{it}, z_{jt}) + \beta\int\overline{V}_{jt+1}(s)m(s|x_{it}, z_{jt})ds\}}{1 + \frac{1}{J}\sum_{l\in\overline{W}_{jt}}\exp\{V(x_{lt}, z_{jt}) + \beta\int\overline{V}_{jt+1}(s)m(s|x_{lt}, z_{jt})ds\}} + o(1)\\ &\geq \frac{n_m}{J}\frac{\exp\{-\bar{V} - \beta\overline{V}^U\}}{1 + \frac{\overline{J}_{mjt}}{J}\exp\{\bar{V} + \beta\overline{V}^U\}} + o(1)\end{aligned}$$

Using the higher bound for J^*_{mj} derived just above and Jensen's inequality, we can finally show that:

$$\mathbb{E}[J^\circ_{wit}] \geq n^{1/2}\frac{\exp\{-\bar{V} - \beta\overline{V}^U + \gamma_m\}}{1 + \exp\{\bar{V} + \bar{U} + \beta\overline{V}^U + \beta\overline{U}^U + \gamma_w\}} + o(1)$$

Following Menzel (2015) we can then also show that the variance of J°_{wit} converges to zero which implies that:

$$n^{-1/2}(J^\circ_{wit} - \mathbb{E}[J^\circ_{wit}]) \to 0$$

This establishes that $J^*_{wit} \geq n^{1/2}\frac{\exp\{-\bar{V} - \beta\overline{V}^U + \gamma_m\}}{1 + \exp\{\bar{V} + \bar{U} + \beta\overline{V}^U + \beta\overline{U}^U + \gamma_w\}}$ with probability approaching 1 as $n \to \infty$. Following the same steps, we can show that symmetrically, we have:

$$J^*_{mjt} \geq n^{1/2}\frac{\exp\{-\bar{U} - \beta\overline{U}^U + \gamma_w\}}{1 + \exp\{\bar{V} + \bar{U} + \beta\overline{V}^U + \beta\overline{U}^U + \gamma_m\}}$$

$$L^*_{wit} \geq n^{1/2} \frac{\exp\{-\bar{U} - \beta\overline{U}^U + \gamma_m\}}{1 + \exp\{\bar{V} + \bar{U} + \beta\overline{V}^U + \beta\overline{U}^U + \gamma_m\}}$$

$$L^*_{mjt} \geq n^{1/2} \frac{\exp\{-\bar{V} - \beta\overline{V}^U + \gamma_w\}}{1 + \exp\{\bar{V} + \bar{U} + \beta\overline{V}^U + \beta\overline{U}^U + \gamma_w\}}$$

with probability approaching 1 as $n \to \infty$. This concludes the proof of Lemma 1.

Exogeneity of Feasible Choice Sets

We now need to show that as $n \to \infty$, the dependence between agents taste shocks and opportunity sets vanishes. As there taste shocks are independent across periods under Assumption 1, the dependence between unobserved preferences and opportunity sets can only arise within period. The proof thus mirrors very closely Menzel (2015) which proves the same result in the static case.

For the rest of the proof, I define the following set of indicator functions $E^*_{ijt} = \mathbb{1}\{i \in W_{jt}(\mu^*_t)\}$ and $D^*_{ijt} = \mathbb{1}\{j \in M_{it}(\mu^*_t)\}$ for all teachers $i = 1, ..., n_w$ and schools $j = 1, ..., n_m$. The first result to establish is that the probability that changing one availability indicator affects another agents' opportunity set converges to zero as $n \to \infty$. I first prove the following result:

Lemma 2 *Suppose Assumption 1-3 hold and suppose we change one availability indicator E^*_{ijt} exogenously to $\tilde{E}_{ijt} = 1 - E^*_{ijt}$ and then iterate the deferred acceptance algorithm from this point until convergence. Denote the resulting availability indicators $\{\tilde{E}_{lkt}, \tilde{D}_{lkt} : l = 1, ..., n_w, k = 1, ..., n_m\}$. We have for any teacher l and school k:*

(i). $\mathbb{P}(\tilde{D}_l \neq D^*_l | D^*_l, D^*_{ij} = 0) = \mathbb{P}(\tilde{E}_k \neq E^*_k | E^*_l, D^*_{ij} = 0) = 0$

(ii). There exist constants $\overline{a} < \infty$ and $0 < \lambda < 1$ such that:

$$\mathbb{P}(\tilde{D}_l \neq D^*_l | D^*_l, D^*_{ijt} = 1) \leq n^{-1/2} \frac{\overline{a}}{1 - \lambda}$$

$$\mathbb{P}(\tilde{E}_k \neq E^*_k | E^*_l, E^*_{ijt} = 1) \leq n^{-1/2} \frac{\overline{a}}{1 - \lambda}$$

The same result holds for an exogenous change of D_{ijt} to $\tilde{D}_{ijt} = 1 - D_{ijt}$.

PROOF: Suppose we change E^*_{jit} exogenously to $\tilde{E}_{jit} = 1 - E_{jit}$ and that we iterate the deferred acceptance algorithm from this stage. This will only trigger a chain of rematches if this affects the indirect utility of either i or j. Suppose $D^*_{ijt} = 0$ and that $E^*_{ijt} = 0$ meaning that school j is not feasible to teacher i and vice versa. Suppose now that $\tilde{E}_{jit} = 1 - E^*_{ijt} = 1$, meaning that suddenly teacher i's preference for school j increase such that teacher i becomes feasible for school j. This will not affect the indirect utility of school j nor teacher i given that school j is not feasible to teacher i. This change will thus not trigger a chain of rematches. A similar argument can be used in the case where E^*_{ijt} changes from 1 to $\tilde{E}_{jit} = 1 - E^*_{ijt} = 0$. This establishes part (i) of Lemma 2.

Now suppose that $D^*_{ijt} = 1$ such that if $\tilde{E}_{ijt} = 1 - E^*_{ijt} = 1$, now school j and teacher i will want to rematch together or if $\tilde{E}_{ijt} = 1 - E^*_{ijt} = 0$ school j and teacher i will break their current match. This will trigger a chain of rematches than can potentially cycle back to teacher i or school j's opportunity set. I start by showing that, at each step s of these subsequent rematches, there is at most one indicator in the vector $D^{(s)}_l$ corresponding to a school k with $E^{(s)}_{lkt} = 1$ that will change. The idea of the proof is the following: suppose that a given teacher l matched to school k in step $(s-1)$ becomes unavailable to school k in step s. This school will then replace this teacher by its most preferred feasible applicant, which will only change the availability indicator of this school to this newly hired teacher. On the other hand, if a given teacher becomes available to a school while this school prefers this teacher to its matched employee, then it will replace them by this new employee, making this school unavailable to the kicked out employee. In both cases, this will only change at most one availability indicator among the teachers who are willing to match with this school. Note that at each of these steps, there is a chance that the chain is terminated if the next preferred feasible option is the outside option. A similar argument can be used symmetrically from the teachers perspective.

The rest of the proof now consists in bounding the probability that the chain is terminated by either (a) school k or teacher l preferring the outside option to any other option in their opportunity set or (b) a change in availability indicators of teacher k D_k. I define μ^s_t the state of the match in iteration s of the deferred acceptance algorithm following an exogenous change of E_{ijt} to $\tilde{E}_{ijt} = 1 - E_{ijt}$. The first step bounds the probability that the chain is terminated by the outside option at stage s.

I start from the following observation: given that $\mathbb{P}(V_{lkt} > V_{k,(q)}(W_k(\mu^s))|x_l, z_k) \geq \mathbb{P}(V_{lk} > V_{k,(1)}(W_k(\mu^s))|x_l, z_k)$ and that $W^\circ_{k,(1)} \subset W^*_k \subset \overline{W}_k$, we have from Proposition 2 and Lemma 1 that for any school k and teacher l:

$$\begin{aligned}
&\mathbb{P}(V_{lkt} > \max_{l \in W_{kt}(\mu_t^s) \cup \{0\}} V_{lkt} | x_{lt}, z_{kt}) \\
&\geq \mathbb{P}(V_{lkt} > \max_{l \in \overline{W}_{kt} \cup \{0\}} V_{lkt} | x_{lt}, z_{kt}) \\
&= n^{-1/2} \frac{\exp(V(z_{kt}, x_{lt}) + \beta \int \overline{V}_{kt+1}(s) m(s|x_{lt}, z_{kt}) ds)}{1 + \frac{1}{J} \sum_{i \in \overline{W}_{kt}} \exp(V(z_{kt}, x_{it}) + \beta \int \overline{V}_{kt+1}(s) m(s|x_{it}, z_{kt}) ds)} + o(1) \\
&\geq n^{-1/2} \frac{\exp(V(z_{kt}, x_{lt}) + \beta \int \overline{V}_{kt+1}(s) m(s|x_{lt}, z_{kt}) ds)}{1 + \exp(\bar{U} + \bar{V} + \beta \overline{V}^U + \beta \overline{U}^U + \gamma_w)} + o(1)
\end{aligned}$$

This implies that, conditional on D_i^* and as n approaches infinity:

$$\mathbb{P}(V_{0kt} > \max_{l \in W_{kt}(\mu_t^s) \cup \{0\}} V_{lkt} | D_i^*, x_{it}, z_{kt}) \geq \frac{1}{1 + \exp(\bar{U} + \bar{V} + \beta \overline{V}^U + \beta \overline{U}^U + \gamma_w)} =: p_s$$

Following now the same steps as Menzel (2015), we have, by Bayes law that:

$$\mathbb{P}(V_{0kt} > \max_{l \in W_{kt}(\mu_t^s) \cup \{0\}} V_{ljt} | D_l^*, \tilde{D}_{lkt}^{(s)} = 1, x_{lt}, z_{kt}) \geq \frac{\underline{L} p_s}{\overline{L}(1 - p_s) + \underline{L} p_s}$$

where $\overline{L}$ and $\underline{L}$ are respectively the upper and lower bounds on L_{mj}^* taken from Lemma 1. From there, we finally get that:

$$1 - \mathbb{P}(V_{0kt} > \max_{l \in W_{kt}(\mu_t^s) \cup \{0\}} V_{ljt} | D_l^*, \tilde{D}_{lkt}^{(s)} = 1, x_{lt}, z_{kt}) \leq \frac{\overline{L} \exp(\bar{U} + \bar{V} + \beta \overline{V}^U + \beta \overline{U}^U + \gamma_w)}{\overline{L} \exp(\bar{U} + \bar{V} + \beta \overline{V}^U + \beta \overline{U}^U + \gamma_w) + \underline{L}} =: \lambda < 1$$

This essentially means that the probability that the chain is not terminated at stage s is bounded away from 1.

Now we bound the probability that the chain leads to a change in D_l at stage s. We can thus bound the following probability using Proposition 2 and Lemma 1:

$$\begin{aligned}
&\mathbb{P}(V_{lkt} > \max_{l \in W_{kt}(\mu_t^s) \cup \{0\}} V_{lkt}) | x_{lt}, z_{kt}) \\
&\leq \mathbb{P}(V_{lkt} > \max_{l \in W_{kt}^{\circ} \cup \{0\}} V_{lkt} | x_{lt}, z_{kt}) \\
&= n^{-1/2} \frac{\exp(V(z_{kt}, x_{lt}) + \beta \int \overline{V}_{kt+1}(s) m(s|x_{lt}, z_{kt}) ds)}{1 + \frac{1}{J} \sum_{i \in W_{kt}^{\circ}} \exp(V(z_{kt}, x_{it}) + \beta \int \overline{V}_{kt+1}(s) m(s|x_{it}, z_{kt}) ds)} + o(1) \\
&\leq n^{-1/2} \exp(\bar{V} + \beta \overline{V}^U) + o(1)
\end{aligned}$$

This implies that for n sufficiently large, we have:

$$\mathbb{P}(\tilde{D}_l^{(s)} \neq D_l^* | D_l^*, \tilde{D}_{lkt}^{(s)} = 1, x_l, z_k)$$
$$\leq \frac{n^{-1/2}\exp(\bar{V} + \beta\overline{V}^U)\overline{L}}{n^{-1/2}\exp(\bar{V} + \beta\overline{V}^U)\overline{L} + \underline{L}} \leq n^{-1/2}\exp(\bar{V} + \beta\overline{V}^U)\frac{\overline{L}}{\underline{L}} = n^{-1/2}\overline{a}$$

Using the law of total probability, we can thus bound as $n \to \infty$ the conditional probability that $\tilde{D}_l \neq D_l^*$

$$\mathbb{P}(\tilde{D}_l \neq D_l^* | D_l^*) \leq \sum_{s=1}^{\infty} \lambda^s n^{-1/2}\overline{a} \leq \frac{n^{-1/2}\overline{a}}{1-\lambda}$$

which proves part (b) of Lemma 2.

