THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITE DE TOULOUSE

Délivré par l'Université Toulouse Capitole

École doctorale : **Sciences Economiques-Toulouse School of Economics**

Présentée et soutenue par

KUN LI

le 28 September 2016

Essays on Productivity and Distortions

Discipline : **Sciences Economiques**

Unité de recherche : **TSE-R (UMR CNRS 5314 – INRA 1415)**

Directeur de thèse :

Christian Hellwig, Professor of Economics, Toulouse School of Economics

JURY

Rapporteurs Pierre Dubois, Professor of Economics, Toulouse School of Econmics
Franck Portier, Professor of Economics, Toulouse School of Economics

Suffragants Thomas Chaney, Professor of Economics, Science Po
Marti Mestieri, Assistant Professor of Economics, Northwestern University

TOULOUSE SCHOOL OF ECONOMICS

Essays on Productivity and Distortions

by

Kun Li

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the
Toulouse School of Economics

August 2016

Acknowledgements

This dissertation could not have been written without Christian Hellwig who not only served as my advisor but also encouraged and challenged me throughout my academic program. I would also like to express my gratitude to other dissertation committee members including Thomas Chaney, Pierre Dubois, Marti Mestieri, and Franck Portier. I thank Jo Van Biesebroeck and Yifan Zhang for sharing the price data and code with me, and answering many questions about the data. For thoughtful comments and insight, I thank Bruno Biais, Patrick Feve, Augustin Landier, Thierry Magnac, Paul Scott, Robert Ulbricht, Jo Van Biesebroeck, Venky Venkateswaran, and various seminar and conference participants. I appreciate the opportunities to work with Michael Peters and Steven Pai Xu on the second chapter, and Bruno Biais and Augustin Landier on the third chapter. All remaining errors are my own.

Contents

List of Figures

List of Tables

To Yangbo, Jingxiu and Lin whose love, support and devotion made all my achievements possible

Chapter 1

Privatization, Distortions, and Productivity

1.1 Introduction

How does different types of ownership affect firm performance, and in turn influence aggregate economic outcomes? In general, firms may suffer from two types of inefficiencies. First, firms may have low *productivity*, which measures the level of output given the level of inputs. Second, firm may also have high *distortions*. Here, the level of distortions measures the extend to which firms' input choices deviate from the profit maximization benchmark.

As a consequence, change of ownership–in particular privatization–may mitigate these inefficiencies, and thus improve the aggregate economic outcomes through increasing productivity and reducing distortions. For instance, privatization may increase productivity by improving firms' management practice. It may also reduce distortions since privatized firms may be less influenced by the state policy. Instead, their objectives may be more aligned with profit maximization.

This paper seeks to understand *to what extent* and *through what channels* privatization has contributed to the rapid growth of GDP and TFP in China's manufacturing sector during this transition. To begin with, I build a framework to separately identify productivity, distortions, and production technology from firm level input-output data. I first build a discrete time dynamic investment model where a firm chooses investment, labor and materials to generate revenue. The model features two kinds of shocks every period: a productivity shock and a distortion shock. Following the misallocation literature[1], the distortion is modelled as a wedge that moves firm's marginal revenues away from its marginal costs for all inputs to the same extent. To account for the direct effect of privatization, the model allows ownership and privatization to affect the evolution of productivity.

Second, I present a novel method to jointly identify distortions, productivity and the parameters in the revenue function. I then estimate these parameters of interest through structural estimation. Following the literature on production function estimation[2], the identification relies on the optimal material choice implied from the model, the timing of play and the evolution of productivity. To distinguish distortion from firm-level productivity uncertainties, I assume that the firm's dynamic input decisions depend on distortion and expected productivity as in Asker et al. [2014], whereas its static input decision depends on both distortion and realized productivity as in Hsieh and Klenow [2009].

Next, I apply this framework to understand the effects of privatization on Chinese manufacturing firms. In particular, I use the dataset of Chinese National Bureau of Statistics (NBS) industrial annual surveys from 1998 to 2007. I first estimate the parameters in productivity and distortion processes, and then I use the estimated model to evaluate the impact of privatization on productivity and distortions at firm and aggregate level. I find that when a firm is privatized, it immediately realizes a 5% increase in productivity on average, which is equivalent to one year of productivity growth for an average manufacturing firm in the sample. Coupled with the estimated persistent productivity process, the long run gains from privatization appear to be even more important: It increases the privatized firms' productivity by 12% at 2007. Meanwhile, privatization immediately reduces distortions by 2%, which results in a 0.2% increase of revenue on average. In particular, firms with positive implicit tax before privatization benefit the most from the reduction of distortions. However, the gains from improved factor reallocation due to privatization appear to have limited effects on firm revenues in the long run.

Now turn to the aggregate effects of privatization on allocative efficiency and aggregate productivity growth (henceforth APG). Based on Melitz and Polanec [2015] and Collard-Wexler and Loecker [2015], I propose a decomposition method for APG to explicitly account for the direct effect of privatization. The decomposition suggests that

[1] In particular, Restuccia and Rogerson [2008], Hsieh and Klenow [2009], and more recently, David et al. [2015].

[2] In particular, the model and the identification augment the empirical production function estimation literature as in Olley and Pakes [1996], Levinsohn and Petrin [2003], Ackerberg et al. [2006], and Gandhi et al. [2012] with micro-level distortions as modelled in Restuccia and Rogerson [2008] and Hsieh and Klenow [2009].

with an annual 5% growth rate within the manufacturing sector, the direct contribution of privatization to APG is about 5% in the sample. Given that only a limited number of firms are privatized, there is a fairly sizeable gain associated with the policy. Privatizations contribute to the aggregate productivity growth mainly through their direct impact on firm level productivity. In contrast, although the privatizations also reduce revenue distortions, the associated reallocation gains only contribute a small fraction to overall TFP growth.

The main contributions of this paper are threefold. First, the paper bridges the *direct* and *indirect* approaches in the misallocation literature and identifies a particular policy's effects on both productivity and distortions[3]. It links an important source of misallocation, ownership and privatization, to the measured distortions and productivity from the *indirect* approach. Moreover, this paper explicitly emphasizes the effect of privatization on both productivity and distortions. By contrast, Brown et al. [2006] and Braguinsky et al. [2015], and De Loecker [2011] focus on policy's productivity effect, while Collard-Wexler and Loecker [2015] and Khandelwal et al. [2013] emphasize the reallocation effects.

Second, the paper extends the production function estimation literature by establishing a method that allows for both unobserved distortions and productivity. Gandhi et al. [2012] shows that *without distortion*, under Cobb-Douglas production function, the revenue elasticity of material is identified as average material revenue shares from firm's static first order condition. Conditional on this revenue elasticity, the rest of revenue function parameters are identified following procedures in Ackerberg et al. [2006]. But with *unobserved distortion*, revenue share of material is a composition of revenue elasticity of material and distortion, therefore the average revenue share is no longer an unbiased estimator of static input elasticity. By contrast, this paper illustrates how to estimate production function with endogeneity problems from both unobserved productivity and distortion, as well as the transmission bias from allowing both static and dynamic inputs. It shows that lagged material, lagged material revenue shares, and current revenue jointly identify the revenue elasticity of material and persistence of productivity. With the estimated technology parameters, it is straightforward to back out the implied distortion and productivity at the firm-year level.

Third, the paper provides a framework to analyze the dispersions of marginal product of inputs with dynamically chosen inputs and distortions. In doing so, it reconciles two distinct views from the literature that whether researchers could measure distortion from

[3] The definitions of *direct* and *indirect* approaches follow Restuccia and Rogerson [2013]. The *direct* approach is to pick and measure one factor that is important for misallocation, and quantify the extent to which this factor generates misallocation and impact aggregate TFP. Instead, the *indirect* approach tries to focus on the net effect of an entire bundle of underlying factors on misallocation without attributing distortions to any specific underlying factor.

marginal product of inputs[4]. This paper combines both distortions and dynamic inputs in a dynamic investment model, and estimates productivity, distortions, and production function parameters. The methodological innovation in this paper corrects the bias from ignoring unobserved distortions. Using the estimated distortions and productivity from the Chinese firm-level data, this paper finds that both channels are important to the dispersion of marginal revenue product of capital (MRPK). More importantly, distortions can be estimated from firm's static input choice, which is a composition of distortions, productivity, and estimated technology parameters.

This paper links to the growing literature on document frictions/distortions in China. Hsieh and Klenow [2009] first studies the distortions in China and India, and proposes using the dispersion of MRPK to measure distortions in the economy. Brandt et al. [2013] further decompose the distortions across time and space. Song et al. [2011] emphasizes the differential credit access between state-owned and private firms in a growth model to explain China's economic transition. Khandelwal et al. [2013] examines the effect of removing the trade quotas when China enters WTO. Cooper et al. [2015] focus on the different objectives between state and private enterprises. Cooper et al. [2013] studies the consequences of new Chinese labor regulations intended to protect workers' employment conditions. Wang [2011] examines the price effects of privatization of housing assets in China. Holmes et al. [2015] focuses on the *quid pro quo* policy in China and quantifies its impact on global innovation and welfare. Kalouptsidi [2014] proposes an empirical framework to detect subsidy in the ship-building industry in China. This paper quantifies both productivity and distortions effects of privatization for Chinese manufacturing firms.

1.2 A Dynamic Investment Model

Consider a single agent dynamic investment model. Time is discrete and infinite. A firms uses 3 inputs–capital K , labor L and material M–to generate revenue. Moreover, there are two types of idiosyncratic shocks that influence the firm's revenue. The first is a productivity shock, which directly influences revenue. Write ω for the log-term of the productivity shock.

The second is the output distortion, which influences revenue through input choices. Write τ for the log-term of the output distortion. In particular, the output distortion

[4] Hsieh and Klenow [2009] uses the dispersion of marginal revenue product of capital (MRPK) to measure distortion with the underlying assumption that firms are static optimizers, which implies the input shares are unbiased estimators of production function parameters. However, Asker et al. [2014] shows that with capital adjustment costs and micro uncertainties, a distortion free economy could generate similar patterns of such dispersion.

increases the marginal product of all inputs by the same proportion but is not part of the revenue, as in Restuccia and Rogerson [2008] and Hsieh and Klenow [2009]. Thus, the distortion can be seen as an unobserved firm-year level implicit tax or subsidy on the output.

Firm i's revenue at period t is of Cobb-Douglas form

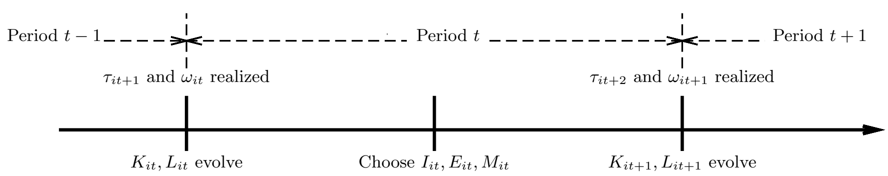
$$R_{it} = K_{it}^{\theta_k} L_{it}^{\theta_l} M_{it}^{\theta_m} \exp\left(\omega_{it}\right). \tag{1.1}$$

Here θ_k, θ_l and θ_m are capital, labor and material coefficients respectively. Moreover, they measure the elasticities of capital, labor and material to revenue. For the rest of the paper, I also refer them as technology parameters.

The timing is shown in figure 1.1:

1. At the beginning of each period t, firm i carries capital K_{it}, labor L_{it}, period t productivity ω_{it}, period t distortions τ_{it}, and period $t+1$ distortions τ_{it+1}. Write w_t for the wage rate, r_t for the interest rate, and p_t^M for the material price at period t.
2. Next, taking a stream of input prices $\{w_t, r_t, p_t^M\}_{t=1}^{t=\infty}$ as given, firm i chooses material M_{it} to generate period t revenue, and makes investment I_{it} and hiring/firing decisions E_{it}. By doing so, the firm sets the period $t+1$ usage of capital K_{it+1} and labor L_{it+1}.
3. After the firm made all its input choices, revenue R_{it} is recorded. Capital, labor, productivity and distortions evolve to their period $t+1$ level. Then firm i begins its period $t+1$ operation.

FIGURE 1.1: The Time Line



Under the timing assumptions, the firm makes decision on current usage of material every period, and the decision depends on both distortions and realized productivity.[5] Thus, the material is a static input. Moreover, in every period, the firm makes decisions on the next period usage of capital and labor, and the decisions depend on distortions and

[5]This is similar to the setting in Hsieh and Klenow [2009].

expected productivity.[6] Thus, both capital and labor are dynamic inputs. In particular, capital and labor evolve as follows:

$$K_{it+1} = (1 - \delta^K)K_{it} + I_{it},$$
$$L_{it+1} = (1 - \delta^L)L_{it} + E_{it},$$

where δ^K is the depreciation rate of capital, and δ^L is the depreciation rate of labor.

The cost of production $C(\omega_{it}, K_{it}, L_{it}, I_{it}, E_{it})$ is a function of the dynamic inputs– capital and labor–and the potential adjustment costs.[7]

We now discuss the evolution of the idiosyncratic shocks. Write $\mathcal{I}_{it}$ for firm i's information set at the beginning of period t. The set consists of the history of its level of inputs, output and productivity until period t, and the level of distortions until period $t+1$. Assume that productivity $\omega_{it} \in \mathcal{I}_{it}$ is a persistent Hicks neutral productivity shock, and follows a Markov process:

$$\omega_{it+1} = \Phi(\omega_{it}, X_{it}) + \xi_{it+1}, \tag{1.2}$$

where $\Phi(\omega_{it}, X_{it})$ is the predictable component of the productivity, $X_{it} \in \mathcal{I}_{it}$ is a vector of observables that directly enter the productivity process 1.2, and ξ_{it+1} is innovation of the productivity process. More specifically, $\xi_{it+1} = \omega_{it+1} - \mathbb{E}[\omega_{it+1}|\mathcal{I}_{it}]$. By construction, $\xi_{it+1} \perp \mathcal{I}_{it}$, i.e., the innovation of productivity is orthogonal to any elements in the information set $\mathcal{I}_{it}$.

Given $\{\tau_{it}, \tau_{it+1}\} \in \mathcal{I}_{it}$, the distortions process is characterized by

$$\tau_{it+2} = \Upsilon(\tau_{it}, \tau_{it+1}, X_{it}) + \zeta_{it+2},$$

where $\Upsilon(\tau_{it}, \tau_{it+1}, X_{it})$ is the predictable component, and ζ_{it+2} is the innovation of distortions process. The innovation component is orthogonal to any elements in the information set $\mathcal{I}_{it}$. Moreover, I assume that the innovations of both productivity and distortions are i.i.d. across firms.

The firm's objective is choosing the usage of inputs to maximize the discounted adjusted (by distortions) profit streams. Formally, the recursive form of firm i's problem is as

[6] This is similar to the setting in Asker et al. [2014].

[7] Since the identification does not require first order condition of dynamic inputs, I do not specify the functional form of adjustment cost at this moment. I only assume that the static input M_{it} is not part of the adjustment cost.

follows

$$\begin{aligned} V(K_{it}, L_{it}, \omega_{it}, \tau_{it}, \tau_{it+1}, t) = \max_{I_{it}, E_{it}, M_{it}} & \exp(\tau_{it} + \omega_{it}) K_{it}^{\theta_k} L_{it}^{\theta_l} M_{it}^{\theta_m} \\ & - w_t L_{it} - p_t^M M_{it} - C(\omega_{it}, K_{it}, L_{it}, I_{it}, E_{it}) \\ & + \beta \mathbb{E}_{\omega,\tau} V(K_{it+1}, L_{it+1}, \omega_{it+1}, \tau_{it+1}, \tau_{it+2}, t+1) \end{aligned}$$

subject to

$$\begin{aligned} K_{it+1} &= (1 - \delta^K) K_{it} + I_{it}, \\ L_{it+1} &= (1 - \delta^L) L_{it} + E_{it}, \\ \omega_{it} &= \Phi(\omega_{it-1}, X_{it-1}) + \eta_{it}, \text{and} \\ \tau_{it+2} &= \Upsilon(\tau_{it}, \tau_{it+1}, X_{it}) + \zeta_{it}. \end{aligned}$$

Here $\mathbb{E}_{\omega,\tau}$ is the expectation for transition of productivity and distortions, β is the discount rate for all firms.