From there, I state the main result that the dependence between taste shocks and agents' opportunity sets vanishes as $n \to \infty$. I first define the joint distribution of $\eta_i = (\eta_{i1}, ..., \eta_{in_m})'$, $\epsilon_j = (\epsilon_{1j}, ..., \epsilon_{n_w j})'$ and the availability indicators D_i^W, E_j^W, D_i^M, E_j^M corresponding to the teacher-optimal and the school-optimal stable matches. Note that I consider these two specific matches since the teacher-optimal and school-optimal stable matches are defined with probability 1 conditional on the realization of the taste shocks η_i and ϵ_j. Indeed, the distribution of availability indicators arising from an arbitrary stable match D_i^* would not be well defined. I also define: $D_{i,-j}^W = (D_{i1}^W, ..., D_{i(j-1)}^W, D_{i(j+1)}^W, ..., D_{in_m}^W)$ and $E_{-i,j} = (E_{1j}^W, ..., E_{(i-1)j}^W, E_{(i+1)j}^W, ..., E_{n_w j}^W)$ with analogous notations for the school optimal match. I then define the conditional c.d.f.s:

$$G_{\eta|D}^W(\eta|\boldsymbol{d}) = \mathbb{P}(\eta_i \leq \eta | D_i^W = \boldsymbol{d}), \quad \boldsymbol{d} \in \{0,1\}^{n_m}$$

$$G_{\eta,\epsilon|D,E}^W(\eta, \epsilon|\boldsymbol{d}, \boldsymbol{e}) = \mathbb{P}(\eta_i \leq \eta, \epsilon_j \leq \epsilon | D_{i,-j}^W = \boldsymbol{d}, E_{-i,j}^W = \boldsymbol{e}), \quad \boldsymbol{d} \in \{0,1\}^{n_m - 1}, \boldsymbol{e} \in \{0,1\}^{n_w - 1}$$

with analogous definitions for the school-optimal stable match and associated p.d.f.s $g_{\eta|D}^W$ and $g_{\eta,\epsilon|D,E}^W$. The main result is the following:

Lemma 3 *Under Assumption 1 and 2, we have:*

(i). $g^W_{\eta|D}$ *and* $g^M_{\eta|D}$ *satisfy:*

$$\lim_n \Big| \frac{g^W_{\eta|D}(\eta|D^W_i)}{g_\eta(\eta)} - 1 \Big| = \lim_n \Big| \frac{g^M_{\eta|D}(\eta|D^M_i)}{g_\eta(\eta)} - 1 \Big| = 1$$

(ii). $g^W_{\eta,\epsilon|D,E}$ *and* $g^M_{\eta,\epsilon|D,E}$ *satisfy:*

$$\lim_n \Big| \frac{g^W_{\eta|D}(\eta,\epsilon|D^W_{i,-j}, E^W_{-i,j})}{g_{\eta,\epsilon}(\eta,\epsilon)} - 1 \Big| = \lim_n \Big| \frac{g^M_{\eta|D}(\eta,\epsilon|D^M_{i,-j}, E^M_{-i,j})}{g_{\eta,\epsilon}(\eta,\epsilon)} - 1 \Big| = 1$$

The same results holds for the school side of the market.

Proof: Let $g^W_{\eta,D}$ be the joint p.d.f. of taste shocks and availability indicators under the teacher optimal stable match. We can rewrite, by definition of a conditional density:

$$\frac{g^W_{\eta|D}(\eta|D^W_i)}{g_\eta(\eta)} = \frac{g^W_{\eta,D}(\eta, D^W_i)}{g_\eta(\eta)P(D^W_i)} = \frac{P(D^W_i|\eta_i = \eta)g_\eta(\eta)}{g_\eta(\eta)P(D^W_i)} = \frac{P(D^W_i|\eta_i = \eta)}{P(D^W_i)}$$

I then follow similar steps as in Menzel (2015) to show that:

$$\Big| \frac{P(D^W_i|\eta_i = \eta)}{P(D^W_i)} - 1 \Big| \leq \sup_{\eta_1,\eta_2} \Big| \frac{P(D^W_i|\eta_i = \eta_1)}{P(D^W_i|\eta_i = \eta_2)} - 1 \Big|$$

such that I only need to bound the probability that shifting η_i from η_1 to η_2 changes teacher i's opportunity set. We know from Lemma 2 that changing an availability indicator will trigger a chain of rematches that could change teacher i's opportunity set with probability less than $\frac{n^{-1/2}\bar{a}}{1-\lambda}$ as n approaches infinity. Here, we can show that shifting agent i's taste shocks would trigger at most two chains of rematches. Indeed, if the shift in taste shocks makes agent i prefers school l with $D_{il} = 1$ instead of her current employer school j, this changes both E_{ij} from 1 to 0 and E_{il} from 0 to 1. Thus, this would trigger two chains of rematches where both school j and the teacher which was displaced from school l by teacher i would need to find a new match. We can thus conclude that:

$$\frac{P(D^W_i|\eta_i = \eta_1)}{P(D^W_i|\eta_i = \eta_2)} - 1 \leq 2\frac{n^{-1/2}\bar{a}}{1-\lambda}$$

as $n \to \infty$ which can be shown to hold also in absolute value. As the right hand side converges to 0 as $n \to \infty$, this proves the first part of claim (i). The same result holds symetrically for the school side.

For part (ii), note that the argument can be extended in a similar way. If you change both school j and teacher i's taste shocks this can trigger at most 4 chains of rematches such that we can bound the probability of a shift in opportunity sets by $n^{-1/2}\frac{4\bar{a}}{1-\lambda}$ which can be made arbitrarily close to 0 as n approaches infinity.

Bounds for Inclusive Values

Since I have established exogeneity of opportunity sets under the school-optimal and teacher-optimal stable matches, the rest of the analysis focuses on characterizing the limit of inclusive values that arise under these extremal matchings. As in Menzel (2015), I show that both converge to a unique limit, implying that inclusive values arising from any stable matching also converge toward this limit.

I define $I^W_{wit} = I_{wit}(\mu^W_t)$ and $I^W_{mjt} = I_{mjt}(\mu^W_t)$ the inclusive values that arise from the sequence of teacher-optimal stable matches $\boldsymbol{\mu}^W$ in period t. Similarly, I define I^M_{wit} and I^M_{mjt} as the inclusive values that arise from the sequence of school-optimal stable matches $\boldsymbol{\mu}^M$ such that for any stable match μ^*_t, we have $I^W_{wit} \geq I_{wit}(\mu^*_t) \geq I^M_{wit}$ and $I^W_{mjt} \leq I_{wit}(\mu^*_t) \leq I^M_{mjt}$ for all t. I state the following result:

Lemma 4 *Under Assumption 1-3:*
(i). For all $i = 1, ..., n_w$ and $j = 1, ..., n_m$:

$$I^M_{wit} \geq \hat{\Gamma}^M_{wt}(x_{it}) + o_p(1) \quad \textit{and} \quad I^M_{mjt} \leq \hat{\Gamma}^M_{mt}(z_{jt}) + o_p(1)$$

where the analogous result holds for the teacher-optimal stable match with the side of inequalities reversed.
(ii). If the weight functions $\omega(x, z) \geq 0$ are bounded and form a Glivenko-Cantelli class in x, then

$$\sup_{x \in \mathcal{X}} \frac{1}{n} \sum_{j=1}^{n_m} \omega(x, z_{jt})(I^M_{mjt} - \hat{\Gamma}^M_{mt}(z_{jt})) \leq o_p(1)$$

and

$$\inf_{z \in \mathcal{Z}} \frac{1}{n} \sum_{i=1}^{n_w} \omega(x_{it}, z)(I^M_{wit} - \hat{\Gamma}^M_{wt}(x_{it})) \geq o_p(1)$$

The analogous conclusion holds for the teacher-optimal stable match where the sign of the inequalities is reversed and if $\omega(x,z) \geq 0$ are bounded and form a Glivenko-Cantelli class in z.

Proof: I first show that we can bound conditional choice probabilities given an opportunity set arising from a stable match using the extremal matchings. I first define the conditional probability that teacher i chooses school j given the realization of opportunity set M^M arising from the school-optimal stable match:

$$\Lambda^M_{wt}(x,z,M^M) = \mathbb{P}(U_{ijt} \geq \max_{k \in M^M_{it} \cup \{0\}} U_{ikt} | (M^M_{i\tau})^T_{\tau=t} = M^M, x_{it} = x, z_{jt} = z)$$

and the expectations about future match payoffs given future opportunity sets as:

$$\overline{U}^M_{t+1}(x, M^M) = \mathbb{E}\left[\max_{k \in M^M_{it+1} \cup \{0\}} U_{ikt+1} | (M^M_{i\tau})^T_{\tau=t+1} = M^M, x_{it+1} = x\right]$$

I also define the conditional choice probabilities and expectations about future payoffs in period t under exogenous opportunity sets as:

$$\Lambda_{wt}(x,z,M) = \mathbb{P}(U_{ijt} \geq \max_{k \in M \cup \{0\}} U_{ikt} | x_{it} = x, z_{jt} = z)$$

$$\overline{U}_{t+1}(x, M) = \mathbb{E}\left[\max_{k \in M \cup \{0\}} U_{ikt+1} | x_{it+1} = x\right]$$

As there are several stable matches such that $M^*_i = M^M_i$ and $W^*_j = W^M_j$ we can show that:

$$J\Lambda^M_{wt}(x,z,(M^M_{i\tau})^T_{\tau=t}) \leq J\Lambda_{wt}(x,z,(M^M_{i\tau})^T_{\tau=t}) + o_p(1)$$

$$\overline{U}^M_{t+1}(x,(M^M_{i\tau})^T_{\tau=t+1}) \geq \overline{U}_{t+1}(x,(M^M_{i\tau})^T_{\tau=t+1}) + o_p(1)$$

Using Proposition 2, we can then show that for $i = 1, ..., n_w$, $l_1 = 1, ..., n_m$ and $l_2 \neq l_1$:

$$\mathbb{E}\Bigg[J\Bigg(D^M_{il_1t}\exp\left\{\beta\int \overline{U}^M_{it+1}(s) w(s|x_{it}, z_{l_1t})ds\right\}$$
$$- \Lambda^M_{mt}(x_{it}, z_{l_1t}, (I^M_{ml_1\tau})^T_{\tau=t})\exp\left\{\beta\int \overline{U}^M_{t+1}(s,(M^M_{i\tau})^T_{\tau=t+1}) w(s|x_{it}, z_{l_1t})ds\right\}\Bigg)\Bigg|(I^M_{ml_1\tau})^T_{\tau=t}, x_{it}, z_{l_1t}\Bigg] \to 0$$

and

$$\mathbb{E}\Bigg[J^2\Bigg(D_{il_1t}^M \exp\left\{\beta\int \overline{U}_{it+1}^M(s)w(s|x_{it},z_{jt})ds\right\}$$
$$-\Lambda_{mt}^M(x_{it},z_{l_1t},(I_{ml_1\tau}^M)_{\tau=t}^T)\exp\left\{\beta\int \overline{U}_{t+1}^M(s,(I_{wi\tau}^M)_{\tau=t+1}^T)w(s|x_{it},z_{l_1t})ds\right\}\Bigg)$$
$$\times D_{il_2t}^M \exp\left\{\beta\int \overline{U}_{it+1}^M(s)w(s|x_{it},z_{l_1t})ds\right\}$$
$$-\Lambda_{mt}^M(x_{it},z_{l_2t},(I_{ml_2\tau}^M)_{\tau=t}^T)\exp\left\{\beta\int \overline{U}_{t+1}^M(s,(I_{wi\tau}^M)_{\tau=t+1}^T)w(s|x_{it},z_{l_2t})ds\right\}$$
$$\Bigg|(I_{ml_1\tau}^M)_{\tau=t}^T,(I_{ml_2\tau}^M)_{\tau=t}^T,(I_{wi\tau}^M)_{\tau=t+1}^T,x_{it},z_{l_1t},z_{l_2t}\Bigg]\to 0$$