Equilibrium Input Choices We now characterize firm's optimal input choices. Given a stream of input prices $\{w_t, p_t^M\}_{t=0}^{\infty}$ and revenue function parameters $\{\theta_k, \theta_l, \theta_m\}$, firm i's optimal choice of material is a function of $\{K_{it}, L_{it}, \omega_{it}, \tau_{it}\}$ that satisfies the static first order condition

$$p_t^M = \exp(\tau_{it} + \omega_{it}) \theta_m K_{it}^{\theta_k} L_{it}^{\theta_l} M_{it}^{\theta_m - 1}. \tag{1.3}$$

The firm determines its next period capital and labor by a system of implicitly defined first order conditions with respect to capital and labor.

From this first order condition, we see the role of distortions τ_{it} in resource allocation. Distortions create wedges between the marginal cost of material (i.e., the price p_t^M) and the marginal revenue of material. When $\tau_{it} < 0$, the marginal cost is less than the marginal revenue, so the firm faces implicit tax. When $\tau_{it} > 0$, the marginal cost is greater than the marginal revenue, so the firm faces implicit subsidy. Only when $\tau_{it} = 0$, the marginal cost equals to the marginal revenue.

1.3 Identification

The goal of the paper is to empirically evaluate the effects of privatization on output and productivity. To do so, this section provides a framework to separately identify productivity, distortions and production technology from a standard production dataset, i.e., firm level input and output data.

More specifically, I develop an identification strategy that builds on Gandhi et al. [2012] to address three empirical concerns. The first concern is the endogeneity issue that all inputs are correlated with unobserved productivity and distortions. (See Olley and Pakes [1996], Levinsohn and Petrin [2003] and Ackerberg et al. [2006].) The second is the transmission-bias issue that the choice of the static input depends on dynamic inputs. (See Marschak and Andrews [1944] and Gandhi et al. [2012].) Third, an additional identification issue arises with distortions in the firm's problem since these distortions are unobserved by the econometrician.

In this section, I first illustrate the identification using a single static production factor. Next, I discuss the general identification with multiple inputs. Finally, I simulate a panel dataset to compare different estimation methods.

1.3.1 Illustration of Identification: Single Static Input

Consider an input-out data that contains an output variable—log revenue r_{it}, and a static input variable—log material m_{it} for firms $i \in \{1, \cdots, N\}$ and year $t \in \{1, \cdots, T\}$.

Since the revenue function is Cobb-Douglas, the log-revenue r_{it} is linear in log-material m_{it} and log-productivity ω_{it}.

$$r_{it} = \theta_m m_{it} + \omega_{it}. \tag{1.4}$$

Following the empirical production function estimation literature as in Olley and Pakes [1996], Levinsohn and Petrin [2003], Ackerberg et al. [2006], and Gandhi et al. [2012], I assume the productivity follows an autoregressive (AR(1)) process

$$\omega_{it+1} = \mu + \gamma \omega_{it} + \alpha \tau_{it+1} + \eta_{it+1}. \tag{1.5}$$

On the right hand of Equation (1.5), the first term μ is the intercept of the AR(1) process. The second term measures how the current productivity contributes to the future productivity, where γ measures the persistence level of the productivity shock. The third term is the future level of distortions τ_{it+1} which serves as a control variable in the process. The last term is the component of productivity innovation η_{it+1} that is orthogonal to the level of distortions τ_{it+1}.[8]

[8]Another way to interpret this formulation is to assume that firms are Bayesian: They are aware of the fact that distortions and productivity are correlated, and use distortions as a signal to forecast productivity.

According to the timing, firms choose material m_{it} after knowing both productivity ω_{it} and distortions τ_{it}.[9] Taking input price p_t^M as given, firm i solve the first order condition of material:

$$\ln p_t^M = \ln \theta_m + (\theta_m - 1)m_{it} + \omega_{it} + \tau_{it} \tag{1.6}$$

Combing Equation (1.6) and Equation (1.4), we have the share equation:

$$\ln s_{it} \equiv \ln \frac{P_t^M M_{it}}{R_{it}} = \ln \theta_m + \tau_{it}. \tag{1.7}$$

The share equation (1.7) illustrates the difference in identification of the elasticity with and without distortions. Without distortions, one can use average material revenue share to pin down the elasticity θ_m. With distortions, it is not quite right to do so: it imposes an additional assumption that the average distortions is zero. Therefore, use average material share to pin down the elasticity θ_m may be problematic in this case. However, by manipulating the revenue function with first order condition for material and productivity evolution equation, one can identify the relevant technology parameters with distortions.

Proposition 1.1 presents the identification results for this special case.

Proposition 1.1. *Suppose the following assumptions are satisfied,*

1. *Productivity ω_{it} is persistent, i.e., $\gamma \neq 0$.*
2. *Distortions τ_{it} varies across firms and years.*

The technology parameter θ_m, productivity ω_{it} and distortions τ_{it} can be identified from Equation (1.4), Equation (1.5), Equation (1.6) and Equation (1.7) from the input-output data.

[9]This timing assumption is consistent with both the production function estimation literature and static misallocation literature (e.g, Hsieh and Klenow [2009]), and is embedded in the dynamic investment model in section 1.2.

Proof. Using Equation (1.5), Equation (1.6) and Equation (1.7), I rearrange the terms in Equation (1.4):

$$\begin{aligned}
r_{it} &= \theta_m m_{it} + \omega_{it} \\
&= \frac{\theta_m}{1-\theta_m}(\omega_{it} + \tau_{it}) + \omega_{it} + \text{Constant} \\
&= \frac{1}{1-\theta_m}\omega_{it} + \frac{\theta_m}{1-\theta_m}\tau_{it} + \text{Constant} \\
&= \frac{\alpha+\theta_m}{1-\theta_m}\tau_{it} + \frac{\gamma}{1-\theta_m}\omega_{it-1} + \frac{1}{1-\theta_m}\eta_{it} + \text{Constant} \\
&= \frac{\alpha+\theta_m}{1-\theta_m}\tau_{it} - \frac{\gamma}{1-\theta_m}\tau_{it-1} + \gamma m_{it-1} + \frac{1}{1-\theta_m}\eta_{it} + \text{Constant} \\
&= \frac{\alpha+\theta_m}{1-\theta_m}\ln s_{it} - \frac{\gamma}{1-\theta_m}\ln s_{it-1} + \gamma m_{it-1} + \frac{1}{1-\theta_m}\eta_{it} + \text{Constant}. \qquad (1.8)
\end{aligned}$$

The first line is the original definition of revenue function Equation (1.4), and the second line uses the material first order condition Equation (1.6) to replace material as a function of productivity and distortions. The third line replaces current productivity by the current distortions and lagged productivity from the AR(1) process described in Equation (1.5). The fourth line uses Equation (1.6) again to replace productivity as a function of material and distortions. The last line uses the share equation Equation (1.7) to replace the unobserved distortions with the observed shares. The final reduced form representation features all observed variables on both side of the equation. Under the assumptions in proposition 1.1, the three regressors at the RHS identify the three unknowns. □

The identification involves two different assumptions. First, the persistence of productivity plays a key role to link lagged inputs with current output. If the productivity is not persistent, i.e., $\gamma = 0$, the reduced form representation of the output is

$$r_{it} = \frac{\alpha+\theta_m}{1-\theta_m}\ln s_{it} + \frac{1}{1-\theta_m}\eta_{it} + \text{Constant}.$$

As a result, we cannot separately identify α and θ_m in this case.

The second assumption requires distortions τ_{it} to vary across firms and years. If τ_{it} is a constant, there are no variation in the reduced form representation to provide information on the coefficients of $\ln s_{it}$ and $\ln s_{it-1}$. Consider a specific case as in Gandhi et al. [2012], where there are no distortions, i.e., $\tau_{it} = 0$. The reduced form representation becomes

$$r_{it} = \gamma m_{it-1} + \frac{1}{1-\theta_m}\eta_{it} + \text{Constant}.$$

Regressing the output r_{it} on input m_{it} provides no information on the technology parameter θ_m. To see this, consider the identification in Gandhi et al. [2012], which uses the share equation to derive the coefficient of material

$$\theta_m = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} s_{it}}{NT}.$$

When τ_{it} is an unobserved nonzero constant, there is no general way to identify all the parameters.

In fact, the identification of this reduced form representation is consistent with empirical production function estimation literature: Productivity has a predictable component and an unpredictable component. By construction the unpredictable component is orthogonal to any element within firm's information set $\mathcal{I}_{it}$. In the current setting, η_{it} is not the entire productivity innovation but the part that is uncorrelated with current distortions.[10] However, it still bears the orthogonal property between productivity innovation η_{it} and elements in firm's information set $\mathcal{I}_{it-1}$. To simplify the language, I refer η_{it} as the productivity innovation for the rest of the paper.

To summarize, introducing unobserved distortions does not further complicate the identification: it even provides more variations that the econometrician can exploit to identify the system. This single-static-input case illustrates that one can identify technology parameters, productivity and distortions by only using the static choices.

1.3.2 General Identification with Multiple Inputs

Instead of restricting to the single-static-input case, now I explore the identification strategy by incorporating dynamic inputs, capital and labor, into the revenue function. In addition, I allow exogenous covariates in the productivity process. Comparing to the previous single input and output data, now there are two more input variables–log capital k_{it} and log labor l_{it}–and a vector of exogenous covariates x_{it} for firm i at year t.

The revenue function is as follows

$$r_{it} = \theta_l l_{it} + \theta_k k_{it} + \theta_m m_{it} + \omega_{it}, \tag{1.9}$$

where θ_l is the elasticity of labor to revenue, and θ_k is the elasticity of capital to revenue. The productivity process is specified as an AR(1) process

$$\omega_{it} = \mu + \gamma\omega_{it-1} + \alpha\tau_{it} + \pi x_{it} + \eta_{it}. \tag{1.10}$$

It is consistent with our general specification in Equation (1.5).

[10] This setup is similar to the one in David et al. [2015].

Now the first order condition of material m_{it} is

$$\ln p_t^M = \ln \theta_m + (\theta_m - 1)m_{it} + \theta_k k_{it} + \theta_l l_{it} + \omega_{it} + \tau_{it}. \tag{1.11}$$

Combing it with the revenue equation, Equation (1.9), we have the same share equation

$$\ln s_{it} \equiv \ln \frac{P_t^M M_{it}}{R_{it}} = \ln \theta_m + \tau_{it}. \tag{1.12}$$

Proposition 1.2 provides general identification results for a model with dynamic inputs and exogenous covariates in productivity process.

Proposition 1.2. *Consider the same assumptions in Proposition 1.1. The technology parameter $\{\theta_k, \theta_l, \theta_m\}$, parameters of covariates in the productivity process $\{\alpha, \pi\}$, the persistence of the productivity process γ, productivity ω_{it} and distortions τ_{it} can be identified from Equation (1.9), Equation (1.10), Equation (1.11) and Equation (1.12) using the input-output data.*

Proof. In this case, the reduced form representation of revenue is given by

$$\begin{aligned}\mathbb{E}[r_{it}] = {}& \gamma m_{it-1} + \frac{\alpha + \theta_m}{1-\theta_m} \ln s_{it} - \frac{\gamma}{1-\theta_m} \ln s_{it-1} \\ & + \frac{\theta_k}{1-\theta_m} k_{it} + \frac{\theta_l}{1-\theta_m} l_{it} - \frac{\gamma\theta_k}{1-\theta_m} k_{it-1} - \frac{\gamma\theta_l}{1-\theta_m} l_{it-1} + \frac{\pi}{1-\theta_m} x_{it-1} + \text{Constant}.\end{aligned} \tag{1.13}$$

The derivation is available in A.1.1. On the right hand side of Equation (1.13), the first line consists of static inputs and input shares, which is similar to Equation (1.8). Thus the identification of θ_m, γ and α are similar to the previous case in Section 1.3.1. In particular, we can identify γ from the variations of m_{it-1}. We can also identify θ_m and α from variations of $\ln s_{it-1}$ and $\ln s_{it}$, which are proxy variables for distortions with differences in the means. The second line of Equation (1.13) consists of dynamic inputs and exogenous covariates. Given θ_m, we can directly identify π, θ_k, θ_l from variations of x_{it-1}, k_{it} and k_{it-1}, l_{it} and l_{it-1} respectively. In fact, introducing dynamic inputs brings additional identification power for γ: Dividing the coefficients of k_{it-1} and l_{it-1} with the coefficients of k_{it} and l_{it} gives us $-\gamma$. Once we have all the technology parameters, we can back out firm-year level productivity ω_{it} and distortions τ_{it} from Equation (1.9) and Equation (1.12). □

It is worth discussing the empirical consequences of ignoring the endogeneity and distortions. First, suppose one is interested in the effect of x_{it-1} on output by computing the correlation between x_{it-1} and r_{it}. This is equivalent to directly estimate an ordinary least square (OLS) model. In this case, $\hat{\pi} = \frac{\pi}{1-\theta_m}$ where $\hat{\pi}$ is the estimated effect. Since $0 < \theta_m < 1$, there will be an upward bias as the OLS model assumes $\theta_m = 0$, and ignoring the static input choices. Second, if one estimates the technology parameter θ_m ignoring distortions, the estimated effect of x_{it-1} would also be biased. For example, suppose on average there are positive distortions. Using average material revenue share would overestimate θ_m and thus underestimate the π.

1.3.3 Simulation

To compare different estimation methods, I first simulate a panel data set with inputs, outputs, productivity and distortions. Next, I estimate the production function parameters using different estimation methods, and then compare their biases and asymptotic properties.

Simulating the Dataset I simulate a firm-year level input and output dataset which has $N = 1000$ firms across $T = 11$ years. I start with simulating the productivity and distortion process. I set the level of productivity to be a AR(1) process with intercept $\mu = 0.1$, persistence $\gamma = 0.7$ and correlation between distortion and productivity $\alpha = -0.4$. I set the level of distortions to be a AR(1) process with intercept $\psi = 0.1$, persistence parameter $\phi = 0.3$. I also specify the residues of these two stochastic process–η_{it} and ζ_{it}–to be drawn from two independent standard normal distributions. Moreover, I draw firms' initial level of both productivity and distortions from a normal distribution with mean 0.2 and variance 0.5. Then I simulate the whole time series of productivity and distortions.

Next, I simulate the series of inputs and revenue. I set the value of the technology parameters to be $\theta_k = 0.1, \theta_l = 0.3, \theta_m = 0.5$ respectively. I randomly draw the dynamic inputs k_{it} and l_{it} from a uniform distribution on $[5, 15]$. Assume material price p_t^m to be 1 through all the sample period. Then I use equation (1.3) to compute the optimal material inputs m_{it} , and use the revenue function (1.9) to generate revenue series r_{it}. To this end, the dataset contains $\{k_{it}, l_{it}, m_{it}, r_{it}, \eta_{it}, \omega_{it}, \tau_{it}\}_{i=1,t=1}^{i=N,t=T}$. The common observed dataset for an econometrician is $\{k_{it}, l_{it}, m_{it}, r_{it}\}_{i=1,t=1}^{i=N,t=T}$.

Estimating Production Function Using the simulated dataset, I go on to estimate the parameters $\{\hat{\theta}_k, \hat{\theta}_l, \hat{\theta}_m, \hat{\gamma}\}$, and implied distribution of productivity $\hat{\omega}_{it}$ and distortions $\hat{\tau}_{it}$ using three different methods. The first is the ordinary least square (OLS), which directly regresses output on inputs. The second is the GNR method proposed by Gandhi et al. [2012] that assumes no distortions in the economy, i.e., $\tau_{it} = 0$ for all i and

t. The third is the method proposed in section 1.3. For the last two methods I implement both Nonlinear least square (NLS) and generalized method of moments (GMM) approaches respectively. I also compute a non-structural measure of productivity–labor productivity–which is defined as $\omega_{it}^{Labor} = r_{it} - l_{it}$. I compare it with the productivity measures implied from the structural methods. Since the estimation requires lagged value of inputs and outputs, I only use observations that are not the first year of the sample. To this end, the sample size reduces to 10000 firm-year observations. I repeat the simulation for 1000 times and then plot the asymptotic distributions of the estimates.

Figure A.1 depicts the results of simulation exercise for technology parameters, namely $\{\hat{\theta}_k, \hat{\theta}_l, \hat{\theta}_m, \hat{\gamma}\}$ with the true parameters at $\{0.1, 0.3, 0.5, 0.7\}$. The red line in each graph represents the true underlying values for the parameters. The simulation results suggest that with unobserved distortions, OLS attributes too much weight to the material inputs, producing an upward bias. The average estimates for θ_k and θ_l from OLS are 0.01 and 0.03, while the true parameter values are at 0.1 and 0.3, respectively. It also attributes too little to the capital and labor inputs, producing a downward bias. The average estimates from OLS for θ_m is 0.93, and the true value is at 0.5. With unobserved distortions, the GNR method performs worse than the OLS method: the bias for the technology parameters are even larger. Finally the method proposed in section 1.3 produces consistent estimates. Two different identification strategies, the nonlinear least square (NLS) and generalized method of moments (GMM) yield quite similar asymptotic properties.