Therefore, since under Assumption 1, we know that $\exp(U_t(x_{it},z_{jt}))$ is bounded, we can thus conclude that:

$$\text{Var}\left(\frac{1}{n}\sum_{k=1}^{n_m}\exp\left\{U_t(x_{it},z_{kt})+\beta\int \overline{U}_{t+1}^M(s,(I_{wi\tau}^M)_{\tau=t+1}^T)w(s|x_{it},z_{kt})ds\right\}J(D_{ik}^M-\Lambda_{mt}^M(x_{it},z_{kt},(I_{mk\tau}^M)_{\tau=t}^T))\right)\to 0$$

which implies that:

$$\frac{1}{n}\sum_{k=1}^{n_m}\exp\left\{U_t(x_{it},z_{kt})+\beta\int \overline{U}_{t+1}^M(s,(I_{wi\tau}^M)_{\tau=t+1}^T)w(s|x_{it},z_{kt})ds\right\}J(D_{ikt}^M-\Lambda_{mt}^M(x_{it},z_{kt},(I_{mk\tau}^M)_{\tau=t}^T))=o_p(1)$$

Given that from Proposition 2:

$$J\Lambda_{mt}^M(x,z,(W_{j\tau}^M)_{\tau=t}^T)\geq\frac{\exp\left\{V_t(x,z)+\beta_m\int \overline{V}_{t+1}^M(s)m(s|\boldsymbol{x},\boldsymbol{z})ds\right\}}{\exp\left\{\beta_m\int \overline{V}_{t+1}^M(s)m_0(s|\boldsymbol{x})ds\right\}+I_{mjt}^M}+o_p(1)$$

This implies that:

$$\frac{1}{n}\sum_{k=1}^{n_m}\exp\{U_t(x_{it},z_{kt})\}\left(JD_{ikt}^M-\frac{\exp\left\{V_t(x_{it},z_{kt})+\beta_m\int \overline{V}_{t+1}^M(s)m(s|x_{it},z_{kt})ds\right\}}{\exp\left\{\beta_m\int \overline{V}_{t+1}^M(s)m_0(s|z_{kt})ds\right\}+I_{mjt}^M}\right)\geq o_p(1)$$

which proves the first claim of part (i) of Lemma 4. Similar steps can be used to bound inclusive values on the school side and for the teacher optimal sequence of stable matches.

Part (ii) follows from part (i) of the Lemma and the boundedness condition on ω which implies pointwise convergence. The Glivenko-Cantelli condition on ω then implies uniform convergence. This concludes the proof of Lemma 4.

The next step consists in establishing uniform convergence with respect to $\Gamma_{wt} \in \mathcal{T}_{wt}$ and $\Gamma_{mt} \in \mathcal{T}_{mt}$ of the fixed point mappings $\hat{\Psi}_{wt}$ and $\hat{\Psi}_{mt}$ to their population counterparts. I define:

$$\hat{\Psi}_{wt}[\mathbf{\Gamma}](x) = \frac{1}{n}\sum_{j=1}^{n_m} \psi_{wt}(z_{jt}, x; \mathbf{\Gamma})$$

where ψ_{wt} is defined as:

$$\psi_{wt}(z_{jt}, x; \mathbf{\Gamma}) = \frac{\exp\left\{U_t(x, z_{jt}) + V_t(x, z_{jt}) + \beta_w \int \overline{U}_{t+1}(s) w(s|x, z_{jt}) ds + \beta_m \int \overline{V}_{t+1}(s) m(s|x, z_{jt}) ds\right\}}{\exp\left\{\beta_m \int \overline{V}_{t+1}(s) m_0(s|z_{jt}) ds\right\} + \Gamma_{mt}(z_{jt})}$$

Similarly, I define:

$$\hat{\Psi}_{mt}[\mathbf{\Gamma}](z) = \frac{1}{n}\sum_{i=1}^{n_w} \psi_{mt}(z, x_{it}; \mathbf{\Gamma})$$

where ψ_{mt} is defined as:

$$\psi_{mt}(z, x_{it}; \mathbf{\Gamma}) = \frac{\exp\left\{U_t(x_{it}, z) + V_t(x_{it}, z) + \beta_w \int \overline{U}_{t+1}(s) w(s|x_{it}, z) ds + \beta_m \int \overline{V}_{t+1}(s) m(s|x_{it}, z) ds\right\}}{\exp\left\{\beta_w \int \overline{U}_{t+1}(s) w_0(s|x_{it}) ds\right\} + \Gamma_{wt}(x_{it})}$$

I define the class of functions $\mathcal{F}_w : \{\psi_w(., x; \mathbf{\Gamma}) : x \in \mathcal{X}, \mathbf{\Gamma} \in \mathcal{T}\}$ and $\mathcal{F}_m : \{\psi_m(z, ; \mathbf{\Gamma}) : z \in \mathcal{Z}, \mathbf{\Gamma} \in \mathcal{T}\}$.

Lemma 5 *Under Assumption 1:*
(i). The classes of functions $\mathcal{F}_w$ and $\mathcal{F}_w$ are Glivenko-Cantelli.
(ii). As $n \to \infty$ and for all t:

$$(\hat{\Psi}_{wt}[\mathbf{\Gamma}](x), \hat{\Psi}_{mt}[\mathbf{\Gamma}](z)) \to (\Psi_{wt}[\mathbf{\Gamma}](x), \Psi_{mt}[\mathbf{\Gamma}](z))$$

uniformly in $\mathbf{\Gamma} \in \mathcal{T}$, and $(x, z) \in \mathcal{X} \times \mathcal{Z}$.

Proof: Under Assumption 1, $\exp\{U(x, z) + V(x, z)\}$ is Lipschitz in x and z such that this class of functions is Glivenko-Cantelli. Γ_{mt} and Γ_{wt} are bounded and have bounded $p \geq 1$

derivatives for all t which makes the class of functions $\mathcal{T}$ Glivenko-Cantelli. Finally, as m and w are continuous densities and that $\overline{U}$ and $\overline{V}$ are continuous note that the transformation $\psi_m(g,h) = \frac{g}{1+h}$ is bounded and continuous since h and g are bounded and continuous and $h \geq 0$. Theorem 3 in van der Vaart and Wellner (2000) implies claim (i) of Lemma 5. Part (ii) of Lemma 5 is a direct implication of part (i).

Proof of Theorem 3.1 (iii)

I finally turn to the proof of part (iii) of Theorem 1. I first apply Lemma 4 to show that for any $q \geq 1$:

$$\hat{\Gamma}^M_{wt}(x_{it}) = \frac{1}{n}\sum_{k=1}^{n_m} \frac{\exp\left\{U_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) + V_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) + \beta\int \overline{U}^M_{it+1}(s)w(s|\boldsymbol{x}_{it}, \boldsymbol{z}_{kt})ds + \beta\int \overline{V}^M_{kt+1}(s)m(s|\boldsymbol{x}_{it}, \boldsymbol{z}_{kt})ds\right\}}{\exp\left\{\beta\int \overline{V}^M_{kt+1}(s)m_0(s|\boldsymbol{z}_{kt})ds\right\} + I^M_{mkt}}$$

$$\geq \frac{1}{n}\sum_{k=1}^{n_m} \frac{\exp\left\{U_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) + V_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{kt}) + \beta\int \overline{U}^M_{t+1}(s)w(s|\boldsymbol{x}_{it}, \boldsymbol{z}_{kt})ds + \beta\int \overline{V}^M_{t+1}(s)m(s|\boldsymbol{x}_{it}, \boldsymbol{z}_{kt})ds\right\}}{\exp\left\{\beta\int \overline{V}^M_{t+1}(s)m_0(s|\boldsymbol{z}_{kt})ds\right\} + \hat{\Gamma}^M_{mt}(z_{kt})} + o_p(1)$$

Analogous bounds can be formed for the inclusive value functions of the teacher-optimal stable match. We thus have that:

$$\hat{\Gamma}^M_{wt} \geq \hat{\Psi}^M_{wt}[\hat{\boldsymbol{\Gamma}}^M] + o_p(1) \quad \text{and} \quad \hat{\Gamma}^M_{mt} \leq \hat{\Psi}^M_{mt}[\hat{\boldsymbol{\Gamma}}^M] + o_p(1)$$

$$\hat{\Gamma}^W_{wt} \leq \hat{\Psi}^W_{wt}[\hat{\boldsymbol{\Gamma}}^W] + o_p(1) \quad \text{and} \quad \hat{\Gamma}^W_{mt} \geq \hat{\Psi}^W_{mt}[\hat{\boldsymbol{\Gamma}}^W] + o_p(1)$$

Given that $\hat{\Psi}_{wt}[\boldsymbol{\Gamma}]$ and $\hat{\Psi}_{mt}[\boldsymbol{\Gamma}]$ are nonincreasing and Lipschitz continuous in $\boldsymbol{\Gamma}$, we have:

$$\hat{\Gamma}^M_{wt} \geq \hat{\Psi}^M_{wt}[\hat{\boldsymbol{\Gamma}}^M] + o_p(1) \geq \hat{\Psi}^M_{wt}[\hat{\boldsymbol{\Psi}}^M[\hat{\boldsymbol{\Gamma}}^M]] + o_p(1)$$

Thus for any $\boldsymbol{\Gamma}^*$ solving the fixed point problem:

$$\Gamma^*_{wt} = \hat{\Psi}_{wt}[\boldsymbol{\Gamma}^*] + o_p(1) \quad \text{and} \quad \Gamma^*_{mt} = \hat{\Psi}_{mt}[\boldsymbol{\Gamma}^*] + o_p(1)$$

we thus have:

$$\hat{\Gamma}^M_{wt} \geq \Gamma^*_{wt} + o_p(1) \quad \text{and} \quad \hat{\Gamma}^M_{mt} \leq \Gamma^*_{mt} + o_p(1)$$

However, we know that the mapping $\hat{\Psi}$ is a contraction in logs, which means that it has a unique fixed point $\mathbf{\Gamma}^*$. In addition, the school-optimal stable match is unanimously preferred by schools while the teacher-optimal stable match is unanimously preferred by teachers (Roth and Sotomayor (1992)). This implies that $M_{it}(\mu^M) \subset M_{it}(\mu^*) \subset M_{it}(\mu^W)$ and $W_{it}(\mu^W) \subset W_{it}(\mu^*) \subset W_{it}(\mu^M)$ which means that for all i and j:

$$I_{wit}^M \leq I_{wit}^* \leq I_{wit}^W \quad \text{and} \quad I_{mjt}^W \leq I_{mjt}^* \leq I_{mjt}^M$$

This in turn implies that for all (x, z):

$$\hat{\Gamma}_{wt}^M(x) \leq \hat{\Gamma}_{wt}^*(x) \leq \hat{\Gamma}_{wt}^W(x) \quad \text{and} \quad \hat{\Gamma}_{mt}^W(z) \leq \hat{\Gamma}_{mt}^*(z) \leq \hat{\Gamma}_{mt}^M(z)$$

which implies that:

$$\Gamma_{wt}^* + o_p(1) \geq \hat{\Gamma}_{wt}^W \geq \hat{\Gamma}_{wt}^M \geq \Gamma_{wt}^* + o_p(1)$$

$$\Gamma_{mt}^* + o_p(1) \leq \hat{\Gamma}_{mt}^W \leq \hat{\Gamma}_{mt}^M \leq \Gamma_{mt}^* + o_p(1)$$

which in turn implies that:

$$\hat{\Gamma}_{wt}^M = \Gamma_{wt}^* + o_p(1) \quad \text{and} \quad \hat{\Gamma}_{mt}^M = \Gamma_{mt}^* + o_p(1)$$

$$\hat{\Gamma}_{wt}^W = \Gamma_{wt}^* + o_p(1) \quad \text{and} \quad \hat{\Gamma}_{mt}^W = \Gamma_{mt}^* + o_p(1)$$

Combining this with Lemma 3, this gives us for all $i = 1, ..., n_w$ and all $j = 1, ..., n_m$:

$$I_{wit}^M = \Gamma_{wt}^* + o_p(1) \quad \text{and} \quad I_{mjt}^M = \Gamma_{mt}^* + o_p(1)$$

$$I_{wit}^W = \Gamma_{wt}^* + o_p(1) \quad \text{and} \quad I_{mjt}^M = \Gamma_{mt}^* + o_p(1)$$

Note that given that inclusive value functions that would arise under any stable match μ_t^* defined as I_{wit}^* and I_{mjt}^* are such that $I_{wit}^M \leq I_{wit}^* \leq I_{wit}^W$ and $I_{mjt}^M \geq I_{mjt}^* \geq I_{mjt}^W$ the equality written above holds also for any I_{wit}^* and I_{mjt}^*.

I have shown that inclusive values can be approximated by the solution of the finite sample fixed point problem. Lemma 5 finally implies that the solution of the finite sample fixed point problem converges toward the solution of its population equivalent. This proves Theorem 1.(iii).

3.E Monte Carlo Simulations

To gain confidence in the validity of the theoretical results described in Section 3.4 and 3.5, I perform two Monte Carlo exercises. First, I simulate data from a market with different numbers of participating agents to verify whether empirical matching frequencies converge to their theoretical limit. Then, I then evaluate the performance of the Maximum Likelihood Estimator proposed in 3.5. I consider a market with three periods $T = 3$, normalize $\gamma_w = \gamma_m = 0$, and set $\beta_w = \beta_m = 0.9$. I then specify the flow payoffs as $U(x, z; \boldsymbol{\theta}) = \theta_1 + \theta_2 z$ and $V(x, z; \boldsymbol{\theta}) = \theta_1 + \theta_3 x$ for all t and set $\boldsymbol{\theta} = (1, 1, 1)$. I assume that $x_{i1} \sim \mathcal{N}(0, 1)$ and $z_{j1} \sim \mathcal{N}(0, 1)$. I assume the following laws of motion for x and z:

$$x_{it+1} = \begin{cases} x_{i1} + 1 & \text{if} \quad \mu_{wt}(i) \neq 0 \\ x_{i1} & \text{if} \quad \mu_{wt}(i) = 0 \end{cases} \quad , \quad z_{jt+1} = \begin{cases} z_{j1} + 1 & \text{if} \quad \mu_{mt}(j) \neq 0 \\ z_{j1} & \text{if} \quad \mu_{mt}(j) = 0 \end{cases}$$

This simulates a setting where teachers and schools become less attractive when they stay unmatched.

3.E.1 Convergence of Matching Frequencies

In this Monte Carlo exercise, I simulate data from the DGP described above for different market sizes indexed by n. In order to simulate the equilibrium, I first solve the fixed point problem described in Equation 3.4 to recover Γ^*_{wt} and Γ^*_{mt} and solve recursively for $\overline{U}_{t+1}$ and $\overline{V}_{t+1}$ for $t = \{1, 2\}$. I then draw a set of taste shocks ϵ_{ijt} and η_{ijt} for each period and each teacher-school pair and construct the lifetime utilities U_{ijt} and V_{ijt}. I then use the Deferred Acceptance algorithm to recover the teacher-optimal stable match in each period. The goal of this exercise is to evaluate whether the observed matching frequencies converge to their limit. More specifically I will look at whether the share of unmatched teachers in each period converges to its limit. Table 3.E.1 shows the results of this exercise. We can clearly see that as the size of the market increases, the share of unmatched teachers observed in the simulated data converges to its limit, which is displayed in the bottom line. This shows that, even with moderate sample sizes, the limit economy seems to be a relatively good approximation for the finite economy.

3.E.2 Estimation

In this second experiment, I simulate data by following the same procedure for different values of n. I then estimate $\boldsymbol{\theta}$ using the procedure described in Section 3.5. Table 3.E.2 shows that the estimator is unbiased even with small sample sizes. It is also consistent given that the standard deviation of the estimator decreases as the sample sizes increases.

Table 3.E.1: Monte Carlo: Share of Unmatched teachers

n	$t = 1$	$t = 2$	$t = 3$
20	0.2600	0.1870	0.2675
50	0.2439	0.1734	0.2447
100	0.2389	0.1640	0.2339
200	0.2314	0.1565	0.2237
500	0.2263	0.1509	0.2147
1000	0.2228	0.1469	0.2095
2000	0.2206	0.1432	0.2053
Model	0.2076	0.1384	0.1965

NOTES. This table reports the average share of unmatched schools and teachers in each period taken over 200 sample draws for different sample sizes n.

Table 3.E.2: Monte Carlo: MLE

n	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$
20	0.952	0.986	1.023
	(0.475)	(0.352)	(0.334)
50	0.962	0.988	1.010
	(0.292)	(0.223)	(0.204)
100	0.969	0.994	1.007
	(0.192)	(0.156)	(0.140)
200	0.977	0.991	1.003
	(0.133)	(0.104)	(0.105)
500	0.984	0.994	1.003
	(0.088)	(0.067)	(0.063)
1000	0.992	0.995	1.002
	(0.060)	(0.047)	(0.046)
True value	1	1	1

NOTES. This table reports the average and standard deviation of the ML estimator of $\boldsymbol{\theta}$ over 500 sample draws for different sample sizes n.

3.F Alternative Model: Irreversible Matches

In this section, I present an alternative to the model discussed in Section 3.4. I consider a setting where matches are irreversible and the match is stable in each period given agents' continuation value of staying unmatched, as in Doval (2022), and show that all the results derived in this paper extend.

3.F.1 Model

In this model, in each period t, agents can either decide to form a match with an agent from the other side or decide to stay unmatched and wait to get better opportunities in period $t+1$. The timing works as follows:

Period 1: The set of teachers $\mathcal{I}_1$ and schools $\mathcal{J}_1$ arrive in the market. A matching μ_1 occurs and all teachers $i \in \mathcal{I}_1$ that stay unmatched such that $\mu_{w1}(i) = 0$ move on to the second period. Similarly, all schools $j \in \mathcal{J}_1$ which choose to leave their slot empty such that $\mu_{m1}(j) = 0$ move on to the second period. I define the set of teachers that choose to stay unmatched in period 1 as $\mathcal{I}_1^0(\mu)$. Similarly I define the set of schools that choose to leave their vacancy empty as $\mathcal{J}_1^0(\mu)$.

Period t: The set of teachers $\mathcal{I}_t$ and schools $\mathcal{J}_t$ arrive in the market along with the teachers that chose to stay unmatched in the previous period $\mathcal{I}_{t-1}^0(\mu)$ and the schools that chose to keep their slots empty in the previous period $\mathcal{J}_{t-1}^0(\mu)$. We define the set of teachers available in period t as $\mathcal{I}_t(\mu) = \mathcal{I}_t \cup \mathcal{I}_{t-1}^0(\mu)$ and the set of school available in period t as $\mathcal{J}_t(\mu) = \mathcal{J}_t \cup \mathcal{J}_{t-1}^0(\mu)$. A matching μ_t occurs and all teachers $i \in \mathcal{I}_t(\mu)$ such that $\mu_{wt}(i) = 0$ and schools $j \in \mathcal{J}_t(\mu)$ such that $\mu_{mt}(j) = 0$ participate in the next period. I define the set of teachers that choose to stay unmatched in period t as $\mathcal{I}_t^0(\mu)$. Similarly, I define the set of schools that choose to leave their vacancy empty as $\mathcal{J}_t^0(\mu)$.

Period T: The set of teachers $\mathcal{I}_T$ and schools $\mathcal{J}_T$ arrive in the market along with the teachers in $\mathcal{I}_{T-1}^0$ and the schools in $\mathcal{J}_{T-1}^0$. We define the set of teachers available in period T as $\mathcal{I}_T(\mu) = \mathcal{I}_T \cup \mathcal{I}_{T-1}^0(\mu)$ and the set of schools available in period T as $\mathcal{J}_T(\mu) = \mathcal{J}_T \cup \mathcal{J}_{T-1}^0(\mu)$. From there a matching μ_T occurs and all teachers and schools choosing the outside option at this stage stay unmatched forever. The resulting matching is defined by $\mu = (\mu_t)_{t=1}^T$.

Firms and teachers are characterized by their observed attributes which collapse into two vectors of random variables $\boldsymbol{x}_{it}$ and $\boldsymbol{z}_{jt}$. I assume that the observed state variables of the new entrants in period t are drawn from the probability distribution functions m_t° and w_t°. I then assume that state variables evolve exogenously according to the Markov transition distribution functions m and w. This implies that aggregate states according to the following rule:

$$w_{t+1}(\boldsymbol{x}, \boldsymbol{\mu}) = \int_{\mathcal{X}_t} w(\boldsymbol{x}|s) f_t(s, *) ds + w_{t+1}^\circ(\boldsymbol{x})$$

$$m_{t+1}(\boldsymbol{z}, \boldsymbol{\mu}) = \int_{\mathcal{Z}_t} m(\boldsymbol{z}|s) f_t(*, h) dh + w_{t+1}^\circ(\boldsymbol{x})$$

I define the lifetime utility that teacher i gets from being matched with school j in period t as:

$$U_{ijt} = U_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{jt}) + \sigma \eta_{ijt}$$

whereas the lifetime utility that school j gets from being matched with teacher i in period t is defined as:

$$V_{ijt} = V_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{jt}) + \sigma \epsilon_{ijt}$$

I then define the lifetime utility that teacher i gets from staying unmatched and that school j gets from leaving its slot empty in period t as U_{i0t} and V_{0jt}:

$$U_{i0t} = \sigma \max_{k=1,\dots,J} \eta_{i0,k} + \beta_w \int \overline{U}_{it+1}(\boldsymbol{x}_{it+1}) w(\boldsymbol{x}_{it+1}|\boldsymbol{x}_{it}) d\boldsymbol{x}_{it+1} - \beta_w \log(J)$$

$$V_{0jt} = \sigma \max_{k=1,\dots,J} \epsilon_{0j,k} + \beta_m \int \overline{V}_{jt+1}(\boldsymbol{z}_{jt+1}) m(\boldsymbol{z}_{jt+1}|\boldsymbol{z}_{jt}) d\boldsymbol{z}_{jt+1} - \beta_m \log(J)$$

I then assume that Assumption 1-4 (ii) hold. I simply adjust 4 (iii) as the law of motion for aggregate states is defined as above. I also slightly modify Assumption 3 (i) such that $|\mathcal{I}_t| = [\exp(\gamma_{wt})n]$, $|\mathcal{J}_t| = [\exp(\gamma_{mt})n]$.

3.F.2 Linking Primitives to Equilibrium Sorting

I follow the same steps as in Section 4.2. I define F for a given random matching μ_t from a finite economy indexed by n as follows:

$$F_{nt}(x_{it}, z_{jt}|\mu_t) = \frac{1}{n} \sum_{i \in \mathcal{I}_t(\mu)} \sum_{j \in \mathcal{J}_t(\mu)} \mathbb{P}(x_{it} \leq x, z_{jt} \leq z, \mu_{wt}(i) = j)$$

I then denote F_t the limit of the distribution function F_{nt} as the size of the market n grows to infinity. I also define the joint density of matched characteristics as f_t.