Figure A.1 is here.

I then turn to the implied distribution of productivity $\hat{\omega}_{it}$ and distortions $\hat{\tau}_{it}$. I use the mean of coefficients from the simulation exercises to compute the implied productivity and distortion distributions from the simulated dataset. Figure A.2 depicts the distribution of implied productivity and distortions. Figure A.3 depicts the correlation between implied productivity from all these methods, labor productivity, and the true productivity. The results are two-folds. First, both the OLS and GNR methods underestimate the mean and the dispersion of productivity. Second, implied productivity from the OLS method has little correlation with the true productivity. The GNR method performs better comparing to OLS method: The implied productivity and distortions distributions are consistent under both NLS and GMM approaches using the method proposed in section 1.3.

Figure A.2 is here.

Figure A.3 is here.

Finally, I look at the implied correlation between productivity and distortions. Figure A.4 depicts the correlation between implied productivity from all three methods and the true distortion. The method proposed in section 1.3 captures the right correlation between productivity and distortions, while the OLS and GNR methods produce much more negative correlation, and labor productivity generates close to zero correlation.

Figure A.4 is here.

1.4 Privatization: Data and Specification

This section applies the framework provided in Section 1.3 to the Chinese manufacturing firm level data from 1998 to 2007. In so doing, we can quantify the effects of privatization in China. I introduce the input-output data used in this paper in Section 1.4.1. Then in Section 1.4.2, I discuss the specification used in quantifying the effects of privatization.

1.4.1 Data

Production Data The data is mainly from the National Bureau of Statistics (NBS) firm level survey. It contains annual firm-level data from 1998 to 2007 across various industries. These firms are either state-owned, or are non-state firms with sales above 5 million RMB (hereafter referred to as the "above-scale"industrial firms). They are from industries of mining, manufacturing or electricity/gas/water producing.

The raw data come as multiple cross-section files for each year. To make it a standard input-output production dataset, I first follow Brandt et al. [2012] to link observations across years and compute capital stocks and produce a panel data with inputs, outputs and other information. [11]

Then I further clean the dataset for estimation. In particular, I impose four restrictions to the data sample used in the estimation: a) I require that firms have a minimum of two consecutive years of data. b) I require that for any pair in the data, firm i should operate in the same two-digit industry. c) I require that there are no missing or negative values for current and lagged log revenue and inputs, i.e., capital, total wage bill, number of workers, material and revenue. d) I require that there are no missing values for the

[11] I thank Jo Van Biesebroeck and Yifan Zhang for sharing the price data and code with me, and answering many questions about the data.

following variables for both year t and year $t-1$: exporting value, firm age, provinces, registration type, and industrial classification.

Measurement I now discuss the measures of key variables constructed from above dataset. The first is the ownership dummy, $\mathbb{1}_{i \in SOE,t}$. To construct this variable, I rely on firm's registration type. There are four general categories: State-owned, Collective, Private, and Foreign (Including Hong Kong, Macau and Taiwan). I record a firm i at year t with registration code 110 or 151 as a state-owned firm regardless of its capital composition.[12] Based on the ownership dummy, the privatization dummy $\mathbb{1}_{privatized,i,t-1}$ equals to 1 if firm i is a state-owned firm at year $t-1$ but a non-state-owned firm at year t.[13]

The second is the inverse mills ratio, which is used to control for the potential selection bias. During the sample period, a substantial number of firms exit from the market, and the firms may exit due to selection effects. Thus, the data may only contain the relatively productive firms that survived from the privatization, and thus induce an upward bias for the effects of privatization on productivity. To address the potential selection bias, I follow Olley and Pakes [1996] to estimate a probit model of exit on firm size, firm age, sales capital ratio, ownership and their interactions as well as year and provincial dummies. Then I use the estimated model to construct the inverse mills ratio as a control variable in the productivity evolution equation.

Finally, I construct two set of geographic dummies based on firms' registered zip code. One at province level and the other at region level. I first group the observations into provinces and then into four economic regions according to The Communist Party of China [2006]. This is because of the large differences between regional and even provincial policy. Figure A.5 depicts the four economic regions.

Figure A.5 is here.

Summary Statistics Table A.1 illustrates the composition of privatization across years by industry. We do observe that privatization is gradual over years. In our sample there are over 1500 privatization events in 1999, and it reduces to 330 in 2006. Moreover, the distribution of the privatization events is uneven across industries: Food processing and Nonmetal Mineral Products industry top the list with more than 1000 privatization events during the sample years.

[12] Recent paper by Song and Hsieh [Forthcoming] and Bai et al. [2009] use capital composition as an alternative measurement of the control right.

[13] There is a small fraction of firms that have multiple switches in terms of their registration types. I drop those firms from the sample.

Table A.1 is here.

Table A.2 decomposes privatization across years and provinces. Most of the privatization occurs in the eastern provinces, specially in Guangdong, Jiangsu and Shandong. By contrast, western provinces in general have small number of privatization.

Table A.2 is here.

Table A.3 provides further summary statistics regarding to the input, output, and other relevant variables used in the empirical analysis.

Table A.3 is here.

1.4.2 Specification

To quantify the effects of privatization on productivity and distortions, I specify the revenue function, productivity and distortion processes as follows. The revenue function (in log-term) of firm i at year t in industry j is

$$r_{it}^{j} = \theta_{k}^{j} k_{it} + \theta_{l}^{j} l_{it} + \theta_{m}^{j} m_{it} + \omega_{it},$$

where $\theta_{k}^{j}, \theta_{l}^{j}$, and θ_{m}^{j} are industry-specific input elasticity for capital, labor and material, respectively.

Productivity Process The evolution process of productivity is

$$\omega_{it} = \mu + \sum_{n=1}^{N} \gamma_n \omega_{it-1}^{n} + \rho^{SOE} \mathbb{1}_{i \in SOE, t-1} + \rho^{Privatized} \mathbb{1}_{Privatized,i,t-1} + \rho^{After} \mathbb{1}_{After,i,t-1} + \alpha \tau_{it} + \phi \tau_{it-1} + \pi x_{it-1} + \eta_{it}. \quad (1.14)$$

On the right hand side, the first term is the intercept of the productivity process. The second term captures how the lagged productivity levels contribute to the current productivity level. Here n is the order of lagged productivity. So I allow for nonlinearity in the productivity process. Following De Loecker [2013] and Braguinsky et al. [2015], I take $N = 3$. The third term captures the effect of lagged state ownership on the productivity process. Here $\mathbb{1}_{i \in SOE, t-1}$ is a dummy variable that equals 1 if and only if firm i is a SOE firm at year $t-1$.

The fourth term captures the transitory effect of privatization on productivity process, and $\mathbb{1}_{Privatized,i,t-1}$ is a dummy variable that equals 1 if and only if firm i is privatized at year $t-1$ (i.e., firm i is a SOE firm at year $t-1$, but a non-SOE firm at year t). The fifth term captures the permanent effect of privatization on productivity. Here $\mathbb{1}_{After,i,t\text{-}1}$ is a dummy variable that equals 1 if and only if firm i was privatized before or at year $t-1$.[14] The seventh term captures the correlation between productivity and distortions for all firms. The eighth term captures how lagged distortion levels contribute to current productivity level. The ninth term captures the effects of exogenous covariates on productivity. In particular,

$$\pi x_{it-1} = \pi^{Export}\mathbb{1}_{Export,i,t-1} + \pi^{IMR} IMR_{it-1} + \pi^{T} Trend_j + \pi^{T2} Trend_j^2 + \pi_t + \pi_p, \quad (1.15)$$

where $\mathbb{1}_{Export,i,t-1}$ is a dummy variable that equals to 1 if firm i reports positive exporting value at year $t-1$; IMR_{it-1} is the inverse mills ratio to control for potential selection bias; $Trend_j$ and $Trend_j^2$ are the industry-specific linear and quadratic time trends; π_t and π_p are the year and province fixed effects.

Now I discuss three aspects of the specification of the productivity process. First, under this specification, both ownership and privatization can potentially change the levels of productivity process. In particular, there are potentially three types of effects on productivity process from privatization: (a) Privatization directly affects productivity process through the term $\rho^{Privatized}\mathbb{1}_{Privatized,i,t-1}$. This is similar to a transitory shock in the productivity process; (b) privatization changes the productivity with a permanent shift of the productivity level through $\rho^{After}\mathbb{1}_{After,i,t-1}$. This is similar to a permanent shock to the mean of the productivity process; and (c) except the above two channels, privatization indirectly changes future productivity via its effect on distortions.

Second, this specification allows both the lagged and current distortions τ_{it} entering the productivity process directly. Recall that firms know current distortions before they know current productivity with information set $\mathcal{I}_{it-1}$. Since there is a correlation between instantaneous productivity and distortions, firms can use current distortions as an additional variable to forecast current productivity. Third, the productivity innovation here is the part of the innovation that is uncorrelated with current distortion.[15]

[14]This is similar to Brown et al. [2006] and Braguinsky et al. [2015].
[15]This is similar to David et al. [2015].

Distortion Process The evolution of the distortion is

$$\begin{aligned}\tau_{it+1} = \mu^\tau + \sum_{n=1}^{N} \gamma_n^{\tau,c} \tau_{it}^n + \sum_{n=1}^{N} \gamma_n^{\tau,l} \tau_{it-1}^n + \\ \mathbb{1}_{\tau_{it}\leq 0} \left(\rho_N^{\tau,SOE} \mathbb{1}_{i\in SOE,t-1} + \rho_N^{\tau,Privatized} \mathbb{1}_{Privatized,i,t-1} + \rho_N^{\tau,After} \mathbb{1}_{After,i,t}\right) + \\ \mathbb{1}_{\tau_{it}>0} \left(\rho_P^{\tau,SOE} \mathbb{1}_{i\in SOE,t-1} + \rho_P^{\tau,Privatized} \mathbb{1}_{Privatized,i,t-1} + \rho_P^{\tau,After} \mathbb{1}_{After,i,t}\right) + \\ \phi^\tau \omega_{it-1} + \pi^\tau x_{it-1} + \eta_{it+1}^\tau. \quad (1.16)\end{aligned}$$

On the right hand side, the first term is the intercept of the distortion process. The second and third terms capture the lagged and current distortions' contributions to the future distortions.

The terms in the second and third line reflect the effects of ownership and privatization on distortions when $\tau_{it} > 0$ (i.e., firm i at year t faces implicit subsidy) and $\tau_{it} < 0$ (i.e., firm i at year t faces implicit tax) respectively. The first term inside the parenthesis captures the effect of lagged state ownership $\mathbb{1}_{i\in SOE,t-1}$ on the distortion process.The second term captures the transitory effects of privatization $\mathbb{1}_{Privatized,i,t-1}$, and the third term reflects the permanent effect of privatization by $\mathbb{1}_{After,i,t}$. The last two terms capture the contribution of lagged productivity and other exogenous covariates on future distortions. The specification of control variables is the same as in the productivity process

$$\pi^\tau x_{it-1} = \pi^{\tau,Export} \mathbb{1}_{Export,i,t-1} + \pi^{\tau,IMR} IMR_{it-1} + \pi^{\tau,T} Trend_j + \pi^{\tau,T2} Trend_j^2 + \pi_t^\tau + \pi_p^\tau. \quad (1.17)$$

where the first term captures the correlation between distortions and exporting status. The second terms is the inverse mills ratio for survivor bias. The thid and fourth terms are the industry-specific linear and quadratic time trend. π_t and π_p are the year and province fixed effects.

Under this specification, both ownership and privatization can potentially change the persistence and levels of distortion processes, conditional on whether the firm is being implicitly taxed or subsidized. In particular, there are potentially three types of effects of privatization on distortions: (a) Privatization directly affects distortion process through the terms $\mathbb{1}_{Privatized,i,t-1}\mathbb{1}_{\tau_{it-1}\leq 0}$ and $\mathbb{1}_{Privatized,i,t-1}\mathbb{1}_{\tau_{it-1}>0}$. Conditional on the sign of τ_{it-1}, privatization acts like a transitory shock in the distortions processes; (b) privatization changes future distortions with a permanent shift of the distortion level through $\mathbb{1}_{After,i,t-1}\mathbb{1}_{\tau_{it-1}\leq 0}$ and $\mathbb{1}_{After,i,t-1}\mathbb{1}_{\tau_{it-1}>0}$, conditional on the sign of τ_{it-1}. This is similar to a permanent shock to the mean of the productivity process; and (c) except the above two channels, privatization indirectly changes future distortions via its effect on productivity.

1.5 Estimation Results

This section presents the estimation results—the joint estimation results of revenue function, productivity process, distortion process and the correlation between the estimated productivity and distortions.

Revenue Function and Productivity To begin with, I estimate revenue functions for 28 two-digit manufacturing industries respectively.Table A.4 reports the estimates of the parameters in revenue function and productivity process for select industries. Column (1) - (5) report the estimation results for the five largest two-digit manufacturing industries.[16] They are Food Processing (13), Textile (17), Raw Chemical and Chemical Products (26), Nonmetal Mineral Products (31), and Ordinary Machinery (35). Column (6) reports the estimation results for all firms across all 28 industries. In addition, I include industry fixed effects in the estimation in column (6).

Table A.4 is here.

I now discuss the estimates that have substantive implications. First, the revenue elasticities of capital, labor and material (i.e., $\theta_k, \theta_l, \theta_m$) vary across industries. On average, elasticities of capital, labor and material range from 0.04 to 0.10, 0.08 to 0.18, and 0.60 to 0.74 respectively. In particular, textile (13) has the highest material elasticity and lowest capital elasticity comparing to the other 4 industries listed in Table A.4.

Second, there is a positive effect of privatization on productivity process. Since the estimates $\rho^{Privatized}$–the coefficient associated with the privatization dummy–in Column (6) indicates the overall effect is 0.093, and the average productivity is around 2, the transitory shock to productivity is about 4.65%. Notice that the coefficient associated with SOE dummy $\rho^{SOE} = -0.09$, which means the productivity of a SOE firm is 0.09 lower than that of an average private firm. The coefficient associated with the after privatization dummy $\rho^{After} = -0.02$ suggests that the privatized firms on average are less productive comparing to the private firms. The difference between ρ^{After} and ρ^{SOE} indicates a permanent increase in productivity for the firms that are privatized.

Third, the estimated γ_1, γ_2, and γ_3 infer that productivity process is persistent. Therefore, the permanent increase of productivity induced by privatization benefits the privatized firms for all the following periods. The higher the persistence level of the productivity process, the larger the effects from privatization.

[16]The top five largest two-digit industries are defined in terms of the number of observations.

Fourth, $\pi^{Exporting}$ captures the correlation between productivity and exporting status. Whether the correlation is positive or negative depends on the industry that the firm belongs to. For instance, the correlation is negative in Textile (13), and positive in Ordinary Machinery (35).[17]

Finally, the estimated ϕ implies that lagged distortions contribute positively to the productivity process. Larger lagged implicit subsidy (tax) is associated with higher (lower) current productivity. The estimated α suggests that the current distortion is negatively correlated with the current productivity. A more productive firm is more likely to have implicit tax instead of implicit subsidy. Consistent with Restuccia and Rogerson [2008], this result suggests that potentially there are large gain from resource reallocation.

To summarize, the estimation of the productivity process suggests that privatization does improve the productivity of the privatized firms by a transitory shock and a permanent increase in productivity. The size of improvement is significant: it immediately realizes a 5% of the productivity increase, which is equivalent to the annual growth rate of the manufacturing sector in China. Since the productivity process is persistent, privatization has long run implications in productivity.

Moreover, I look at the cross sectional distribution of productivity. I decompose productivity based on different firm characteristics such as ownership and geographic differences.[18] Figure A.6 depicts the estimated productivity distributions conditional on ownership (left) and economic regions (right). It is consistent with the estimation results that private firms are more productive comparing to the SOE firms. In particular, firms in Eastern region are more productive than firms in other regions.

Figure A.6 is here.

Distortion Process Now I discuss the estimation results of distortion process. Beginning with the time-series regression, table A.5 reports the estimation results of the distortion process. I report estimates of the five largest industries in column (1)-(5) respectively, and the estimates of the whole sample in column (6).

Table A.5 is here.

There are two aspects of the estimation results. First, SOE firms have larger distortions comparing to private firms. The estimated $\rho_P^{\tau,SOE}$ and $\rho_N^{\tau,SOE}$ capture the asymmetric

[17]The estimates are consistent with the explanation in Lu [2010] that exporting firms are less productive in high labor intensity industries.