As in the standard setting, I define the opportunity set faced by a given teacher $i \in \mathcal{I}_t$ in period t given a match μ as:

$$M_{it}(\mu) = \{j \in \mathcal{J}_t : V_{ijt} \geq V_{\mu_{mt}(j)jt}\}$$

Similarly, I define the opportunity set of school $j \in \mathcal{J}_t$ as:

$$W_{jt}(\mu) = \{i \in \mathcal{I}_t : U_{ijt} \geq U_{i\mu_{mt}(i)t}\}$$

The analogous of Proposition 1 follows directly from Assumption 4:

Proposition F.1 *Consider a match μ^* satisfying Assumption 4, for all $i \in \mathcal{I}_t$ and $j = \mathcal{J}_t$:*
(i) For all $t = 1, ..., T$:

$$U_{i\mu^*_{wt}(i)t} = \max_{k \in M_{it}(\mu^*) \cup \{0\}} U_{ikt} \quad \textit{and} \quad V_{\mu^*_{mt}(j)jt} = \max_{l \in W_{jt}(\mu^*) \cup \{0\}} V_{ljt}$$

(ii) Under Assumption 2, for all $t < T$:

$$\overline{U}_{it+1}(x) = \mathbb{E}_{\mathcal{S}_t}\left[\max_{k \in M_{it+1}(\mu^*) \cup \{0\}} U_{ikt+1} | x_{it+1} = x\right]$$

$$\overline{V}_{jt+1}(z) = \mathbb{E}_{\mathcal{S}_t}\left[\max_{l \in W_{jt+1}(\mu^*) \cup \{0\}} V_{ljt+1} | z_{jt+1} = z\right]$$

The proof is identical to the proof of Proposition 1. This result implies that an equilibrium match μ^* can be rewritten as the outcome of two dynamic discrete choice models where each agent's choice set is its opportunity set. However, each alternative, except the option of staying unmatched, is a terminating action.

I characterize the limit of conditional choice probabilities (CCPs) and expected future payoffs under arbitrary exogenous choice sets and by fixing the aggregate states distributions. I assume that $M_{it} = \{1, ..., J\}$ and $W_{jt} = \{1, ..., J\}$ for all t and I fix m_t and w_t for all t.

Proposition F.2 *Consider a given teacher $i \in \mathcal{I}_t$. Under Assumption 1-3 we have:*
(i) For all t, as $J \to \infty$:

$$J\mathbb{P}(U_{ijt} \geq U_{ikt}, k = \{0, 1, ..., J\} | \boldsymbol{x}_{it}, \boldsymbol{z}_{jt}) \longrightarrow \frac{\exp\Big\{U_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{jt})\Big\}}{\exp\Big\{\beta_w \int \overline{U}_{t+1}(s) w(s|\boldsymbol{x}_{it}) ds\Big\} + \int \exp\Big\{U_t(\boldsymbol{x}_{it}, h)\Big\} m_t(h) dh}$$

$$\mathbb{P}(U_{i0t} \geq U_{ikt}, k = \{0, 1, ..., J\} | \boldsymbol{x}_{it}) \longrightarrow \frac{\exp\Big\{\beta_w \int \overline{U}_{t+1}(s) w(s|\boldsymbol{x}_{it}) ds\Big\}}{\exp\Big\{\beta_w \int \overline{U}_{t+1}(s) w(s|\boldsymbol{x}_{it}) ds\Big\} + \int \exp\Big\{U_t(x_{it}, h)\Big\} m_t(h) dh}$$

(ii) For all t:

$$\overline{U}_{t+1}(x) = \log\Bigg(\exp\Big\{\beta_w \int \overline{U}_{t+2}(s) w_0(s|x) ds\Big\} + \int \exp\Big\{U_{t+1}(x, h)\Big\} m_{t+1}(h) dh\Bigg) + \gamma + o(1)$$

where $\gamma \approx 0.5772$ is Euler's constant. Again, the proof is identical to the proof of Proposition 2. The same result holds symmetrically for the school side.

I now introduce that opportunity sets are unobserved and endogenous and show that the implications of Proposition F.2 allow us to tackle both of these issues. Using the same argument as in the standard case, Proposition F.2 implies that: (i) the probability that school j rematches with a specific teacher i vanishes to zero as the size of opportunity sets increases to infinity and (ii) the probability of choosing the outside option instead is nondegenerate in the limit. This implies that the dependence between taste shocks and opportunity sets vanishes in the limit.

I now consider a sequence of school-optimal stable matches μ^M. As opportunity sets' endogeneity vanishes in the limit for extremal matchings, we can then use Proposition F.2

(i) to bound teachers' CCPs in period t, assuming that we would observe the corresponding opportunity set $M_{it}(\mu_t^M)$ and future expected payoff function $\overline{U}_{it+1}^M$:

$$n^{1/2}\mathbb{P}(U_{ijt} \geq \max_{k\in M_{it}(\mu_t^M)\cup\{0\}} U_{ikt}|x_{it}, z_{jt}, (z_{kt})_{k\in M_{it}(\mu_t^M)}, M_{i\tau}(\mu_t^M), \overline{U}_{it+1}^M) \tag{3.8}$$
$$\leq \frac{\exp\left\{U_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{jt})\right\}}{\exp\left\{\beta_w \int \overline{U}_{it+1}^M(s)w(s|\boldsymbol{x}_{it})ds\right\} + n^{-1/2}\sum_{k\in M_{it}(\mu_t^M)} \exp\left\{U_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{kt})\right\}} + o(1)$$

Similar bounds can be computed for a sequence of teacher-optimal stable match μ^W where the direction of the inequality is reversed. The same result also holds for the school side with the direction of the inequality reversed. Using Proposition F.2 (ii), we can also bound agents' expectations about their match payoff under a sequence of school-optimal stable matches μ^M as follows:

$$\overline{U}_{it}^M(x) \geq \log\left(\exp\left\{\beta \int \overline{U}_{it+1}^M(s)w(s|x)ds\right\} + n^{-1/2}\sum_{k\in M_{it}(\mu_t^M)} \exp\left\{U_t(x, \boldsymbol{z}_{kt})\right\}\right) + \gamma + o(1) \tag{3.9}$$

where again similar bounds can be computed for the teacher-optimal stable match and for the school side with the direction of the inequality reversed.

In Equations (3.8) and (3.9), $n^{-1/2}\sum_{k\in M_{it}(\mu_t^M)} \exp\left\{U_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{kt})\right\}$ serves as a sufficient statistic that collapses all the information contained in opportunity sets which is needed to approximate CCPs and expectations about future payoffs.

I define teacher i's inclusive value given a sequence of realized matches μ^* as:

$$I_{wit}^* = n^{-1/2}\sum_{k\in M_{it}(\mu_t^*)} \exp\left\{U_t(\boldsymbol{x}_{it}, \boldsymbol{z}_{kt})\right\}$$

Similarly, I define school j's inclusive value given μ^* as:

$$I_{mjt}^* = n^{-1/2}\sum_{l\in W_{jt}(\mu_t^*)} \exp\left\{V_t(\boldsymbol{x}_{lt}, \boldsymbol{z}_{jt})\right\}$$

I also define I_{wit}^M and I_{mjt}^M as the inclusive values that would arise under a sequence of school-optimal stable matches μ^M in period t and I_{wit}^W and I_{mjt}^W as the inclusive values that would arise under a sequence of teacher-optimal stable matches μ^W in period t.

Inclusive values arising from a sequence of school-optimal and teacher-optimal stable matches in a given period t can be approximated by *expected inclusive value functions.* I rewrite I^M_{wit} as:

$$
\begin{aligned}
I^M_{wit} &= \frac{1}{n}\sum_{k\in\mathcal{J}_t(\mu)} \exp\left\{U(\boldsymbol{x}_{it},\boldsymbol{z}_{kt})\right\} \times \sqrt{n}\mathbb{1}\{k\in M_{it}(\mu^M_t)\} \\
&= \frac{1}{n}\sum_{k\in\mathcal{J}_t(\mu)} \exp\left\{U(\boldsymbol{x}_{it},\boldsymbol{z}_{kt})\right\}\sqrt{n}\mathbb{1}\{V_{ikt} \geq \max_{l\in W_{kt}(\mu^M_t)\cup\{0\}} V_{lkt}\}
\end{aligned}
$$

The inclusive value of a given teacher is determined by the set of schools that would accept her, which in turn depends on the preferences of all schools as well as their opportunity sets. Using the school analogous of Equation (3.1), I thus show that:

$$
I^M_{wit} \geq \hat{\Gamma}^M_{wt}(x_{it}) + o_p(1) \quad \text{and} \quad I^M_{mjt} \leq \hat{\Gamma}^M_{mt}(z_{jt}) + o_p(1)
$$

where $\hat{\Gamma}^M_{wt}$ and $\hat{\Gamma}^M_{mt}$ are the school-optimal expected inclusive value function of teachers and schools in period t which are defined as:

$$
\hat{\Gamma}^M_{wt}(x_{it}) = \frac{1}{n}\sum_{k\in\mathcal{J}_t(\mu)} \frac{\exp\left\{U_t(\boldsymbol{x}_{it},\boldsymbol{z}_{kt}) + V_t(\boldsymbol{x}_{it},\boldsymbol{z}_{kt})\right\}}{\exp\left\{\beta\int \overline{V}^M_{kt+1}(s)m(s|\boldsymbol{z}_{kt})ds\right\} + I^M_{mkt}}
$$

$$
\hat{\Gamma}^M_{mt}(z_{jt}) = \frac{1}{n}\sum_{l\in\mathcal{I}_t(\mu)} \frac{\exp\left\{U_t(\boldsymbol{x}_{lt},\boldsymbol{z}_{jt}) + V_t(\boldsymbol{x}_{lt},\boldsymbol{z}_{jt})\right\}}{\exp\left\{\beta\int \overline{U}^M_{lt+1}(s)w(s|\boldsymbol{x}_{lt})ds\right\} + I^M_{wlt}}
$$

where I define $\overline{U}^M_{it+1}$ and $\overline{V}^M_{jt+1}$ as follows:

$$
\overline{U}^M_{it+1}(x) = \log\left(\exp\left\{\beta\int \overline{U}^M_{it+2}(s)w_0(s|x)ds\right\} + I^M_{wit+1}\right)
$$

$$
\overline{V}^M_{jt+1}(z) = \log\left(\exp\left\{\beta\int \overline{V}^M_{jt+2}(s)m_0(s|z)ds\right\} + I^M_{mjt+1}\right)
$$

Note that similar bounds can be established for the inclusive values that would arise under the teacher-optimal stable match:

$$I_{wit}^{W} \leq \hat{\Gamma}_{wt}^{W}(x_{it}) + o_p(1) \quad \text{and} \quad I_{mjt}^{W} \geq \hat{\Gamma}_{mt}^{W}(z_{jt}) + o_p(1)$$

The proof follows the same steps as the proof of Lemma 4 in Appendix 3.D.4.