[18]See discussions in Hsieh and Klenow [2009] and Brandt et al. [2013].

effects of lagged state ownership on distortions when firms have implicit tax (i.e., $\tau_{it} \leq 0$) and implicit subsidy (i.e., $\tau_{it} > 0$), respectively. The estimated $\rho_P^{\tau,SOE}$ is 0.036. This implies that when restricting at firms with positive implicit subsidy at year t, SOE firms have higher implicit subsidy comparing to private firms. This effect is significant in almost all the industries and in the regression for all the firms. On the other hand, the estimates of $\rho_N^{\tau,SOE}$ is 0.041 for all the firms, which suggests that SOE firms do have significantly higher implicit tax when restricting to firms with implicit tax at year t.

Second, the estimated $\rho_N^{\tau,Privatized}$ and $\rho_P^{\tau,Privatized}$ capture the asymmetric transitory effects of privatization on distortions when firms have implicit tax and subsidy, respectively. Overall, it suggests that privatization lowers the implicit subsidy and tax but these effects vary across industries. The estimates for all the firms in column (6) suggest that privatization results in an immediate 0.08 reduction of the implicit tax and a 0.02 reduction of the implicit subsidy. Notice that the implicit tax (subsidy) for privatized firm at one year before privatization is 0.38(-0.38). These estimates result in a 5% decrease of the implicit subsidy and a 20% decrease of the implicit tax, respectively. However, the regressions for individual industries suggest less significant results. Weighted by the number of firms with implicit tax or subsidy, privatization reduces distortions about 2%. And if we immediately reduce the distortions by 2% on material usage and assume dynamic inputs are fixed, the reduction of distortions results in a 0.2% in terms of total revenue. Therefore, even though in the overall regressions there are statistical significant effects of privatization on distortions, the economic effects are at best second-order. I further estimate the productivity and distortion effects using propensity score matching in appendix A.2. The matched results produce similar patterns.

To summarize, the estimation of the distortion process suggests that privatization does improve the distortions of the privatized firms by a transitory shock and a permanent shift in distortions. It results in significant reduction of implicit tax and subsidy faced by those privatized firms. The magnitudes of the effects are large, especially for firms with implicit tax.

I then turn to the cross-sectional distributions of distortions. Figure A.7 depicts the estimated distortion distributions conditional on ownership (left) and economic regions (right). In the sample, most of the observations are associated with implicit subsidy, while only a small fraction of the observations is associated with implicit tax. Moreover, it is consistent with the time-series estimation results that SOE firms have more dispersed distortion comparing to private firms. Last, firms in Eastern region have higher implicit subsidy than other regions.

Figure A.7 is here.

I further correlate the estimated distortions with firm characteristics like ownership, exit dummy, exporting status and subsidy dummies. I regress the value and the absolute value of distortions $|\tau_{it}|$ on firm characteristics while controlling for firm level covariates including firm size, age, age square, year fixed effect, region fixed effect and industry fixed effect. Table A.6 reports these estimates. These results suggest that a) the exit firms have larger absolute value of distortions; b) similar to the previous results, SOE firms have larger dispersion of distortions; c) exporting firms have smaller absolute value of distortions, and d) the subsidy dummy correlates positively with the level of distortion measure, which validates the implicit subsidy explanation when $\tau_{it} > 0$.

Table A.6 is here.

Finally I analyze the correlation between productivity and distortions at the micro-level. Restuccia and Rogerson [2008] suggests that with uncorrelated distortions with productivity, the welfare consequences of distortions are relative small. With correlated distortions, there is a larger potential gain to remove distortions. Figure A.8 plots micro level productivity and distortions by industry. It is quite obvious that productivity are negatively correlated with distortions, i.e., more productive firms face implicit tax, and less productive firms in general receives implicit subsidies.

Figure A.8 is here.

1.6 Counterfactual Analysis

This section conducts counterfactual analysis under different environments. To do so, I take a balanced sample of firms that were eventually privatized within the sample period. This sample consists of 4000 firms from 1998 to 2007, which is about half of the privatized firms in the full sample. During the sample period, these firms experienced significant productivity growth. The median productivity is 2.19 in 1998 and 2.61 in 2007. The firm level distortions also change significantly. The median distortion is 0.23 and standard deviation is 0.42 in 1998. In 2007 they change to 0.19 and 0.37, respectively.

Next, I use the estimated productivity and distortion processes to generate the counterfactual predictions of productivity and distortion under three different environments. Then I compare the predicted distributions with the realized ones. The first and second counterfactual experiments alter the timing of privatization. More specifically, the first counterfactual experiment asks what would change if there is no firm privatized. The second counterfactual experiment asks what would change if all SOE firms privatized at the beginning of the sample. The third counterfactual takes the timing of privatization as given, and decompose the effects of privatization into several components. Notice that privatization affects productivity and distortions via three effects: (a) The transitory effects captured by the privatization dummy $\mathbb{1}_{Privatized,i,t-1}$. (b) The permanent effects captured by the after dummy $\mathbb{1}_{After,i,t}$. (c) The indirect effects from lagged distortions and productivity. The third counterfactual asks how much do transitory, permanent and the indirect effect of privatization affect the end of sample productivity and distortions distributions. I report the counterfactual experiments as follows.

No Privatization I simulate the productivity and distortion process assuming privatization has no effects on both process. Alternatively this could be interpreted to have delayed privatization till end of the sample. Figure A.11 left panel illustrates the counterfactual productivity and distortion distributions. It shows that without privatization, the productivity distribution shifts toward left with the median productivity drops from 2.61 to 2.31. This shows that privatization overall increase productivity by 14% for the privatized firms.However, even though the distortion distribution moves closer to 0, the optimal margin revenue equals marginal cost case, privatization does not contribute significantly to the reduction of distortions. Therefore, the productivity effects is much more important comparing to the reallocation effects from privatization.

Early Privatization I then alter the timing of privatization for these eventually privatized firms. The fractions of privatized firms are quite smooth over year: With on average 500 firms per year are privatized between 1998 to 2003, and 200 firms per year between 2004 to 2006. Figure A.11 right panel illustrate the counterfactual productivity and distortion distributions when every firm in this sample is privatized at 1998. The overall productivity distribution shifts to the right with median productivity increases from 2.61 to 2.70. This suggests that early privatization would bring additional 4% of productivity. Similar to the previous counterfactuals, altering the timing of privatization does not results in significant change of distortion distributions.

This results highlight the differences between the effects of privatization on productivity and distortion process. In productivity process, the transitory component is significant and the timing of privatization contributes significantly to the productivity process. Here though the estimates suggest a statistical significant result from privatization on distortions, the decomposition results suggest that the reduction of distortions may be driven by factors other than distortions.

Decompose Privatization in Productivity Process The counterfactual results indicate that the timing of privatization does matter since the productivity process is persistent and privatization has both significant transitory and permanent effects on productivity. In the final counterfactual, I further decompose the contributions from the transitory and permanent components while fixing distortions and the true privatization timing.

Conditional on the true privatization date, I decompose the productivity gain from privatization into various components. From the section 1.4, there are three main channels: a) privatization acts as an transitory shock; b) privatization shifts the mean productivity; c) privatization changes the persistence of productivity process. To see how much the transitory component contributes to the productivity, I conduct additional two counterfactual experiments. The first one is to set $\mathbb{1}_{Privatized,i,t-1} = 0$ to shut down the transitory component. Figure A.12 left panel depicts the counterfactual productivity distribution, with the median productivity is about 2% lower from 2.61 to 2.55. The second one is to set $\mathbb{1}_{After,i,t} = 0$ to shut down the permanent effect. Figure A.12 right panel depicts this counterfactual productivity distribution. Even with the transitory component, the mean productivity drops significantly: it changes from 2.61 to 2.36, which is about a 10% decrease in mean productivity. The decomposition results suggest that the permanent effects accounts for over 80% of total privatization effects on productivity, whereas the rest 20% comes from the transitory component.

Figure A.12 is here.

To summarize, when holding distortions the same, privatization affects firm level productivity the most by shifting the mean of the productivity process. The transitory component, though with a smaller magnitude, still contributes significantly to the productivity distribution. These two effects, joint with the persistence of productivity shape the realized productivity distribution at 2008.

1.7 Aggregate Productivity Growth Decomposition

To understand the effect of privatization on aggregate productivity growth (APG), I extend the decomposition method proposed by Melitz and Polanec [2015] and Collard-Wexler and Loecker [2015] to explicitly account for the change in ownership. I then apply this method using the estimated productivity to decompose the APG in the data from 1999 to 2007.

Denote $\triangle\Omega$ the aggregate productivity growth between 1999 and 2007 in the sample. I show in appendix A.3 that the aggregate productivity growth can be decomposed into various components using the following formula

$$\triangle\Omega = \frac{1}{2}\sum_{\phi}\Bigg[\underbrace{\triangle\bar{\omega}_t(\phi)}_{\text{Productivity Growth}} + \underbrace{\triangle\Gamma_t^{OP}(\phi)}_{\text{Reallocation Within Ownership}}\Bigg] + \underbrace{\triangle\Gamma_t^{B}}_{\text{Reallocation Between Ownership}} + \underbrace{\triangle\Gamma_t^{T}}_{\text{Transition Between Ownership}} + \underbrace{s_{E2}(\Omega_{E2}-\Omega_{S2})}_{\text{Entry}} + \underbrace{s_{X1}(\Omega_{S1}-\Omega_{X1})}_{\text{Exit}}. \quad (1.18)$$

Denote ϕ the index of ownership with $\phi = 1$ for SOE firms and $\phi = 0$ for private firms. The first term on the RHS captures the average within firm productivity improvement denoted as $\triangle\bar{\omega}_t(\phi)$, and reallocation effects within the particular ownership as $\triangle\Gamma_t^{OP}(\phi)$. The second term captures the reallocation between the two ownerships, denoted as $\triangle\Gamma_t^{B}$. And the third term $\triangle\Gamma_t^{T}$ captures the direct contribution from ownership switches to the aggregate productivity growth. The fourth and fifth terms capture the contributions from entry and exit, respectively. And notice that this decomposition uses the survivor groups to benchmark the entrants and exiters. I use the employment share as weight s_{it} in this exercise.[19]

Table A.7 shows the decomposition results for the aggregate productivity growth using equation 1.18. There are three main results from this decomposition exercise.

Table A.7 is here.

First, there is a large aggregate productivity growth within the manufacturing sector in the sample. During 1998 to 2007, the APG within the sample is 58%, which translates to a 5.2% annual growth rate.

Second, further decompose the productivity within each ownership, I find an average SOE firm became 44 percent more productive between 1999 and 2007. In addition, within the state sector, reallocation toward more-productive firms contributes positively but much less in terms of magnitude: It only accounts for 2 percent the increase of aggregate productivity. In the meanwhile, an average private firm became 37 percent more productive, which is lower than an average SOE firm.[20] However, the reallocation

[19] Collard-Wexler and Loecker [2015] uses market share for the weight. Here since I use revenue productivity, I use a specific production factor for the weight. The results are similar using other production factor or revenue.

[20] Song and Hsieh [Forthcoming] also found similar patterns using different approach.

process plays a large role within the private sector. The reallocation within the private sector contributes nearly 10% of the aggregate productivity growth. Overall, this finding indicates that during the transition period, within firm productivity growth dominates the reallocation process in aggregate productivity growth.

Third, I find that the between-ownership reallocation component contributes negatively to the aggregate productivity growth. In 1998, the between covariance was 0.5 percent. However, it changes to -6 percent at the end of the sample.

Fourth, firm dynamics is still an important source of aggregate productivity growth. In total it contributes 6 percent to the aggregate productivity growth. However, the exit margin contributes positively and entry margin contributes negatively. This is a general feature of Melitz and Polanec [2015] as we use survivors as benchmark. The results suggest that the exiters are inferior in terms of their productivity, and so as the entrants.[21]

Fifth, privatization alone accounts for 4 percent in the aggregate productivity growth. Given in the sample there are over 8000 firms are privatized and only half of them are presented in the decomposition analysis, this number is large. From the previous discussion, privatization affects aggregate productivity mainly through its impact on firm-level productivity.

To summarize, privatizations contribute to the aggregate productivity growth mainly through their direct impact on firm level productivity. The decomposition results show that while within firm technology improvements account for more than 80% of aggregate TFP growth, the privatizations accounted for almost all of the contributions from the transition firms. Given that only a limited number of firms are privatized, there is a fairly sizeable gain associated with the policy. In contrast, although the privatizations also reduce revenue distortions, the associated reallocation gains only contribute a small fraction to overall TFP growth.

1.8 Discussion: Distortions and Dynamic Inputs

The analysis can also address an important question in the misallocation literature: How to measure distortions? Literature typically uses the dispersion of Marginal Revenue Product of Capital (henceforth MRPK) to measure distortions. (See Hsieh and Klenow [2009].) But Asker et al. [2014] argues that, when capital is a dynamic input, even

[21]The results are consistent with recent paper by Garcia-Macia et al. [2015]. However, notice that I am using revenue productivity, the inferior productivity could be due to the low prices entrants charge, but not their quantity productivity.

in a distortion-free economy, dispersion in productivity contributes to the dispersion in MRPK. This implies that the dispersion of MRPK may not be a good measure of distortions with dynamic inputs.

Since my framework allows both distortions and dynamic inputs, I can address how much the dispersion of MRPK can be attributed to the distortions with dynamic inputs. Notice that I estimated productivity innovation and distortions in Section **??**. Once I construct the measure of MRPK, I can evaluate how productivity innovation and the dispersion of distortions influence the dispersion of MRPK. Now I discuss how I measure MRPK.

Follow Asker et al. [2014], I measure the firm-year level MRPK as follows

$$MRPK_{it} = \log(\theta_k) + r_{it} - k_{it}, \tag{1.19}$$

where r_{it} is log revenue, k_{it} is log capital, and θ_k is the underlying revenue elasticity for capital.

Then for each industry-year, I compute the the dispersion of MRPK. Notice that the term $\log(\theta_k)$ of Equation (1.19) is a constant for industry level data. Thus, when I compute the dispersion of MRPK, it is equivalent to compute the unadjusted firm-year level MRPK

$$MRPK^U_{it} = r_{it} - k_{it}.$$

Figure A.9 illustrates how productivity innovation (i.e., the volatility of TFPR) and the dispersion of distortions correlate with the distortion of MRPK. Each dot on the graph representing a specific industry year pair, and the size of the dot reflects the relative weights by total revenue of this particular industry year combination. The red solid line is an unweighted linear fitting line, and the blue dash line is an weighted linear fitting line. The left graph suggests that the dispersion MRPK is positively correlated to productivity innovation. The right suggests that the dispersion of MRPK is also positively correlated to the dispersion of distortions .

Figure A.9 is here.

Next, I regress the dispersion of MRPK on productivity innovation and dispersion of distortions. I include industry and year fixed effects to control for other potential d-ifferences across industries and years. Table A.8 reports the regression results. The first two columns present the regression results of dispersion of MRPK on productivity innovation. The last two columns present the regression results of dispersion of MRPK on dispersion of distortions. The results show that both productivity innovation and the dispersion of distortions have a positive effect on the dispersion of MRPK.

Table A.8 is here.

The fact that there is a significant positive relationship between the dispersion of M-RPK and productivity innovation is consistent with the result in Asker et al. [2014]. This implies that, at industry level, productivity innovation does feed into dispersion of MRPK. Therefore, dynamic inputs are important factors contributing to the dispersion of MRPK. But this does not necessary imply that dispersion of distortions does not contribute to dispersion of MRPK. In fact, the dispersion of MRPK does indicate distortions. I now explain the evidence.

Figure A.10 plots the relationship between dispersion of distortions and productivity innovation. It shows that there is a significant positive relationship between the two. Thus, in the region with higher value productivity innovation, there is larger dispersion of distortions. This fact provides a rationale to support the idea that unproductive firms survive because they receive more implicit subsidy, i.e., larger τ_{it}. To this end, these regression results clearly suggest that the dispersion of MRPK can provide information on both productivity and distortions, and these two channels are not mutually exclusive. To sum up, the results are consistent with both Hsieh and Klenow [2009] and Asker et al. [2014]: First, the dispersions of MRPK do indicate distortions in the economy. Second, productivity innovation indeed amplifies dispersion of MRPK. In total, these two forces could explain a large proportion of dispersion in MRPK in the data.

Figure A.10 is here.