The rest of the proof entails characterizing the fixed point problem and showing that inclusive values arising from an equilibrium match μ^* can be approximated by its solution. I define the fixed point mappings as follows:

$$\hat{\Psi}_{wt}[\mathbf{\Gamma}](x) = \frac{1}{n} \sum_{k \in \mathcal{J}_t(\mu)} \frac{\exp\left\{U_t(x, \boldsymbol{z}_{kt}) + V_t(x, \boldsymbol{z}_{kt})\right\}}{\exp\left\{\beta \int \overline{V}_{t+1}[\mathbf{\Gamma}](s) m(s|\boldsymbol{z}_{kt}) ds\right\} + \Gamma_{mt}(z_{kt})}$$

$$\hat{\Psi}_{mt}[\mathbf{\Gamma}](z) = \frac{1}{n} \sum_{l \in \mathcal{I}_t(\mu)} \frac{\exp\left\{U_t(\boldsymbol{x}_{lt}, z) + V_t(\boldsymbol{x}_{lt}, z)\right\}}{\exp\left\{\beta \int \overline{U}_{t+1}[\mathbf{\Gamma}](s) w(s|\boldsymbol{x}_{lt}) ds\right\} + \Gamma_{wt}(x_{lt})}$$

$$\overline{U}_{t+1}[\mathbf{\Gamma}](x) = \log\left(\exp\left\{\beta \int \overline{U}_{t+2}[\mathbf{\Gamma}](s) w_0(s|x) ds\right\} + \Gamma_{wt+1}(x)\right)$$

$$\overline{V}_{t+1}[\mathbf{\Gamma}](z) = \log\left(\exp\left\{\beta \int \overline{V}_{t+2}[\mathbf{\Gamma}](s) m_0(s|z) ds\right\} + \Gamma_{mt+1}(z)\right)$$

For a given equilibrium match μ^*, for any $x \in \mathcal{X}$ and $z \in \mathcal{Z}$ in each period t:

$$\hat{\Gamma}_{wt}^{*}(x) = \hat{\Psi}_{wt}[\hat{\mathbf{\Gamma}}^{*}](x) + o_p(1) \quad \text{and} \quad \hat{\Gamma}_{mt}^{*}(z) = \hat{\Psi}_{mt}[\hat{\mathbf{\Gamma}}^{*}](z) + o_p(1) \tag{3.10}$$

meaning that inclusive values in period t arising from an equilibrium match μ^* can be approximated by fixed points of the mappings $\hat{\Psi}_{wt}, \hat{\Psi}_{mt}$. To characterize the limit of inclusive values, I then consider the limit version of this fixed point problem:

$$\Gamma_{wt} = \Psi_{wt}[\mathbf{\Gamma}] \quad \text{and} \quad \Gamma_{mt} = \Psi_{mt}[\mathbf{\Gamma}] \quad \forall t \tag{3.11}$$

where

$$\Psi_{wt}[\mathbf{\Gamma}](x) = \int \frac{\exp\left\{U_t(x, h) + V_t(x, h)\right\}}{\exp\left\{\beta \int \overline{V}_{t+1}[\mathbf{\Gamma}](s) m(s|h) ds\right\} + \Gamma_{mt}(h)} m_t[\mathbf{\Gamma}](h) dh$$

$$\Psi_{mt}[\mathbf{\Gamma}](z) = \int \frac{\exp\left\{U_t(h,z) + V_t(h,z)\right\}}{\exp\left\{\beta \int \overline{U}_{t+1}[\mathbf{\Gamma}](s) w(s|h) ds\right\} + \Gamma_{wt}(h)} w_t[\mathbf{\Gamma}](h) dh$$

$$\overline{U}_{t+1}[\mathbf{\Gamma}](x) = \log\left(\exp\left\{\beta \int \overline{U}_{t+2}[\mathbf{\Gamma}](s) w_0(s|x) ds\right\} + \Gamma_{wt+1}(x)\right)$$

$$\overline{V}_{t+1}[\mathbf{\Gamma}](z) = \log\left(\exp\left\{\beta \int \overline{V}_{t+2}[\mathbf{\Gamma}](s) m_0(s|z) ds\right\} + \Gamma_{mt+1}(z)\right)$$

$$w_t[\mathbf{\Gamma}](x) = \int_{\mathcal{X}_t} w(x|s) f_{t-1}[\mathbf{\Gamma}](s,*) ds + w_t^\circ(x)$$

$$m_t[\mathbf{\Gamma}](z) = \int_{\mathcal{Z}_t} m(z|s) f_{t-1}[\mathbf{\Gamma}](*,h) dh + m_t^\circ(z)$$

The final step of the proof shows that this population fixed point problem has a unique solution and that the approximate solution of the finite sample fixed point problem converges to it. This is stated in the following result:

Theorem F.1 *Under Assumption 1-4:*
(i) The mapping $(\log \mathbf{\Gamma}_{\boldsymbol{w}}, \log \mathbf{\Gamma}_{\boldsymbol{m}}) \mapsto (\log \mathbf{\Psi}_{\boldsymbol{m}}[\mathbf{\Gamma}], \log \mathbf{\Psi}_{\boldsymbol{w}}[\mathbf{\Gamma}])$ *is a contraction.*
(ii) The fixed point problem described in Equation (3.11) always has a unique solution $\mathbf{\Gamma}^*_{\boldsymbol{m}}, \mathbf{\Gamma}^*_{\boldsymbol{w}}$.
(iii) For any equilibrium μ^*, $I^*_{wit} \longrightarrow \Gamma^*_{wt}(x_{it})$ *and* $I^*_{mjt} \longrightarrow \Gamma^*_{mt}(z_{jt})$ *for all* i, j *and* t.

The complete proof of this result can be found in Appendix 3.F.4. Finally, from Theorem F.1 and Proposition F.2, we can fully characterize analytically the equilibrium of the model as a function of teachers' and schools' payoff functions. The limit joint density of matched characteristics f_t can be derived from the limit of conditional choice probabilities and has the following expression:

$$\frac{f_t(x,z)}{w_t(x)m_t(z)} = \frac{\exp\left\{U_t(x,z) + V_t(x,z) + \gamma_{wt} + \gamma_{mt}\right\}}{\left(\exp\left\{\beta \int \overline{U}^*_{t+1}(s) w(s|x) ds\right\} + \Gamma^*_{wt}(x)\right)\left(\exp\left\{\beta \int \overline{V}^*_{t+1}(s) m(s|z) ds\right\} + \Gamma^*_{mt}(z)\right)}$$

$$\frac{f_t(x,*)}{w_t(x)} = \frac{\exp\left\{\beta \int \overline{U}^*_{t+1}(s) w_0(s|x) ds + \gamma_{wt}\right\}}{\left(\exp\left\{\beta \int \overline{U}^*_{t+1}(s) w(s|x) ds\right\} + \Gamma^*_{wt}(x)\right)}$$

$$\frac{f_t(*,z)}{m_t(z)} = \frac{\exp\left\{\beta\int \overline{V}^*_{t+1}(s)m_0(s|z)ds + \gamma_{mt}\right\}}{\left(\exp\left\{\beta\int \overline{V}^*_{t+1}(s)m(s|z)ds\right\} + \Gamma^*_{mt}(z)\right)}$$

where $f_t(x,*)$ and $f_t(*,z)$ are, respectively, the density of the characteristics of unmatched teachers and unmatched schools. I define the equilibrium expected future payoff functions $\overline{U}^*_{t+1}$ and $\overline{V}^*_{t+1}$ recursively as:

$$\overline{U}^*_{t+1}(x) = \log\left(\exp\left\{\beta\int \overline{U}^*_{t+2}(s)w(s|x)ds\right\} + \Gamma^*_{wt+1}(x)\right)$$

$$\overline{V}^*_{t+1}(z) = \log\left(\exp\left\{\beta\int \overline{V}^*_{t+2}(s)m(s|z)ds\right\} + \Gamma^*_{mt+1}(z)\right)$$

and the equilibrium aggregate states distribution w^*_t and m^*_t as:

$$w^*_t(x) = \int_{\mathcal{X}_t} w(x|s)f_{t-1}[\mathbf{\Gamma}](s,*)ds + w^\circ_t(x)$$

$$m^*_t(z) = \int_{\mathcal{Z}_t} m(z|s)f_{t-1}[\mathbf{\Gamma}](*,h)dh + m^\circ_t(z)$$

3.F.3 Identification

The identification strategy follows the same steps as Section 3.5. I thus fix the value of the discount factors and consider two polar cases: (i) $T < \infty$ and nonstationarity and (ii) $T = \infty$ and stationarity.

Finite horizon

The identification argument in the finite horizon case can be done by backward induction. Starting from the last period T, we can identify the joint surplus as follows:

$$U_T(x,z) + V_T(x,z) = \log\left(\frac{f_T(x,z)}{f_T(x,*)f_T(*,z)}\right)$$

We can also identify Γ^*_{wT} and Γ^*_{mT} from the distribution of unmatched teachers and schools:

$$\Gamma^*_{wT}(x) = \frac{w_T(x)\exp(\gamma_{wT})}{f_T(x,*)} - 1$$

$$\Gamma^*_{mT}(z) = \frac{m_T(z)\exp(\gamma_{mT})}{f_T(*, z)} - 1$$

$\overline{U}_T$ and $\overline{V}_T$ can then be computed by backward induction:

$$\overline{U}_T(x) = \log(1 + \Gamma^*_{wT}(x)) + \gamma$$

$$\overline{V}_T(z) = \log(1 + \Gamma^*_{mT}(z)) + \gamma$$

From there, we can then repeat the same steps to identify the inclusive value functions and the joint surplus in period $T-1$. Finally, we iterate the procedure to identify the joint surplus and the inclusive value functions in all periods t. This results in the following proposition.

Proposition F.3 *Under Assumption 1-4 and for $T < \infty$:*
(i) The joint surplus function $U_t + V_t$ and the inclusive value functions Γ^*_{wt} and Γ^*_{mt} are identified for all t from f_t, the limiting joint distribution of matched characteristics in period t.
(ii) Without further restrictions, we cannot separately identify U_t and V_t for all t.

Infinite horizon

To allow for $T = \infty$, I impose Assumption 5 which implies $\Gamma_{mt} = \Gamma_m$ and $\Gamma_{wt} = \Gamma_w$ for all t, $\overline{U}_t = \overline{U}$ and $\overline{V}_t = \overline{V}$. This implies that we can write:

$$\begin{aligned} \frac{f(x,*)}{w(x)} &= \frac{\exp\left\{\beta \int \overline{U}^*(s) w_0(s|x) ds\right\}}{\left(\exp\left\{\beta \int \overline{U}^*(s) w_0(s|x) ds\right\} + \Gamma^*_w(x)\right)} \\ &= \frac{\exp\left\{\beta \int \overline{U}^*(s) w_0(s|x) ds\right\}}{\exp\left\{\overline{U}^*(x) - \gamma\right\}} = \exp\left\{\beta \int \overline{U}^*(s) w_0(s|x) ds - \overline{U}^*(x) + \gamma\right\} \end{aligned}$$

From there, we can invert this mapping to recover $\overline{U}^*$. We can follow the same steps to recover $\overline{V}$ from $f(*, z)$. It is then immediate to see that we can identify $U + V$ from $f(x, z)$.

Proposition F.4 *Under Assumption 1-5 and for $T = \infty$:*
(i). The joint surplus function $U + V$ and the inclusive value functions Γ^*_w and Γ^*_m are

identified from the limiting joint distribution of matched characteristics in each period f.
(ii). Without further restrictions, we cannot separately identify U and V.

3.F.4 Proof Theorem F.1

I will start by proving part (i) of Theorem F.1. A first step is to restrict the space of functions in which the solutions to the fixed point problem described in Equation 3.11 can belong to. Namely, I will start by showing to we can restrict ourselves to a Banach space of continuous functions.