1.9 Conclusion

This paper investigates the effects of privatization on productivity and distortions at both firm and aggregate levels. To do so, the paper first provides a novel method to identify and jointly estimate distortions, productivity and production function parameters from a standard production dataset. Next it goes on to provide evidence on the direct effects of privatization. Finally, the results suggest that aggregate productivity growth is mainly driven by the productivity improvement within firms. Despite much privatization, state-owned sector has grown faster relatively to private sector.

The results indicate that privatization does have large implications for productivity and distortions. The transitory and permanent increases of productivity after privatization stand out to be the most important forces that shape both firm-level and aggregate outcomes. Though it reduces distortions at firm level, the effects from distortions are quite limited in the data.

There are many caveats required in this paper. First, in this paper I only allow firm-year level idiosyncratic output distortions. That is, the distortions move firm's marginal revenues away from its marginal costs for all inputs to the same extent. However, there are large capital and labor distortions at firm year level in China as well. Since the current model only employs first order condition from the static input, it is possible to allow both capital and labor distortions.

Second, the framework in this paper could be applied to other policies, for instance, trade liberalization, subsidy program and government regulations. In doing so, it articulates the trade off of implementing certain policy–productivity enhancing and distortion reduction.

Third, there are many interpretations for the deviation of the firm level first order conditions: Different technologies, measurement error, dispersed input prices and many others. This paper emphasizes one particular interpretation: the output distortions. To account for other plausible sources that lead to similar findings is an important topic for future research.

Chapter 2

Understanding Transitions using Directed Search

This paper explains how a directed search model can be used to understand worker transition data in labor markets. The basic theory provides a dynamic extension of the model in Peters [2010] in which workers have privately known types that are observable to firms once workers apply. The purpose is primarily to derive the search outcome distribution and to show how the wage distribution and search outcome distribution are related, and how this relationship yields a very testable theory of worker transitions between jobs. We suggest a way to test this relationship.

The basic idea is that type information is incorporated into workers' search decisions, so that the wage at which worker is currently employed reveals something about his or her type. In particular, workers who are currently employed at high wage firms are more likely to have high types. As a consequence, if they are forced to move to new jobs, or move as a result of on the job search, they are more likely to receive higher wages at these new jobs as well.

This idea is not new. The main purpose of the exercise here is to try to establish whether directed search can impose additional structure on this relationship that might ultimately make it possible to distinguish between directed search and other models like random matching that might yield a similar result. The main results in the paper consist of a series of theorems relating the wage offer distribution to the the properties of the relationship between workers wages before and after a job transition.

For example, suppose that the wage offer distribution is given by $G(w)$ on some compact interval $[\underline{w}, \overline{w}]$of wages. We show that the search outcome distribution for any worker type is given by

$$\tilde{G}_{w_0}(w) = \frac{\int_{w_0}^{w} \frac{dG(\tilde{w})}{\tilde{w}}}{\int_{w_0}^{\overline{w}} \frac{dG(\tilde{w})}{\tilde{w}}}$$

for some wage w_0.

This fact can be used to suggest various empirical tests. For example, we show that if the variance of this truncated distribution is a decreasing function of w_0, then the variance of a worker's future wage should be a decreasing function of the wage at the job the worker currently holds. Similarly, we can show conditions on this derived distribution in which the relationship between the workers old and new wage is convex or concave.

The model we provide here is richer in empirical content that most models of directed search. In its most basic variant, directed search assumes all firms are identical and offer the same wage in equilibrium. Models that do allow for heterogeneity among firms (for example, Peters [2000]), still assume workers are identical but use mixed application strategies when they apply to firms, applying with highest probability at the firms who offer the highest prices. In the steady state of such a model, if workers use the same mixed strategy in every period, their wages will be variable over time. However, there will be no correlation at all between the wage they receive in the match that they leave and the match that they move to.

At the other extreme, models that support pure assortative matching (for exampleShi [2001] or Eeckhout and Kircher [2010]) will predict that workers who land high wage jobs in one period will do so again in future periods. Theoretically, outcomes are perfectly correlated over time. The same kind of outcome could be expected from wage-ladder like models (e.g., Delacroix and Shi [2006]) in which homogeneous workers search on the job and implicitly use the current wage as a way of coordinating applications. Workers who are employed at some wage will apply to firms offering slightly higher wages until they are successful at finding a new job. This provides a high correlation between the wages of workers who move between jobs without an unemployment spell. This correlation is broken when workers' matches are terminated exogenously and they experience unemployment. They then fall to the bottom of the wage ladder. The basic prediction is very high correlation in wages for job to job transitions, and no correlation (or perhaps a negative correlation) for movements between jobs that involve an unemployment spell.

Older models of directed search then, either seem to predict no correlation of an individual workers' wages over time, or a nearly perfect correlation. The model we develop here produces a correlation in between that is more in line with empirical evidence (some of which we show below).

Random matching models with distributions of worker types can also be used to produce an intermediate correlation between wages across transitions, though we are not aware of a model that has explicitly studied this. For example, Burdett and Mortensen [1998] studies a model with wage posting in which firms are identical but workers are described by an atomless distribution of supply prices. Workers who need high wages to convince them to work will tend to be paid high wages in each of their jobs, tending to support high, but not perfect, correlation in wages over time. However, they do not explicitly study transitions.

Our approach differs from their in two ways. First, we do not assume that firms are identical. This generalization means that our model can be consistent with a larger set of wage distributions. Second, our distribution of supply prices for workers is endogenously determined by the wage offer distribution. It is this relationship between supply prices and the wage offer distribution that supports all of the empirical implications in our model.

Postel-Vinay and Robin [2002] study a model with random matching and bargaining with heterogeneity in worker (and firm) types. After being randomly matched with some worker, a firm makes the worker a take it or leave it offer. The firm is assumed to observe the worker's productivity type, which is the same assumption we make here. Workers with higher types have better outside options. As a consequence their wages tend to be high in all their matches. However, their purpose is not to study these correlations, so they do not derive the outcome distribution for workers as we do here.

There are two major differences between our model and theirs. First, from the theoretical perspective, we do not use discounting. Instead we assume that workers maximize the limit of their average expected wage payments. Random matching models need discounting to support match frictions. Without it, workers and firms simply wait around until they are assortatively matched. Directed search doesn't require the same device because frictions are built directly into the matching process because of workers inability to coordinate their search strategies. The advantage of our approach is that it allows us to draw a much closer connection between the wage distribution and workers search outcomes than is possible with discounted payoffs.

One of the implications of our assumption is that search strategies no longer depend on the wage at which a worker is currently working. In typical wage ladder models, workers' search decisions collapse to what are effectively pure strategies in which workers with higher wages only apply to higher wages. This supports too much autocorrelation in wages of workers who search on the job. As this effect disappears in our model, wage variation after on the job transitions is instead attributable to workers' unobservable types.

Second, of course, is the fact that we use directed search. This assumption makes it possible to directly tie the search outcome distribution to the wage offer distribution in a way that is not possible when the two are indirectly connected through a value function. This approach makes it possible for us to provide a non-parametric test of the model.

2.1 Fundamentals

A labor market consists of measurable sets of positions and workers. It will be assumed that all the workers in this market are identical in terms of observable qualifications, including education, experience, past performance, etc. So all workers are acceptable employees at all firms. However, workers also possess observable but non-verifiable types that are potentially valuable to firms.

Types might be things like the potential employee's charm and articulation, or references that the employee gets from outsiders. An example of what we have in mind may be the academic market for newly graduated phd's where employers need to see reference letters and conduct interviews before they make offers. As in the academic market, the 'market' doesn't know workers' types in the sense that firms don't know workers' types before they apply, and workers don't know each others' types. However, we assume that firms can identify workers' types and rank them once they contact them with an application. So rather than randomly selecting among applications as is typical in models of directed search, firms choose among workers on the basis of these types. Firms rank these types the same way, so a type moves between jobs with a worker. This is the property of the model that connects the outcomes for workers as they move between positions.

Workers types are contained in a compact subset Y of $\mathbb{R}_+$. The measure of the set of searching workers with types less than or equal to y is given by $F(y)$, where F is a monotonically increasing and differentiable function defined on the interval $Y = \left[\underline{y}, \overline{y}\right]$. Types have nothing to do with worker preferences, which are assumed to be the same for all workers. The measure of the set of workers will be normalized to 1.

We'll model jobs as renewable short term contracts offering fixed wage payments. Workers compete for these contracts by applying for them. Following the random matching literature, we'll assume that when a worker's application to a new job is successful, she will give up the lowest paying job that she currently holds. We'll differ from the random matching literature by assuming that she will have to fulfill an existing contract, unless it is terminated by her employer. So a worker who searches on the job and wins a new contract will temporarily hold two jobs, one of which she will relinquish at the end of the period. This is convenient, since multiple job holders do exist in the French labor market data we use.

The assumption that workers will accept a second job when they already have one is reasonable - though it is far from the norm, it seems to occur with some frequency in our data. Contract workers will typically seek new work while working on an existing job simply because existing contracts are subject to termination. This might well involve a period of overlap between the old job and the new one. This is the behavior we are trying to capture with this assumption. Of course, employees moonlight in order to augment there income. However, an implication of our assumption is that workers will give up their lower paying contract even though it pays a wage that is higher than what they can expect to earn by searching. This assumption reflects possible restrictions imposed by employers preventing employees from working at other firms, or employee capacity constraints that prevent them from working multiple contracts for long periods of time.

So we assume that a worker begins each period either unemployed, or in a match. In either case, workers apply to one and only one position. We view the 'single application' assumption as a modelling device used to approximate a frictional matching process. As such it is no better or worse than assuming that a worker is randomly contracted by a firm. So we do not discuss it further.

If the worker is already employed under some contract, we call this search on the job. If the worker is offered the job to which he or she applied, they accept it and earn the corresponding wage in that period. After the outcome of the worker's search decision is made, the worker's existing match, if he has one, may be exogenously terminated. This occurs with probability γ. If termination occurs, the worker is either unemployed, or earns the wage associated with any new job the worker managed to get during the period. If the original job does not terminate, the worker is paid by both firms, then resigns from the lower wage job at the end of the period.

We use the limit of the average of expected payoffs as the objective for workers. Since we will focus entirely on steady state equilibrium, this simply means maximizing expected payments, period by period.

At the end of each period, firms either have an unfilled position caused by exogenous termination, an unfilled position caused by the fact that their existing worker has resigned to move to a higher paying firm, or have a continuing employee. If the position is unfilled, the firm enters the market at the beginning of the next period, and advertises an opening for their position.

Positions are parameterized by some characteristic $x \in X$, where X is a compact subset of $\mathbb{R}$. Associated with each position is an optimal wage offer. As mentioned above, we aren't much interested in the firm's optimization problem. So we'll characterizes the population of firms with the distribution of wages these firms offer. We'll use E to represent the accepted wage distribution and G to represent the wage offer distribution. Generally, both distributions will be assumed monotonically increasing and differentiable. The notation dE and dG will be used to refer to the densities of these distributions.

Firms offer contracts that specify the expected wage payment w a worker will receive in each period during which he or she is employed in the position. Firms set wages to maximize the expected profit they earn from whichever worker they hire. A position of type x filled by a worker of type y under a contract that provides expected payment w to a worker, generates an expected per period profit to the firm of $v(w, x, y)$.

2.2 The Market

We model the labour market as a large game in which the payoffs that players receive depend on their own actions, and on the distributions of actions taken by the other players and focus on a steady state equilibrium in which the distribution of wages on offer from open positions supports a distribution of expected payments that does not change over time.

The market begins each period with a set of matched workers and firms, and a set of vacant positions. The presumption in the paper is that each of the worker firm matches consists of a wage type pair. the notation $e(w, y)$ will be used to denote the density of the joint distribution of matched workers and firms. The unknown distribution of types employed at wage w given by $\psi(\cdot|w)$, so $E(w) = \int_{\underline{w}}^{w} \int_{\underline{y}}^{\overline{y}} e(\tilde{w}, y)\, d\psi(y|\tilde{w})\, d\tilde{w}$.

Firms have to decide what wage to offer, while workers have to decide where to apply. We'll focus on symmetric equilibrium, so we can write the firm's strategy rule as $\rho : X \to \mathbb{R}$. As in any directed search model, we expect the workers to use mixed application strategies. So write $\pi : [\underline{w}, \overline{w}] \times [\underline{y}, \overline{y}] \to [0, 1]$ to be the probability that a worker of type y applies to a firm offering a wage less than or equal to w. We assume that for each y, $\int_{\underline{w}}^{\overline{w}} d\pi(w, y) \leq 1$.

A symmetric worker application strategy π gives rise to a distribution P of applications, where $P(w, y)$ is the measure of the set of applications made to firms whose wages are not higher than w by workers whose types are no higher than y. This distribution is given by

$$P(w, y) = \int_{\underline{y}}^{y} \int_{\underline{w}}^{w} d\pi(\tilde{w}, \tilde{y})\, dF(\tilde{y}). \tag{2.1}$$

Since P is absolutely continuous with respect to Lebesque measure on $\left[\underline{y}, \overline{y}\right]$ and with respect to G on the interval $[\underline{w}, \overline{w}]$, we can write

$$P(w, y) = \int_{\underline{w}}^{w} \int_{\underline{y}}^{y} p_{\tilde{w}}(\tilde{y})\, d\tilde{y} dG(\tilde{w})$$

for some (Lebesque) measurable function $p_{\tilde{w}}(\tilde{y})$[1].

Heuristically, the function $p_w(y)$ is the ratio of the measure of the set of workers of type y who apply to firms offering wage w to the measure of firms offering wage w. In other words, it is a variant of the 'queue size' that is so commonly used in directed search. In an urn ball matching model, the probability that a worker is hired when he or she applies at a firm offering the wage w is the exponential of the negative of the queue size.

An analogous formula applies when workers have unverifiable types that are used to determine who is hired. The difference is that in the standard model with identical workers all the the other workers who apply at the same wage are potential competitors. In the model here, if a worker has type y, only workers who apply at the same wage and have higher types are competitors.[2] So the appropriate queue size is the ratio of the measure of the set of workers who apply at wage w and have types higher than y to the measure of the set of firms offering wage w. In other word

$$\int_{y}^{\overline{y}} dp_w(\tilde{y}),$$

is the appropriate queue size. So we use the familiar formula $e^{-\int_y^{\overline{y}} dp_w(\tilde{y})}$ to give the probability that the worker will be hired if he applies at wage w.[3]

Since workers maximize average expected wages, the payoff to a worker of type y who is employed at wage w and applies at a firm offering wage w' as

$$w' e^{-\int_y^{\overline{y}} dp_{w'}(\tilde{y})} + (1-\gamma) w. \tag{2.2}$$

Of course, if the worker is unemployed, the w term is just 0. Since any wage that maximizes the first term maximizes this expression for any w, a worker will maximize his expected payments from firms by maximizing the first term in every period.

Given the measure P, we can write down the probability that a worker leaves his current position at the end of any period. It is the probability the worker applies to and is hired

[1]This follows from the Radon-Nikodym theorem

[2]In Peters [2010] it is shown that the formulas that follow coincide with limits of the payoffs that workers receive in finite markets.

[3]A formal derivation of this probability as the limit of the probability of being hired in a large finite game is given in Peters [2010].

by a firm paying a wage that is higher than his current wage, plus the probability that the match is exogenously terminated. This formula is

$$Q(w,y) = \gamma + (1-\gamma)\int_w^{\overline{w}} e^{-\int_y^{\overline{y}} dp_{w'}(\tilde{y})} d\pi\left(w',y\right). \tag{2.3}$$

A firm who hires a worker retains him or her until the match terminates. Firms who have multiple applications hire the highest type worker who applies and set wages to maximize the expected profit generated by whatever worker they hire. So an unfilled position has value

$$\max_w \left\{ \int_{\underline{y}}^{\overline{y}} \frac{v(w,x,y) e^{-\int_y^{\overline{y}} dp_w(\tilde{y})}}{Q(w,y)} dp_w(y) \right\} \tag{2.4}$$

where $Q(w,y)$ is defined by (2.3).

Finally, in a steady state, a firm who has hired a worker at wage w and loses the worker, either because the worker leaves for higher pay, or because the match is terminated for some exogenous reason, should post a new offer with same wage w that it offered before. The reason is simply that the wage that the firm pays its existing worker is the one that maximized the expression in (2.4) when the firm attracted that worker in the first place.

The steady state condition is then given in a manner similar to the other formulas above. Firms who offer wage w and employed a worker of type y in the last period enter the market looking for a new hire if their worker decided to move during the last period. The 'measure' of firms in this position is $dE(w)\int_{\underline{y}}^{\overline{y}} Q(w,y)\, d\psi(y|w)$. Also joining the market is a set of firms whose employee left two periods ago, but who were unable to hire a new worker last period. The measure of this set is

$$dE(w)\int_{\underline{y}}^{\overline{y}} Q(w,y)\, d\psi(y|w)\left(1 - \int_{\underline{y}}^{\overline{y}} e^{-\int_x^{\overline{y}} dp_w(\tilde{y})} dp_w(x)\right).$$

Similarly, there are firms who lost their worker three periods ago, but failed to hire in the previous two periods, and so on. Adding all these gives the measure of firms who are looking for new workers and offering wages less than or equal to w as

$$G(w) = \int_{\underline{w}}^{w} \frac{\int_{\underline{y}}^{\overline{y}} Q(\tilde{w},y)\, d\psi(y|\tilde{w})}{\int_{\underline{y}}^{\overline{y}} e^{-\int_x^{\overline{y}} dp_{\tilde{w}}(\tilde{y})} dp_{\tilde{w}}(x)} dE(\tilde{w}). \tag{2.5}$$

In this formulation, both $Q(w,y)$, $\psi(\cdot|\cdot)$ and p_w depend on the wage offer distribution G, so that the steady state wage offer distribution is a fixed point. This fixed point is developed formally below.

An equilibrium for this model is a collection $\{E, G, \pi\}$ satisfying three conditions

- (optimality of search strategies) for every y w maximizes (2.2) for every w in the support of $\pi(\cdot, y)$;
- (optimality of wage offers) for every x, $W(x)$ maximizes (2.4); and
- (steady state condition) The relation (2.5) holds almost everywhere for G.

2.3 Continuation Strategies

The approach we are going to take here is somewhat unusual. Rather than starting with a fixed distribution of firm types, then deriving the equilibrium distributions E and G, we will instead take the wage distribution E to be exogenously given. The search strategies needed to support that distribution can then be derived by solving a fixed point problem to find G. These strategies will, in turn determine firms' profit functions. At this point, we just imagine that firms profit functions are distributed in a way that supports the observable wage distribution.

To find the wage offer distribution, we begin by assuming that we know it, then work out the search strategies that satisfy (2.2). These strategies determine the transition function (2.5) which provides a fixed point problem whose solution identifies the wage offer distribution.

What the following theorem says is that if the wage offer distribution is given by G, workers apply at every wage above a type dependent reservation wage $\omega(y)$ with *equal* probability. This reservation wage is increasing in type. In this sense, the model resembles random search and matching models with worker types in which higher type workers hold out for higher wages in the future because they know they can get them. The logic here differs in that workers trade off wage against trading probability as they do in all directed search models.

Theorem 2.1. *For any differentiable wage offer distribution G, there is a continuation equilibrium characterized by a monotonically increasing reservation wage strategy $\omega(y)$ in which each worker applies with equal probability at every wage at or above* $\max[\underline{w}, \omega(y)]$. *Formally, for every y*

$$\int_{\underline{w}}^{w} d\pi(\tilde{w}, y) = \int_{\omega(y)}^{w} d\pi(\tilde{w}, y) = \int_{\omega(y)}^{w} \frac{dG(\tilde{w})}{G(\overline{w}) - G(\omega(y))}.$$

The reservation wage is characterized by the solution to the differential equation

$$\omega'(y) = \frac{\omega(y) F'(y)}{G(\overline{w}) - G(\omega(y))} \tag{2.6}$$

through the point $(\overline{y}, \overline{w})$. *Finally for every wage* w *in the support of* G, *the queue size faced by a worker of type* y *who applies for a position offering wage* w *is*

$$\int_y^{\overline{y}} dp_w(\tilde{y}) = \int_y^{\omega^{-1}(w)} \frac{1}{G(\overline{w}) - G(\omega(y'))} dF(y') \tag{2.7}$$

The proof of this theorem (which follows the logic in Peters [2010]) is given in the appendix. What gives this theorem most of its power is the fact that the only way that a worker's search strategy depends on y is through the reservation wage. We exploit this property extensively in what follows.

Full Equilibrium

The wage that firms offer determines the quality of their applicants as well has how long an applicant stays in a job. The results in the previous section provide a useful way to view this trade off.

Lemma 2.2. *In a symmetric steady state equilibrium, the function* $Q(w, y)$ *is equal to*

$$\gamma + (1-\gamma) \frac{\omega'(y)}{F'(y)} \int_w^{\overline{w}} \frac{dG(\tilde{w})}{\tilde{w}}. \tag{2.8}$$

Proof. From Theorem 2.1, the function $Q(w, y)$ can be written as

$$\gamma + (1-\gamma) \int_w^{\overline{w}} e^{-\int_y^{\overline{y}} dp_{\tilde{w}}(\tilde{y})} \frac{dG(\tilde{w})}{G(\overline{w}) - G(\omega(y))} =$$

$$\gamma + (1-\gamma) \int_w^{\overline{w}} e^{-\int_y^{\omega^{-1}(\tilde{w})} \frac{1}{G(\overline{w}) - G(\omega(y'))} dF(y')} \frac{dG(\tilde{w})}{G(\overline{w}) - G(\omega(y))}. \tag{2.9}$$

Since workers used mixed application strategies, they must receive the same payoff from every seller. So

$$e^{-\int_y^{\overline{y}} dp_{\tilde{w}}(\tilde{y})} = \frac{\omega(y)}{w}.$$

Substituting this into (2.9) gives

$$Q(w, y) = \gamma + (1-\gamma) \int_w^{\overline{w}} \frac{\omega(y)}{\tilde{w}} \frac{dG(\tilde{w})}{G(\overline{w}) - G(\omega(y))}.$$

From (2.6) this becomes

$$Q(w, y) = \gamma + (1-\gamma) \int_w^{\overline{w}} \frac{\omega'(y)}{F'(y)\tilde{w}} dG(\tilde{w}) =$$

$$\gamma + (1-\gamma)\frac{\omega'(y)}{F'(y)}\int_{w}^{\overline{w}}\frac{dG(\tilde{w})}{\tilde{w}}.$$

□

An immediate corollary of this Lemma gives the first useful result:

Theorem 2.3. *In a symmetric steady state equilibrium, a worker is more likely to leave a job the higher is his or her type.*

Proof. By (2.6) and the fact that ω is increasing, $\frac{\omega'(y)}{F'(y)}$ is an increasing function of y. The theorem then follows immediately from (2.8). □

To put this another way, job duration is a declining function of type.

This result is an implication of on the job search. Whether on the job search is important in any particular labor market is an empirical issue. In fact, the work below, we use this result. Since the wage at which a worker is employed is informative about his or her type. It follows that in markets where duration doesn't vary with the wage, transitions are likely occurring as the result of exogenous termination. As we show, the wage offer distribution is readily identified in this case.

To end this section, we add an additional useful result.

Lemma 2.4. *Almost everywhere* $dp_w(\tilde{y}) = \frac{\omega'(\tilde{y})}{\omega(\tilde{y})}d\tilde{y}$

Proof. Using the fact that workers are indifferent about applying at all wages above their reservation wage, we have

$$we^{-\int_{y}^{\overline{y}}dp_w(\tilde{y})} = \omega(y).$$

Taking logs gives for every $w > \omega(y)$

$$\log w - \log\omega(y) = \int_{y}^{\omega^{-1}(w)} dp_w(\tilde{y}).$$

Writing the difference between the logs as the integral of the derivative

$$\int_{y}^{\omega^{-1}(w)} dp_w(\tilde{y}) = \int_{y}^{\omega^{-1}(w)}\frac{\omega'(\tilde{y})}{\omega(\tilde{y})}d\tilde{y}$$

from which the result follows. □

For those who are interested, we explain in the appendix how firms wages can be modelled. As a part of this, we explain why it is without loss to assume that workers types are uniformly distributed. The gist of the argument is that any observed behavior can be rationalized for any distribution of worker types by modify the profit function.

Employment Histories

For the rest of the paper, we'll focus on workers who move between firms with an intervening unemployment spell. The formulas can be adapted for workers who transit directly from one job to another because of on the job search. However, the formulas are more complex and do not change the basic logic.

The logic developed above is simply that worker types are unobservable, but outcomes provide some information about type. In particular the wage at which a worker is currently employed will say something about the worker's type provided type actually matters to firms. If types matter, high type workers will be more likely to get jobs with high type firms. The equilibrium conditions allow us to derive the distribution of search outcomes for workers who leave jobs at different wages.

Our particular interest is to establish conditions on the wage offer distribution that determine the shape of the relationship between the workers current wage and his expected future wage, as well as conditions under which the variance of the worker's future wage will be declining with the wage at which he or she is currently employed.

The model has other implications about transitions. Of course, workers who are employed at high wage firms are less likely to apply to and be hired with firms making higher offers, no matter what their type. So duration of employment will be longer at high wage firms.

At any given wage, the highest type workers who are employed at that wage are more likely to apply to and receive offers from higher wage firms. So high type workers at any wage will move more frequently than low type workers. This effect should be reflected in wages after a job transition. In particular, the longer the duration of a worker's employment with a firm, the lower his wage after transition is likely to be. One reason is that a low type worker is more likely to suffer a wage cut after an exogenous termination. The other is simply that the lower type worker is just less likely to be hired at the higher wage firms.

Relationship between current wage and type.

The core argument used above is that the wage at which a worker is currently employed is positively (but not perfectly positively) correlated with his type. To see why, note that by Theorem 2.1, workers make applications to every firm whose wage is above their reservation wage with, so to speak, equal probability. This means that when a worker moves from one job to another (in other words, conditional on moving), the wage of a worker of type y moves to is a random variable. Our first task is to compute this distribution.

The calculation is slightly different for workers who have unemployment spells than it is for those who make direct transitions on the job. So we'll start with unemployed workers.

According to Theorem 2.1, a worker of type y could end up being hired at a lot of different wages, since the support of his equilibrium mixed application strategy includes all wages above his reservation wage $\omega(y)$. Since the worker applies to a firm offering wage w with density $\frac{dG(w)}{G(\overline{w})-G(\omega(y))}$ and is then hired with probability $\frac{\omega(y)}{w}$ (because the expected payoff at each wage must be the same as the payoff he gets from applying at his reservation wage and being hired for sure), the probability density with which the worker is hired at a wage w (his outcome distribution) when he eventually leaves unemployment is given by

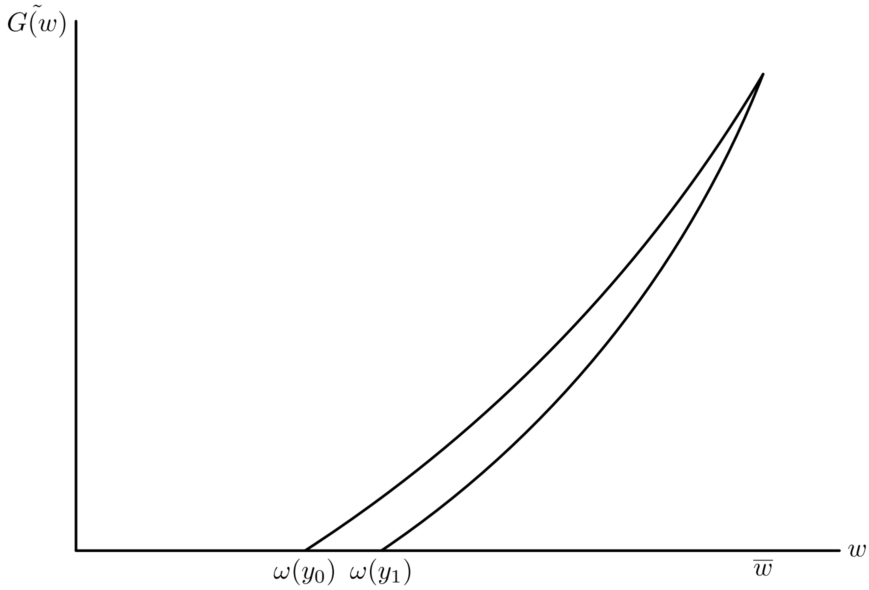
$$\sum_{t=0}^{\infty} \frac{dG(w)}{G(\overline{w}) - G(\omega(y))} \frac{\omega(y)}{w} \left[\int_{\omega(y)}^{\overline{w}} \left(1 - \frac{\omega(y)}{\tilde{w}}\right) \frac{dG(\tilde{w})}{G(\overline{w}) - G(\omega(y))} \right]^t =$$

$$\frac{\frac{dG(w)}{G(\overline{w})-G(\omega(y))} \frac{\omega(y)}{w}}{\int_{\omega(y)}^{\overline{w}} \frac{\omega(y)}{\tilde{w}} \frac{dG(\tilde{w})}{G(\overline{w})-G(\omega(y))}} =$$

$$\frac{dG(w)\frac{1}{w}}{\int_{\omega(y)}^{\overline{w}} \frac{1}{\tilde{w}} dG(\tilde{w})}.$$

Notice that the numerator in this term is independent of the workers type, y. Since ω is a strictly increasing function by Theorem 2.1, this density is higher at every wage the higher the worker's type. Since the cumulative distribution function is monotonically increasing and reaches a value 1 at $\overline{w}$, the distribution function for a higher type worker must first order stochastically dominate that of a lower type worker.

Then if we compare the probability distributions over future wages for two workers of types $y_1 < y_2$, they look like the distributions given in the following figure.



It is straightforward to calculate the mean wage received by a worker of type y when he or she eventually moves to a new job. It is given by

$$\frac{\int_{\omega(y)}^{\overline{w}} w \cdot \frac{1}{w} dG(w)}{\int_{\omega(y)}^{\overline{w}} \frac{1}{\tilde{w}} dG(\tilde{w})} = \frac{G(\overline{w}) - G(\omega(y))}{\int_{\omega(y)}^{\overline{w}} \frac{1}{\tilde{w}} dG(\tilde{w})} \tag{2.10}$$

while the variance of this distribution of future wages is

$$\frac{\int_{\omega(y)}^{\overline{w}} w^2 \cdot \frac{1}{w} dG(w)}{\int_{\omega(y)}^{\overline{w}} \frac{1}{\tilde{w}} dG(\tilde{w})} - \left[\frac{G(\overline{w}) - G(\omega(y))}{\int_{\omega(y)}^{\overline{w}} \frac{1}{\tilde{w}} dG(\tilde{w})}\right]^2. \tag{2.11}$$

The advantage of these two formulas is that they give a simple relationship between the wage offer distribution and unemployed workers' experience when they return to the workforce. Of course, we don't observe the type directly, but do get some information about it from the wage at which the worker was previously employed.

The next step then, is to find the distribution of types $\psi(y|w)$ that are hired at each different wage.

Lemma 2.5. *The conditional distribution function $\psi(y|w)$ satisfies*

$$\psi(y|w) =$$

$$\frac{\int_{\underline{y}}^{y} \omega'(\tilde{y})\, d\tilde{y}}{\int_{\underline{y}}^{\omega^{-1}(w)} \omega'(\tilde{y})\, d\tilde{y}} = \frac{\omega(y) - \underline{w}}{w - \underline{w}}.$$

if $y \leq \omega^{-1}(w)$, and is equal to 1 otherwise.

Proof. The probability with which a firm offering a wage w hires a worker of type y can be derived using the help of Theorem 2.1. When a firm has vacancy, the probability that it immediately hires a worker whose type is y or less is given by

$$\int_{\underline{y}}^{y} e^{-\int_{y}^{\overline{y}} dp_w(\tilde{y})} dp_w(y)$$

when y is less than or equal to $\omega^{-1}(w)$, and is 1 otherwise. Using Theorem 2.1 and the fact that the probability with which a worker of type y gets a job at wage w is $\frac{\omega(y)}{w}$, this can be written as

$$\int_{\underline{y}}^{y} \frac{\omega(\tilde{y})}{w} \frac{\omega'(\tilde{y})}{\omega(\tilde{y})} d\tilde{y} = \int_{\underline{y}}^{y} \frac{\omega'(\tilde{y})}{w} d\tilde{y}.$$

The probability with which the firm fails to hire initially is

$$1 - \int_{\underline{y}}^{\omega^{-1}(y)} \frac{\omega'(\tilde{y})}{w} d\tilde{y} = 1 - \frac{w - \omega\left(\underline{y}\right)}{w} = \frac{\omega\left(\underline{y}\right)}{w}.$$

Of course, if the firm fails to hire a worker on its initial try, it will continue to try in future. Using the steady state reasoning given above, the probability that a worker who is working at wage w has type less than or equal to y is then

$$\sum_{t=0}^{\infty} \int_{\underline{y}}^{y} \frac{\omega'(\tilde{y})}{w} d\tilde{y} \left[\frac{\omega\left(\underline{y}\right)}{w}\right]^t =$$

$$\frac{\int_{\underline{y}}^{y} \frac{\omega'(\tilde{y})}{w} d\tilde{y}}{\int_{\underline{y}}^{\omega^{-1}(w)} \frac{\omega'(\tilde{y})}{w} d\tilde{y}}$$

if $y \leq \omega^{-1}(w)$, and is equal to 1 otherwise. $\square$

Since the wage appears as a constant in both the numerator and denominator, they can be canceled. Giving the first result - if $w_1 > w_0$, then the probability distribution over types employed at w_1 first order stochastically dominates the corresponding distribution at w_0.