We start by constructing bounds for the solutions of this fixed point problem. Note that for all t, we can see that $\Psi_{wt}[\boldsymbol{\Gamma_m}](x) \geq 0$ and $\Psi_{mt}[\boldsymbol{\Gamma_w}](x) \geq 0$ for all (x, z) which implies that the solutions of this fixed point problem must be bounded from below by 0. To construct an upper bound we first need to construct a lower bound on $\overline{U}_t$ and $\overline{V}_t$. We proceed by backward induction. We know that $\overline{U}_T(x) = \log(1 + \Gamma_{wT}(x)) + \gamma$ which implies that $\overline{U}_T(x) \geq \gamma$ for all x. Iterating this procedure, we can then show that $\overline{U}_{T-1}(x) \geq \gamma(1 + \beta_w)$ and more generally that $\overline{U}_{t+1}(x) \geq \gamma \sum_{\tau=0}^{T-t} \beta_w^\tau$ and $\overline{V}_{t+1}(z) \geq \gamma \sum_{\tau=0}^{T-t} \beta_m^\tau$ for all (x, z). We also know from Assumption 1, that U_t and V_t are bounded from above. We can thus show that

$$\Psi_{wt}[\boldsymbol{\Gamma_m}](x) \leq \frac{\exp\{\overline{U_t} + \overline{V_t}\}}{\gamma \sum_{\tau=1}^{T-t} \beta_m^{\tau-1}} \quad \forall x \in \mathcal{X}$$

$$\Psi_{mt}[\boldsymbol{\Gamma_w}](z) \leq \frac{\exp\{\overline{U_t} + \overline{V_t}\}}{\gamma \sum_{\tau=1}^{T-t} \beta_w^{\tau-1}} \quad \forall z \in \mathcal{Z}$$

To prove continuity of the mappings $\Psi_{wt}[\boldsymbol{\Gamma_m}]$ and $\Psi_{mt}[\boldsymbol{\Gamma_w}]$ we proceed by backward induction. Starting from $t = T$, we can rewrite $\Psi_{wT}[\boldsymbol{\Gamma_m}]$ as:

$$\Psi_{wT}[\boldsymbol{\Psi_m}[\boldsymbol{\Gamma_w}]](x) = \int \frac{\exp\{U_T(x, s) + V_T(x, s)\}}{1 + \int \frac{\exp\{U_T(t,s)+V_T(t,s)\}}{1+\Gamma_{wT}(t)} w_T[\boldsymbol{\Gamma_w}](t)dt} m_T[\boldsymbol{\Psi_m}[\boldsymbol{\Gamma_w}]](s)ds$$

which shows that continuity of the solution of $\Gamma_{wT} = \Psi_{wT}[\boldsymbol{\Psi_m}[\boldsymbol{\Gamma_w}]]$ follows directly from continuity of U_T and V_T as stated in Assumption 1. From there we can infer that $\overline{U}_T(x)$ is also continuous and we know that it is a non negative function which implies that $\Psi_{wT-1}[\boldsymbol{\Gamma_m}]$ will also be continuous. We can then iterate this argument to prove that the solutions of the fixed point problem described in Equation 3.11 must be continuous and bounded functions.

We now turn to the proof that the mapping $(\log \boldsymbol{\Gamma_w}, \log \boldsymbol{\Gamma_m}) \mapsto (\log \boldsymbol{\Psi_m}[\boldsymbol{\Gamma_w}], \log \boldsymbol{\Psi_w}[\boldsymbol{\Gamma_m}])$

is a contraction. We will start by showing that for alternative sets of functions $\boldsymbol{\Gamma_m} = (\Gamma_{mt})_{t=1}^T$ and $\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}} = (\tilde{\Gamma}_{mt})_{t=1}^T$, there always exist a constant $\lambda < 1$ such that:

$$\Big|\Big| \log \boldsymbol{\Psi_w}[\boldsymbol{\Gamma_m}] - \log \boldsymbol{\Psi_w}[\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}}]\Big|\Big|_\infty \leq \lambda \Big|\Big| \log \boldsymbol{\Gamma_m} - \log \tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}}\Big|\Big|_\infty$$

The mean value inequality for vector valued functions defined on Banach spaces implies that:

$$\Big|\Big| \log \boldsymbol{\Psi_w}[\boldsymbol{\Gamma_m}](x) - \log \boldsymbol{\Psi_w}[\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}}](x)\Big|\Big|_\infty \leq \sup_{a\in[0,1]} \Big|\Big| D\log \boldsymbol{\Psi_w}[a \log \boldsymbol{\Gamma_m} + (1-a)\log \tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}}](x)\Big|\Big|_\infty \Big|\Big| \log \boldsymbol{\Gamma_m}(x) - \log \tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}}(x)\Big|\Big|_\infty$$

where $D\log \boldsymbol{\Psi_w}$ are the Gateaux derivatives of $\log \boldsymbol{\Psi_w}$. I will thus characterize and bound the following object for any $t \in [0,1]$ and any $x \in \mathcal{X}$:

$$D\log \boldsymbol{\Psi_w}[a \log \boldsymbol{\Gamma_m} + (1-a)\log \tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}}](x)$$

Note first that we can rewrite $\log \Psi_{wt}[\log \boldsymbol{\Gamma_m}](x)$ as:

$$\log \int \frac{\exp\left\{U_t(x,h) + V_t(x,h)\right\}}{\exp\left\{\beta \int \overline{V}_{t+1}[\log \boldsymbol{\Gamma_m}](s) m(s|h) ds\right\} + \exp\{\log \Gamma_{mt}(h)\}} m_t[\log \boldsymbol{\Gamma_m}](h) dh$$

where

$$m_t[\log \boldsymbol{\Gamma_m}](z) = \int_{\mathcal{Z}_t} m(z|s) f_{t-1}[\log \boldsymbol{\Gamma_m}](*,h) dh + m_t^\circ(z)$$

$$\overline{V}_{t+1}[\log \boldsymbol{\Gamma_m}](z) = \log\left(\exp\left\{\beta \int \overline{V}_{t+2}[\log \boldsymbol{\Gamma_m}](s) m(s|z) ds\right\} + \Gamma_{mt+1}(z)\right)$$

$$f_{t-1}[\log \boldsymbol{\Gamma_m}](*,z) = \frac{\exp\left\{\beta \int \overline{V}_t^*(s) m(s|z) ds\right\}}{\left(\exp\left\{\beta \int \overline{V}_t^*(s) m(s|z) ds\right\} + \Gamma_{mt-1}^*(z)\right)} m_{t-1}[\log \boldsymbol{\Gamma_m}](z)$$

Using the same steps as in the proof of Theorem 1 (i), we can show that:

$$\sup_{a\in[0,1]} \Big|\Big| D\log \boldsymbol{\Psi_w}[a \log \boldsymbol{\Gamma_m} + (1-a)\log \tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}}](x)\Big|\Big|_\infty < 1$$

which implies that for any alternative sets of functions $\boldsymbol{\Gamma_m} = (\Gamma_{mt})_{t=1}^T$ and $\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}} = (\tilde{\Gamma}_{mt})_{t=1}^T$

there always exist a constant $\lambda < 1$ such that:

$$\Big|\Big|\log \boldsymbol{\Psi}_{\boldsymbol{w}}[\boldsymbol{\Gamma}_{\boldsymbol{m}}] - \log \boldsymbol{\Psi}_{\boldsymbol{w}}[\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}}]\Big|\Big|_{\infty} \leq \lambda \Big|\Big|\log \boldsymbol{\Gamma}_{\boldsymbol{m}} - \log \tilde{\boldsymbol{\Gamma}}_{\boldsymbol{m}}\Big|\Big|_{\infty}$$

Symmetrical arguments can be applied to find that there for any alternative sets of functions $\boldsymbol{\Gamma}_{\boldsymbol{w}} = (\Gamma_{wt})_{t=1}^{T}$ and $\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{w}} = (\tilde{\Gamma}_{wt})_{t=1}^{T}$ always exist a constant $\lambda < 1$ such that:

$$\Big|\Big|\log \boldsymbol{\Psi}_{\boldsymbol{m}}[\boldsymbol{\Gamma}_{\boldsymbol{w}}] - \log \boldsymbol{\Psi}_{\boldsymbol{m}}[\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{w}}]\Big|\Big|_{\infty} \leq \lambda \Big|\Big|\log \boldsymbol{\Gamma}_{\boldsymbol{w}} - \log \tilde{\boldsymbol{\Gamma}}_{\boldsymbol{w}}\Big|\Big|_{\infty}$$

This concludes the proof of part (i) of Theorem F.1 and shows that the mapping $(\log \boldsymbol{\Gamma}_{\boldsymbol{w}}, \log \boldsymbol{\Gamma}_{\boldsymbol{m}}) \mapsto (\log \boldsymbol{\Psi}_{\boldsymbol{m}}[\boldsymbol{\Gamma}_{\boldsymbol{w}}], \log \boldsymbol{\Psi}_{\boldsymbol{w}}[\boldsymbol{\Gamma}_{\boldsymbol{m}}])$ is a contraction.

Part (ii) of Theorem F.1 directly follows from part (i) and from the Banach fixed point theorem. Part (iii) follows from the same steps as the proof of Theorem 1 (iii).

References

Abdulkadiroğlu, Atila, Nikhil Agarwal, and Parag A Pathak, "The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match," *American Economic Review*, 2017, *107* (12), 3635–89.

Agarwal, Nikhil, "An Empirical Model of the Medical Match," *American Economic Review*, 2015, *105* (7), 1939–78.

_ , "Policy Analysis in Matching Markets," *American Economic Association P&P*, 2017, *107* (5), 246–50.

_ **and Eric Budish**, "Market Design," in "Handbook of Industrial Organization," Vol. 5, Elsevier, 2021, pp. 1–79.

_ **and Paulo J Somaini**, "Demand Analysis under Latent Choice Constraints," *NBER Working Paper 29993*, 2022.

_ **and Paulo Somaini**, "Revealed Preference Analysis of School Choice Models," *Annual Review of Economics*, 2020, *12*, 471–501.

_ **and** _ , "Empirical Models of Non-Transferable Utility Matching," in "Online and Matching-Based Market Design" Forthcoming.

_ **, Itai Ashlagi, Michael A Rees, Paulo Somaini, and Daniel Waldinger**, "Equilibrium Allocations Under Alternative Waitlist Designs: Evidence from Deceased Donor Kidneys," *Econometrica*, 2021, *89* (1), 37–76.

Ahn, Tom, Esteban Aucejo, and Jonathan James, "The Importance of Matching Effects for Labor Productivity: Evidence from Teacher-Student Interactions," Technical Report, California Polytechnic State University, Department of Economics 2021.

Ajzenman, Nicolás, Eleonora Bertoni, Gregory Elacqua, Luana Marotta, and Carolina Méndez Vargas, "Altruism or money? Reducing teacher sorting using behavioral strategies in Peru," 2021.

Araujo, M Caridad, Pedro Carneiro, Yyannú Cruz-Aguayo, and Norbert Schady, "Teacher quality and learning outcomes in kindergarten," *The Quarterly Journal of Economics*, 2016, *131* (3), 1415–1453.

Araujo, Maria Daniela, Guido Heineck, and Yyannu Cruz-Aguayo, "Does Test-Based Teacher Recruitment Work in the Developing World? Experimental Evidence from Ecuador," IZA Discussion Papers 13830, Institute of Labor Economics (IZA) October 2020.

Azevedo, Eduardo M and Jacob D Leshno, "A Supply and Demand Framework for Two-Sided Matching Markets," *Journal of Political Economy*, 2016, *124* (5), 1235–1268.

Baccara, Mariagiovanna and Leeat Yariv, "Dynamic Matching," *Online and Matching-Based Market Design*, 2021.

Barrera-Osorio, Felipe and Dhushyanth Raju, "Teacher performance pay: Experimental evidence from Pakistan," *Journal of Public Economics*, 2017, *148* (C), 75–91.

Bates, Michael D, Michael Dinerstein, Andrew C Johnston, and Isaac Sorkin, "Teacher Labor Market Equilibrium and Student Achievement," *NBER Working Paper 29728*, 2022.

Bau, Natalie and Jishnu Das, "Teacher value added in a low-income country," *American Economic Journal: Economic Policy*, 2020, *12* (1), 62–96.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 2004, *112* (1), 68–105.

Bertoni, Eleonora, Gregory Elacqua, Carolina Mendez, and Humberto Santos, "Teacher Hiring Instruments and Teacher Value Added: Evidence from Peru," IDB Publications (Working Papers) 11118, Inter-American Development Bank March 2021.

_ , _ , Diana Hincapie, Carolina Mendez, and Diana Paredes, "Teachers' Preferences for Proximity and the Implications for Staffing Schools: Evidence from Peru," IDB Publications (Working Papers) 01073, Inter-American Development Bank October 2019.

_ , _ , Luana Marotta, Matías Martinez, Humberto Santos, and Sammara Soares, "Is School Funding Unequal in Latin America? A Cross-Country Analysis. CEPA Working Paper No. 20-11.," *Stanford Center for Education Policy Analysis*, 2020.

Biasi, Barbara, Chao Fu, and John Stromme, "Equilibrium in the Market for Public School Teachers: District Wage Strategies and Teacher Comparative Advantage," NBER Working Papers 28530, National Bureau of Economic Research, Inc March 2021.

_ , _ , and _ , "Equilibrium in the Market for Public School Teachers: District Wage Strategies and Teacher Comparative Advantage," *NBER Working Paper 28530*, 2021.