However, the main purpose of this result is to compute the expected future wage of an unemployed worker whose previous employer paid a wage w_1. This is given by

$$\int_{\underline{y}}^{\omega^{-1}(w_1)} \frac{G(\overline{w}) - G(\omega(y))}{\int_{\omega(y)}^{\overline{w}} \frac{1}{\tilde{w}} dG(\tilde{w})} \frac{\omega'(y)\, dy}{\int_{\underline{y}}^{\omega^{-1}(w_1)} \omega'(\tilde{y})\, d\tilde{y}} =$$

$$\frac{1}{\left(w_1 - \omega\left(\underline{y}\right)\right)} \int_{\underline{y}}^{\omega^{-1}(w_1)} \left(\frac{G(\overline{w}) - G(\omega(\tilde{y}))\, \omega'(\tilde{y})}{\int_{\omega(\tilde{y})}^{\overline{w}} \frac{1}{\tilde{w}} dG(\tilde{w})}\right) d\tilde{y}.$$

A change of variable in the integration gives the following:

Theorem 2.6. *Let $\phi(w_1)$ be the expected future wage of an unemployed worker who was previously employed at wage w_1. Then,*

$$\phi(w_1) = \frac{1}{(w_1 - \underline{w})} \int_{\underline{w}}^{w_1} \left(\frac{G(\overline{w}) - G(w)}{\int_w^{\overline{w}} \frac{1}{\tilde{w}} dG(\tilde{w})} \right) dw. \tag{2.12}$$

This is the main theorem in the paper. It says that the expected future wage can be calculated from knowledge of the wage offer distribution alone. Apart from the fact that the expected future wage is a monotonically increasing function of the current wage w_1, this formula suggests that the relationship between current and future wage is highly non-linear.

The function given by (2.12) can be estimated directly from the accepted wage distribution as we explain below. Conceptually, one way to check this is to estimate the relationship between the wage a worker received at his last job, and his current wage. For example, in the dataset considered below, if we estimate the regression

$$w_2 = \alpha_0 + \alpha_1 w_1 + \alpha_2 w_1^2 + \epsilon, \tag{2.13}$$

where w_1 and w_2 are the previous wage and the current wage, respectively. We can tell several different things from such a regression. For example, whether future wage and current wage is correlated and how strong is such a correlation. Moreover, α_2 should be telling us whether the relation between wages is non-linear. One could try to compare this regression with the estimated relationship given by (2.12). Of course, (2.12) suggests a highly non-linear relationship, so ideally the relationship between w_2 and w_1 should be estimated non-parametrically, then compared with (2.12). As a matter of fact, this roughly describes what we attempt to achieve in the empirical section in this paper.

Tedious calculations provide some special cases. When the wage offer distribution is uniform (on $[0,1]$) the relationship between current and future wage is slightly concave. When the wage offer distribution has cdf x^2 on $[0,1]$, the relationship is linear, while the wage offer distribution x^3 gives a slightly convex relationship. If $G(x) = x^2$, for example, (2.12) reduces to

$$\frac{1}{w} \int_0^w \frac{1 - \tilde{w}^2}{2(1 - \tilde{w})} d\tilde{w} = \frac{1}{2w} \int_0^w (1 + \tilde{w})\, d\tilde{w} = \frac{1}{2} + \frac{1}{2} w.$$

A similar analysis can be applied to the variance. The variance of the worker's future wage, when his or her previous wage was w_1 is (following the argument associated with the theorem (2.6))

$$\frac{1}{(w_1-\underline{w})}\frac{\int_{w_1}^{\overline{w}} wdG(w)}{\int_{w_1}^{\overline{w}}\frac{1}{\tilde{w}}dG(\tilde{w})}-\left[\frac{1}{(w_1-\underline{w})}\frac{G(\overline{w})-G(w_1)}{\int_{w_1}^{\overline{w}}\frac{1}{\tilde{w}}dG(\tilde{w})}\right]^2 \tag{2.14}$$

Again this relationship suggests a fairly complicated relationship between wage and variance. Heuristically, regressing w on w_1 will lead to a relationship that exhibits a lot of heteroskedasticity.

The formulas given above require information about the wage offer distribution G, while available data only provides information on the accepted wage distribution E. The following theorem relates these two distributions.

Theorem 2.7. *The wage offer distribution is the solution to*

$$G(w)=\int_{\underline{w}}^{w}\left(\gamma+(1-\gamma)\int_{\tilde{w}}^{\overline{w}}\frac{dG(w')}{w'}\right)\frac{\tilde{w}}{(\tilde{w}-\underline{w})}dE(\tilde{w}). \tag{2.15}$$

The proof is in appendix. For a direct use of this equation in a general context, it is tempting to examine whether the contract mapping theorem may apply so that one can recall the fixed-point theorem to point-wise identify and therefore to estimate $G(w)$. This amounts to verifying several properties of such a structural equation. Instead, we would rather undertake an easier path by considering a special case where the distribution $E(w)$ is differentiable. This is found to be innocent assumption in our data after all.

We differentiate (2.15) to get

$$g(w)=\left(\gamma+(1-\gamma)\int_{w}^{\overline{w}}\frac{dG(w')}{w'}\right)\frac{w}{w-\underline{w}}e(w). \tag{2.16}$$

Let

$$h(w)\equiv\int_{w}^{\overline{w}}\frac{dG(w')}{w'}. \tag{2.17}$$

Now substituting (2.17) into (2.16) gives

$$-h'(w)=\frac{(\gamma+(1-\gamma)h(w))}{w-\underline{w}}e(w). \tag{2.18}$$

The above equation suggests a way to find wage offer distribution. We can first solve the differential equation (2.18), then integrating the solution $h'(w)$. This is the method we follow in our empirical exercise and we defer the detailed discussion on estimation strategy.

2.4 Comparison with other models

This prediction suggests a way to compare three different models of directed search. Each of these three models can be thought of as special cases of the model discussed here.

For example, one special case of the model above occurs when the type of an employee determines whether or not he is hired just as in the model above, but where type is not retained from one period to another. For example, each employees type could be redrawn each period by selecting randomly from the distribution F. This is nothing more than a restatement of a standard model of directed search in which workers use mixed application strategies.[4] In particular, since type is not persistent, this means the wage at which a worker is currently employed should have no relationship at all with the wage a worker gets when he or she moves on to a new job.

Conversely, worker types might again determine the probability of being hired, but these types may be public observable to other workers. If everyone knows who the highest type worker is, they will also know where he or she will apply given any distribution of wages. As a result, mixing will break down, and workers will match assortatively, as in special cases discussed by Shi [2001] or Eeckhout and Kircher [2010]. Wages received by workers as they move between jobs will be very highly correlated in this case.

The following picture may help make the results in the empirical section, and the connections between the various search models clearer.

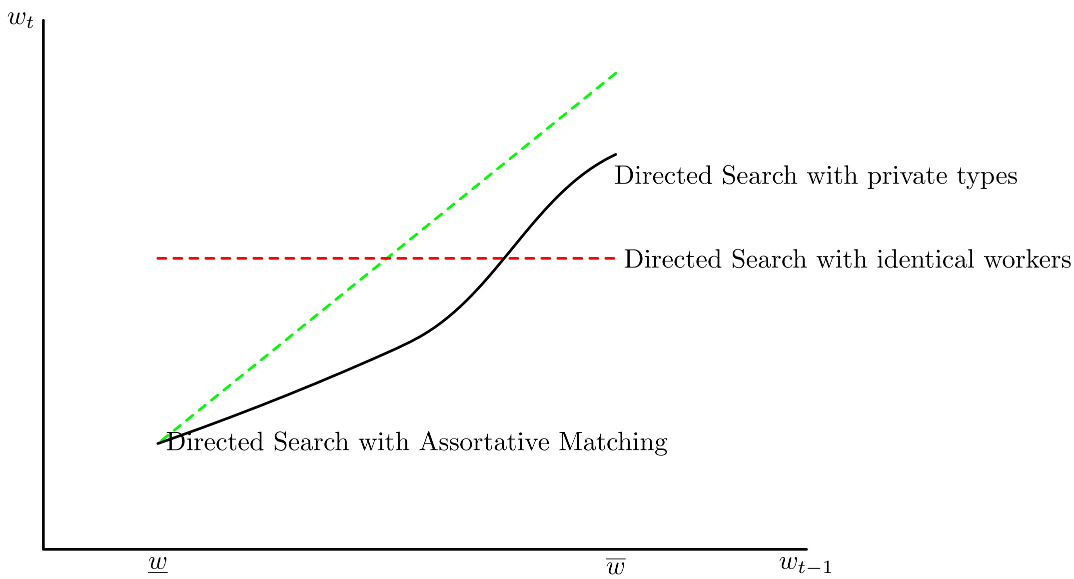


FIGURE 2.1: Comparison of Models

[4] For example Peters [2000].

The horizontal axis in the picture represents the current wage of a worker, while the vertical axis represents the wage in the job that the worker moves to after a match is terminated. All the predictive consequences of the model we consider here emerge from studying the predictive content of current wage. The most basic directed search model in which workers are all identical, but firms offer different wages is represented by the dashed red line in the picture above. It is flat because a lucky draw in one period will not last - when a worker moves to a new job, he or she will receive the same wage on average no matter what their current wage happens to be.

On the other hand, pure assortative matching leads to the relationship indicated by the green line in the picture - whatever wage a worker gets today, he will also get tomorrow, simply because assortative matching will continue to place him or her at the same point in the wage distribution.

Finally, the model described above will lead to a correlation between current and future wage, but this correlation will be much weaker than that predicted by assortative matching. The reason for the weaker correlation is that the wage is a very imperfect signal of worker type because of the mixed application strategies that workers use. On the other hand, higher type workers are more likely to be employed at higher wage firms.

In general, the model says the relationship between past and future wage is non-linear. The shape of the relationship depends on the shape of the wage offer distribution, as we described above.

2.5 Empirical Application

2.5.1 Data

In order to examine transitions in applications, we use a data set on the French labor market. In this section, we'll first describe the data, illustrate the estimation methodology, and report all the empirical results.

Our main data source is the DADS (Declarations Annuelles de Donnees Sociales), a large scale administrative database of matched employer-employee information collected by INSEE (the French National Institute of Statistics and Economic Studies). The data are collected in accordance with the mandatory employer reports of the employment status and gross earnings, for the payroll tax purposes.[5]

[5] Same data source and similar data construction have been used in literature. Refer to Abowd et al (1999), Postel-Vinay and Robin (2002), Cahuc et al (2006) for more detailed elaboration of data source.

Each observation in the DADS file corresponds to a unique individual-establishment-year match. The observation includes an identifier for each employee, an identifier for each establishment, and an identifier for the parent enterprise of the establishment. For each data entry, we observe the beginning and ending days within the calendar year the employee worked in the establishment, together with other match specific information such as the working hours and job nature (whether it is a permanent contract or temprory contract). Furthermore, we observe the employee's gender, age, occupation, total income (both before and after tax), as well as the location and industry of the employing establishment.

Such a database is ideal for our purpose, as one can easily trace workers' movements between jobs. Another advantage of this data is that the employment information is at the plant (establishment) level. Indeed, we identify different labor markets using plant level information. Plants within a firm are assumed to be independent in terms of technological diffusion and production choices.

The data from DADS are reported in two formats. One is cross-sectional (DADS postes). This means we have the census data of employment for each given year. However, each individual, worker or firm, would be assigned different identifiers from different years. Therefore, it makes infeasible to trace a worker's employment history over the time horizon more than a year. The other format DADS provides is of panel nature (DADS panel). It reports the employment histories of 1978-2010 for the workers who were born in Octobers of even years (for example, 1978, 1982, 1990, etc). Since we consider only permanent jobs, we use the panel data of 2005-2010 from DADS for our empirical analysis.[6] We also use the cross sectional data of year 2007 for referencing, when we try to understand our empirical results.

Worker Transitions

We start our data work by looking for worker transitions in the sample of panel data. As the case considered in theory section, we focus on workers who move between jobs with an intervening unemployment spell. For each matched worker we trace her employment history by looking for her earlier and later matchings (i.e., the same worker identifier but different firm identifiers). We regard a pair of jobs to involve a transition if the end of previous job and the start of new employment are at least one month apart.[7]

Using this approach, each employee's work history can be reduced to a series of transitions from one job to another during the period 2005 to 2010. The wage variable we

[6] The job nature of permanency is only provided by DADS from 2005.

[7] We experiment with other length of unemployment for robustness check, for example, 21 days, or 14 days. Our results qualitatively remain to hold.

shall use in the analysis is the hourly wage rate, which is defined by total wage bill received by the worker divided by the effective working hours.[8]

Defining Labor Markets

In the theory, a labor market consists of a set of workers who are all equally qualified for a certain set of jobs. Our prior belief concerning the French labour market is that various segments of the entire labour force may entail different properties and natures of market interaction. It is likely that skill, occupation, industry, and even geographical characteristics may all contribute to forming local or regional markets. These in turn should imply different information structures for search strategy and matching outcome to work in practice.

Rather than breaking the data into markets in an ad hoc way, we employ a flexible and data-driven approach to identify markets. We borrow a commonly used method in statistics and computer science for community detection. In particular, we develop a measure that combines two dimensions of variation, (4-digit) occupational codes and geographic units. The geographic unit is referred to “commune”, which means civic township in French data. A market is basically defined by mobility - set of jobs and workers is defined to be a market when there is evidence that a majority of workers are mobile between the various positions offered in these markets.

In other words, we exploit transition data and objectively find a set of communities which should reflect the connection between cells of occupation-geographic unit in data. Evidence of sufficient mobility then suggests the pair of these cells (occupation-geography) share certain common properties. In turn, we regard these homogeneous groups as local labour markets in our context. More technical notes are provided in Appendix.

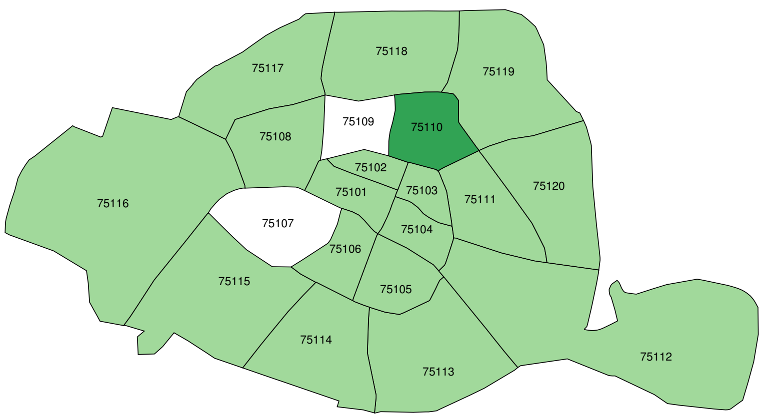


FIGURE 2.2: Examples of Local Labor Markets

[8]The nominal hourly wage rate is normalized by the dollar value of 2005.

The figure illustrates two examples of identified local labor markets. One is that the market covers same commune but different occupations at 75110 (zip code). These occupations are under different category code of jobs – 633c: Electriciens, électroniciens qualifiés en maintenance entretien, réparation : automobile; and 633d: Electriciens, électroniciens qualifiés en maintenance, entretien : équipements non industriels. The other market on the figure covers same occupation but different communes at 75107 and 75109. The occupation in this case is 627d: Ouvriers qualifiés de scierie, de la menuiserie industrielle et de l'ameublement.

To proceed with empirical analysis, we further refine the data requirement and derive a working sample. First, we keep only the transitions involving move between jobs in the same local labor markets in our empirical analysis. This requirement effectively drops the transitions across different markets. Those transitions might occur from other reasons, other than the matching mechanism we study here in this paper.

Next, we aggregate worker transitions at 2-digit occupational codes level to illustrate the empirical relevance of the model. Such aggregation suppresses the heterogeneity at micro-level occupations. To put it differently, it assumes all occupations under same 2-digit category share sufficient common features in data generating process. For what we are trying to achieve, we consider it as a reasonable simplification. Table 2.1 lists the definition for these 2 digits occupations. We concern mostly whether skillfulness inherited in professions is associated with any subtle difference in properties of job matching outcomes.