Bo, Ernesto Dal, Frederico Finan, and Martin A. Rossi, "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service," *The Quarterly Journal of Economics*, 2013, *128* (3), 1169–1218.

Bobba, Matteo, Tim Ederer, Gianmarco Leon-Ciliotta, Christopher Neilson, and Marco G Nieddu, "Teacher Compensation and Structural Inequality: Evidence from Centralized Teacher School Choice in Perú," *NBER Working Paper 29068*, 2022.

Bold, Tessa, Deon Filmer, Gayle Martin, Ezequiel Molina, Brian Stacy, Christophe Rockmore, Jakob Svensson, and Waly Wane, "Enrollment without Learning: Teacher Effort, Knowledge, and Skill in Primary Schools in Africa," *Journal of Economic Perspectives*, November 2017, *31* (4), 185–204.

_ , Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur, "Experimental evidence on scaling up education reforms in Kenya," *Journal of Public Economics*, 2018, *168* (C), 1–20.

Bonhomme, Stéphane, Grégory Jolivet, and Edwin Leuven, "School Characteristics and Teacher Turnover: Assessing the Role of Preferences and Opportunities," *The Economic Journal*, 2016, *126* (594), 1342–1371.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff, "Explaining the Short Careers of High-Achieving Teachers in Schools with Low-Performing Students," *American Economic Review*, 2005, *95* (2), 166–171.

_ , _ , _ , **and** _ , "Analyzing the Determinants of the Matching of Public School Teachers to Jobs: Disentangling the Preferences of Teachers and Employers," *Journal of Labor Economics*, 2013, *31* (1), 83–117.

Brown, Christina and Tahir Andrabi, "Inducing positive sorting through performance pay: Experimental evidence from Pakistani schools," *University of California at Berkeley Working Paper*, 2020.

Buhl-Wiggers, Julie, Jason Kerwin, Jeffrey Smith, and Rebecca Thornton, "The impact of teacher effectiveness on student learning in Africa," in "Centre for the Study of African Economies Conference" 2017.

Cabrera, Jose Maria and Dinand Webbink, "Do Higher Salaries Yield Better Teachers and Better Student Outcomes?," *Journal of Human Resources*, 2020, *55* (4), 1194–1221.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik, "Robust Nonparametric Confidence Intervals for Regression?Discontinuity Designs," *Econometrica*, November 2014, *82*, 2295–2326.

_ , _ , **and** _ , "Optimal Data-Driven Regression Discontinuity Plots," *Journal of the American Statistical Association*, December 2015, *110* (512), 1753–1769.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma, "Simple Local Polynomial Density Estimators," *Journal of the American Statistical Association*, 2020, *115* (531), 1449–1455.

Chetty, Raj, John Friedman, and Jonah Rockoff, "Measuring the Impact of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 2014, *104(9)*, 2633–2679.

_ , _ , **and** _ , "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 2014, *104(9)*, 2593–2632.

__ , **John N Friedman, and Jonah E Rockoff**, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 2014a, *104* (9), 2593–2632.

__ , __ , **and** __ , "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 2014b, *104* (9), 2633–79.

Choo, Eugene and Aloysius Siow, "Who Marries Whom and Why," *Journal of Political Economy*, 2006, *114* (1), 175–201.

Clotfelter, Charles, Elizabeth Glennie, Helen Ladd, and Jacob Vigdor, "Would Higher Salaries Keep Teachers in High-Poverty Schools? Evidence from a Policy Intervention in North Carolina," *Journal of Public Economics*, 2008, *92* (5), 1352–1370.

de Ree, Joppe, Karthik Muralidharan, Menno Pradhan, and Halsey Rogers, "Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia," *The Quarterly Journal of Economics*, 2018, *133* (2), 993–1039.

Deserranno, Erika, "Financial Incentives as Signals: Experimental Evidence from the Recruitment of Village Promoters in Uganda," *American Economic Journal: Applied Economics*, 2019, *11* (1), 277–317.

__ , "Financial Incentives as Signals: Experimental Evidence from the Recruitment of Village Promoters in Uganda," *American Economic Journal: Applied Economics*, 2019, *11* (1), 277–317.

Diamond, Rebecca, "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, 2016, *106* (3), 479–524.

__ **and Cecile Gaubert**, "Spatial Sorting and Inequality," *Annual Review of Economics*, 2022, *14*, 795–819.

Diamond, William and Nikhil Agarwal, "Latent Indices in Assortative Matching Models," *Quantitative Economics*, 2017, *8* (3), 685–728.

Doval, Laura, "Dynamically Stable Matching," *Theoretical Economics*, 2022, *17* (2), 687–724.

Duflo, Esther, Pascaline Dupas, and Michael Kremer, "School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools," *Journal of Public Economics*, 2015, *123*, 92–110.

Ederer, Tim, "Two-Sided Matching Without Transfers: A Unifying Empirical Framework," *TSE Working Paper*, 2022.

Estrada, Ricardo, "Rules versus Discretion in Public Service: Teacher Hiring in Mexico," *Journal of Labor Economics*, 2019, *37* (2), 545–579.

Fack, Gabrielle, Julien Grenet, and Yinghua He, "Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions," *American Economic Review*, 2019, *109* (4), 1486–1529.

Falch, Torberg, "Teacher Mobility Responses to Wage Changes: Evidence from a Quasi-Natural Experiment," *American Economic Review*, 2011, *101* (3), 460–65.

_ **and Bjarne Strøm**, "Teacher Turnover and Non-Pecuniary Factors," *Economics of Education Review*, 2005, *24* (6), 611–631.

Finan, Frederico, Benjamin A. Olken, and Rohini Pande, "The Personnel Economics of the State," in Abhijit Banerjee and Esther Duflo, eds., *Handbook of Field Experiments, Volume II*, North Holland, December 2017.

Fox, Jeremy T, "Identification in Matching Games," *Quantitative Economics*, 2010, *1* (2), 203–254.

Galichon, Alfred and Bernard Salanié, "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models," *The Review of Economic Studies*, 2022, *89* (5), 2600–2629.

_ **and** _ , "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models," *Review of Economic Studies*, Forthcoming.

Gallegos, Sebastian, Christopher Neilson, and Franco Calle, "Screening and Recruiting Talent At Teacher Colleges Using Pre-College Academic Achievement," 2019.

Gerard, François, Miikka Rokkanen, and Christoph Rothe, "Bounds on treatment effects in regression discontinuity designs with a manipulated running variable," *Quantitative Economics*, 2020, *11* (3), 839–870.

Gilligan, Daniel O., Naureen Karachiwalla, Ibrahim Kasirye, Adrienne M. Lucas, and Derek Neal, "Educator Incentives and Educational Triage in Rural Primary Schools," *Journal of Human Resources*, 2022, *57* (1), 79–111.

Glewwe, P. and K. Muralidharan, "Chapter 10 - Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications," in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds., *Handbook of the Economics of Education*, Vol. 5, Elsevier, 2016, pp. 653–743.

Gualdani, Cristina and Shruti Sinha, "Partial Identification in Matching Models for the Marriage Market," *arXiv:1902.05610*, 2019.

Hanushek, Eric A, Steven G Rivkin, and Jeffrey C Schiman, "Dynamic Effects of Teacher Turnover on the Quality of Instruction," *Economics of Education Review*, 2016, *55*, 132–148.

Hatfield, John William and Paul R. Milgrom, "Matching with Contracts," *American Economic Review*, September 2005, *95* (4), 913–935.

He, YingHua, Shruti Sinha, and Xiaoting Sun, "Identification and Estimation in Many-to-one Two-sided Matching Without Transfers," *arXiv:2104.02009*, 2021.

Jackson, C. Kirabo, "What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes," *Journal of Political Economy*, 2018, *126* (5), 2072–2107.

Jackson, C Kirabo, Jonah E Rockoff, and Douglas O Staiger, "Teacher effects and teacher-related policies," *Annu. Rev. Econ.*, 2014, *6* (1), 801–825.

Judd, Kenneth L, *Numerical Methods in Economics*, MIT press, 1998.

Kapor, Adam J, Christopher A Neilson, and Seth D Zimmerman, "Heterogeneous Beliefs and School Choice Mechanisms," *American Economic Review*, 2020, *110* (5), 1274–1315.

Kelso, Alexander S. and Vincent P. Crawford, "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica*, 1982, *50* (6), 1483–1504.

Kennan, John and James R Walker, "The effect of expected income on individual migration decisions," *Econometrica*, 2011, *79* (1), 211–251.

Larroucau, Tomás and Ignacio Rios, "Dynamic College Admissions and the Determinants of Students' College Retention," 2020.

Leaver, Clare, Owen Ozier, Pieter Serneels, and Andrew Zeitlin, "Recruitment, Effort, and Retention Effects of Performance Contracts for Civil Servants: Experimental Evidence from Rwandan Primary Schools," *American Economic Review*, July 2021, *111* (7), 2213–46.

_ , _ , _ , and _ , "Recruitment, Effort, and Retention Effects of Performance Contracts for Civil Servants: Experimental Evidence from Rwandan Primary Schools," *American Economic Review*, 2021, *111* (7), 2213–46.

Magnac, Thierry and David Thesmar, "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 2002, *70* (2), 801–816.

Manning, Alan and Barbara Petrongolo, "How Local are Labor Markets? Evidence From a Spatial Job Search Model," *American Economic Review*, 2017, *107* (10), 2877–2907.

McFadden, Daniel et al., "Conditional Logit Analysis of Qualitative Choice Behavior," 1973.

Menzel, Konrad, "Large Matching Markets as Two-Sided Demand Systems," *Econometrica*, 2015, *83* (3), 897–941.

_ , "Bootstrap With Cluster-Dependence in Two or More Dimensions," *Econometrica*, 2021, *89* (5), 2143–2188.

Moretti, Enrico, "Real Wage Inequality," *American Economic Journal: Applied Economics*, 2013, *5* (1), 65–103.

Moscarini, Giuseppe and Fabien Postel-Vinay, "The Cyclical Job Ladder," *Annual Review of Economics*, 2018, *10*, 165–188.

Muralidharan, Karthik and Venkatesh Sundararaman, "Teacher Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, 2011, *119* (1), 39–77.

Pugatch, Todd and Elizabeth Schroeder, "Teacher pay and student performance: evidence from the Gambian hardship allowance," *Journal of Development Effectiveness*, April 2018, *10* (2), 249–276.

Resnick, Sidney I, *Extreme Values, Regular Variation, and Point Processes*, Vol. 4, New York: Springer, 1987.

Rivkin, Steven G, Eric A Hanushek, and John F Kain, "Teachers, Schools, and Academic Achievement," *Econometrica*, 2005, *73* (2), 417–458.

Rockoff, Jonah E, "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *American Economic Review*, 2004, *94* (2), 247–252.

Rosen, Sherwin, "The theory of equalizing differences," *Handbook of labor economics*, 1986, *1*, 641–692.

Roth, Alvin E, "Marketplaces, Markets, and Market Design," *American Economic Review*, 2018, *108* (7), 1609–58.

_ **and Marilda Sotomayor**, "Two-Sided Matching," *Handbook of Game Theory with Economic Applications*, 1992, *1*, 485–541.

Tincani, Michela M., "Teacher labor markets, school vouchers, and student cognitive achievement: Evidence from Chile," *Quantitative Economics*, 2021, *12* (1), 173–216.

Tincani, Michela M, "Teacher Labor Markets, School Vouchers, and Student Cognitive Achievement: Evidence from Chile," *Quantitative Economics*, 2021, *12* (1), 173–216.

Verdier, Valentin and Carson Reeling, "Welfare Effects of Dynamic Matching: An Empirical Analysis," *The Review of Economic Studies*, 2022, *89* (2), 1008–1037.

Waldinger, Daniel, "Targeting in-kind Transfers Through Market Design: A Revealed Preference Analysis of Public Housing Allocation," *American Economic Review*, 2021, *111* (8), 2660–96.

World Bank, "World Development Report 2018 : Learning to Realize Education's Promise," Technical Report, Washington, DC: World Bank 2018.