TABLE 2.1: Occupation List at Two-digit Level

Occupation	Definition
PCS:62	Skilled industrial type worker
PCS:63	Skilled artisanal worker
PCS:64	Drivers
PCS:65	Skilled handling, storage and transport workers
PCS:67	Unskilled industrial type worker
PCS:68	Unskilled artisanal worker

We now obtain a working sample for our empirical analysis that follows. Table 2.2 provides the basic summary statistics for the sample. There we reported by occupation, both the mean and standard deviation of old and new wages, and unemployment spells.

2.5.2 Estimation Methodology

For each of labour markets $l \in \{1, 2, ...L\}$, our working sample consists of a series of paired wages $\{(w_1, w_2)_i^l\}$ where w_1 is the wage of previous job, w_2 is the wage at new job,

TABLE 2.2: Summary Statistics for Transition Sample

Occupation	Obs	w_{t-1}		w_t		$dur_{t-1,t}$	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
62	10,321	10.577	3.271	11.807	4.402	345.176	217.967
63	12,554	9.992	2.796	10.977	3.859	367.024	256.275
64	7,100	9.351	2.247	10.111	3.063	333.990	225.628
65	2,357	10.247	3.257	11.490	4.405	347.684	225.622
67	5,744	8.795	2.276	9.851	3.187	341.330	225.526
68	11,474	8.088	1.832	8.744	2.459	335.152	239.744

and i is the index for individual worker in the market l. Our fundamental assumption of the empirical exercise is the observed market data are generated by the directed-search model in this paper.

The equilibrium structure of our search model implies the relations of these wages should hold by (2.12) and (2.15). In this setup, econometrians observe past wage w_1, future wage w_2, and from which, the accepted wage distribution E. The structural elements to be estimated are wage offer distribution G and the exogenous rate of separation γ. Our primary interest is to examine the validity of assumed data generating process in various labour markets. To put more simply, whether the variation of data supports the proposed model in this paper.

We take a fully non-parametric approach and proceed with following three steps. First, for a fixed value of γ, we estimate the wage offer distribution $\hat{G}$, with which we can estimate the expected future wage for any given current wage, $\hat{\phi}(w_1)$. In the end, we estimate γ so that our choice of data fitting criteria is optimized. We next illustrate these steps in details.

Step 1. Estimating Wage Offer Distribution G.

In theory section, we derived a differential equation (2.18) as follows, which can solve for $h(w)$ in principle.

$$h'(w) = -\left[\gamma + (1-\gamma)h(w)\right]\frac{e(w)}{w - \underline{w}}$$

We show in appendix that $h(w)$ can be solved analytically. The solution can be written as

$$h(w) = \frac{\gamma}{1-\gamma}\left[\exp\left((1-\gamma)\int_w^{\overline{w}} \frac{e(\tilde{w})}{(\tilde{w} - \underline{w})}d\tilde{w}\right) - 1\right] \tag{2.19}$$

We use standard kernel estimation for $e(w)$, the density of accepted wage distribution. In particular, we choose gaussian kernal function, and follow Silverman's rule of thumb bandwidth (ie, $1.06 * n^{-1/5} * std(w)$). For this step we assume that we know γ for now. Therefore, we get an estimate of $\hat{h}(w)$ according to (2.19).

Recall $h(w) \equiv \int_w^{\overline{w}} \frac{dG(w')}{w'}$. This implies that

$$g(w) = -wh'(w) \tag{2.20}$$

Since we know that

$$h'(w) = -(\gamma + (1-\gamma)h(w))\frac{e(w)}{w - \underline{w}}$$

our estimates of $\hat{e}$ and $\hat{h}$ allow us to get an estimate of $h'(w)$ and therefore $g(w)$ according to (2.20). And in turn we can estimate

$$\hat{G}(w) = \int_{\underline{w}}^{w} g(\tilde{w})d\tilde{w}$$

Step 2. Computing Expected Future Wage ϕ.

Recall from (2.12) the expected future wage is

$$\begin{aligned}\phi(w_1) = \mathbb{E}[w_2|w_1] &= \frac{1}{w_1 - \underline{w}}\int_{\underline{w}}^{w_1} \frac{G(\overline{w}) - G(w)}{\int_{w}^{\overline{w}} \frac{dG(\tilde{w})}{\tilde{w}}}dw \\ &= \frac{1}{w_1 - \underline{w}}\int_{\underline{w}}^{w_1} \frac{1 - G(w)}{h(w)}dw\end{aligned}$$

where the second line holds by the definition of $h(w)$ and the fact that $G(\overline{w}) = 1$. Then, for any w_1, we substitute estimates of $\hat{h}(w_1)$ and $\hat{G}(w_1)$ for an estimate of the expected future wage $\hat{\phi}(w_1)$.

Step 3. Estimating the exogenous rate of separation γ.

As we do observe all the transitions, the realized value of future wage is then indeed observable to econometricians. We computed expected future wage as a function of current wage. Since our previous steps used kernel estimation technique and our estimation may suffer from boundary bias. We first trim off observations which involve current wages below 5% and above 95%. In other words, we only keep current wages between 5-th and 95-th percentiles. We then evenly divide remaining current wages into T bins.

For each bin, we take the middle point as the value of current wage, and then compute its expected future wage in the sample. Consequently, we obtain a sample of T observations. Each observation t corresponds to a value of current wage (w_1^t) and its mean future wage in sample ($\tilde{\phi}(w_1^t)$).

We then use GMM estimation to look for an estimate of γ between 0 and 1. We choose moment conditions as $\mathbb{E}(u_t) = 0$ and $\mathbb{E}(w_1^t \cdot u_t) = 0$, where $u_t = \tilde{\phi}(w_1^t) - \hat{\phi}(w_1^t)$.

A few remarks are in order. First, conducting inference for the estimate γ is non-standard, as the objective function involves $\hat{\phi}$, a complicated estimate from previous steps. The computation itself can be burdensome. We therefore follow a method proposed by Armstrong et al (2014), which provides a feasible but fast algorithm for situation like ours. For balance between speed and accuracy, we choose 500 repetition for bootstrap procedure.

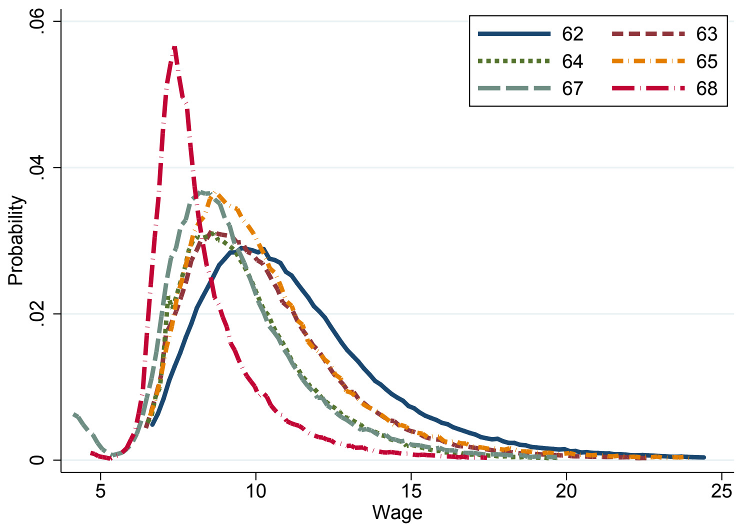
Second, the objective of estimating γ can be modified by including more moment conditions to improve estimation efficiency. For example, one can include variance of future wage as in (2.14). It however can introduce more estimation bias when computing the predicted variance of future wage by the model. Subsequently, it should make our task of computing standard error more involved and less tractable.

Lastly, we implement the estimation procedure by choosing $T = 90$. The estimation results of $\hat{\gamma}$ is listed in the table 2.3. On a robustness check, we also repeat the step with $T = 250$, and find almost identical estimates. We therefore decide to choose this set of outcome to report in the paper.

2.5.3 Empirical Results

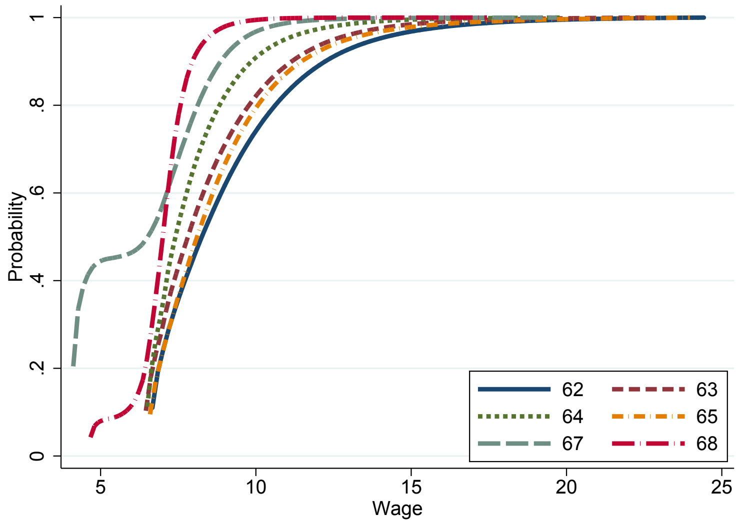
The density functions of accepted wage distributions are plotted in Figure 2.3. They mostly seem to follow log-normal distributions. Occupation of unskilled artisanal workers is the most concentrated at lower average wage. The occupations with more skillfulness feature appear with larger variance and longer tail.

FIGURE 2.3: Densities of Accepted Wage Distribution



We next show our estimates of cumulative distribution functions of wage offer distributions. Such an estimate can only be done for a given value of γ. We therefore plot our estimates in Figure 2.4 at estimated $\hat{\gamma}$. The pattern of spread across occupations is carried over to these wage offer distributions. That is, the occupation of artisanal workers appears to have most concentrated distribution, while the wages for skilled industrial workers are most spread.

FIGURE 2.4: Wage Offer Distribution by Occupation



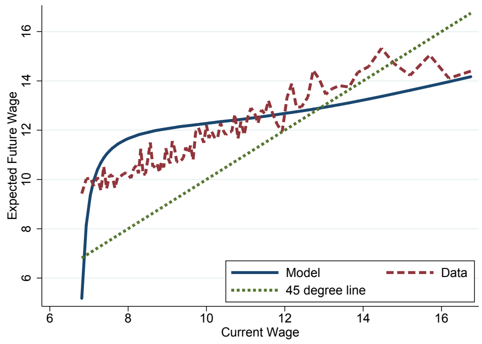
Our GMM estimation results are reported in table 2.3. The first column lists the estimate of γ for each occupation. For all skilled occupations (62-65), we got interior point at our optimal estimates. These rates of exogenous separation vary between 3.5% and 6.5%. The labour market of skilled workers for transport, storage etc (PCS 65) has a much larger rate of separation, 34.3%. For both unskilled labour markets (67 & 68), we could not find an estimate between zero and one. This suggests that there does not exist a reasonable separation rate to justify the model. We therefore take this estimate outcome as an indication of model misspecification.

TABLE 2.3: Estimation Results for γ

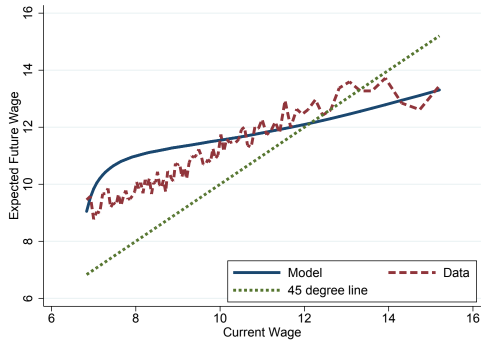
PCS	γ	Std	Obj	J test	P-value
62	0.065	0.010	0.127	11.558	0.003
63	0.085	0.008	0.136	12.345	0.002
64	0.035	0.002	0.499	45.447	0.000
65	0.343	0.037	0.013	1.207	0.547
67	0.001	0.001	0.325	29.556	0.000
68	0.000	0.000	33.318	3031.943	0.000

Our estimation results are plotted in figure 2.5. There each sub-figure represents an occupation. The blue lines are our fitted model, red lines are data plots, and green dotted ones are 45-degree lines. It is obvious that, for occupation 67 and 68, imposing

FIGURE 2.5: Expected future wage: Data and Model

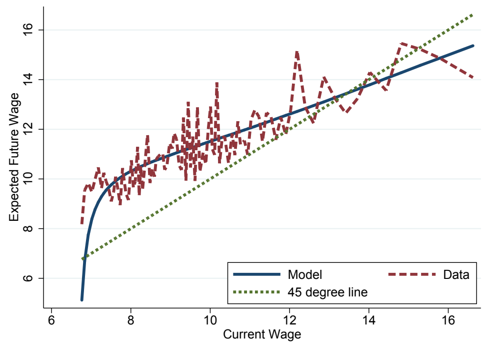


(a) PCS:62

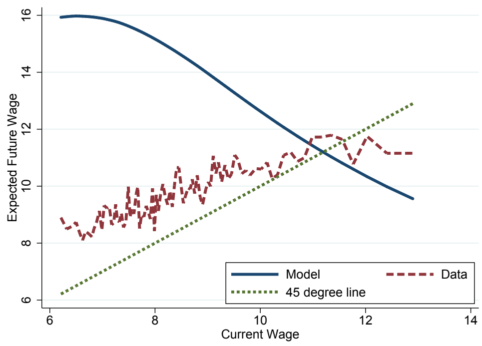


(b) PCS:63

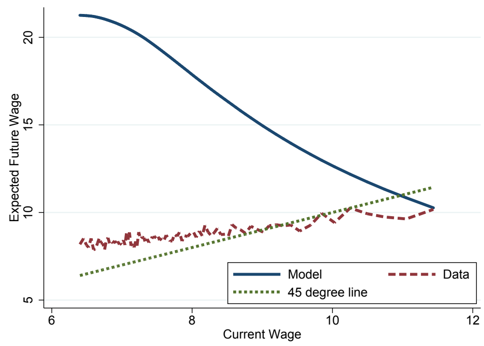
(c) PCS:64



(d) PCS:65



(e) PCS:67



(f) PCS:68

our model structure would not allow one to find a reasonable estimate approximating the data variation. Again, we take this as indication of model mis-specification. For other occupations with skillfulness features, our model seems to generate good approximation to data patterns.

To cross check our estimates on exogenous separation rate, we resort to the cross-sectional data set of DADS. We computed rate of separation for each month and each occupation we considered. They are reported in table 2.4. We use the employment data for the year of 2007 in DADS. What we were hoping to see is whether they would show any specific patterns that coincide with our estimation outcome. For example, whether at all occupation 65 exhibits much larger rate of separation than any other markets. Or, incidentally, whether markets for unskilled workers may show any significant difference than the skilled ones. If so, our estimates on γ may not purely reflect how our model structures restrict data fit. As the table 2.4 shows, there is no clear evidence of how these markets should differ in separation rates.[9]

Table 2.4: Seperation Rate from Cross-sectional Data

PCS	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Overall
62	0.016	0.017	0.028	0.020	0.021	0.037	0.019	0.022	0.029	0.022	0.023
	(0.002)	(0.003)	(0.003)	(0.001)	(0.003)	(0.004)	(0.001)	(0.002)	(0.002)	(0.001)	(0.007)
63	0.028	0.023	0.040	0.026	0.028	0.042	0.028	0.036	0.042	0.030	0.032
	(0.010)	(0.005)	(0.003)	(0.002)	(0.001)	(0.006)	(0.002)	(0.002)	(0.004)	(0.001)	(0.008)
64	0.019	0.021	0.029	0.020	0.023	0.039	0.019	0.027	0.032	0.023	0.025
	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.009)	(0.001)	(0.003)	(0.007)	(0.002)	(0.007)
65	0.024	0.022	0.034	0.025	0.028	0.045	0.023	0.032	0.033	0.027	0.029
	(0.007)	(0.005)	(0.008)	(0.006)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.008)
67	0.016	0.017	0.030	0.019	0.020	0.032	0.022	0.027	0.028	0.022	0.023
	(0.003)	(0.005)	(0.009)	(0.001)	(0.003)	(0.001)	(0.001)	(0.003)	(0.001)	(0.003)	(0.006)
68	0.018	0.016	0.024	0.017	0.019	0.021	0.013	0.022	0.021	0.017	0.019
	(0.005)	(0.005)	(0.003)	(0.003)	(0.004)	(0.004)	(0.001)	(0.002)	(0.003)	(0.002)	(0.004)

2.6 Model prediction revisited

Our empirical estimation so far has followed a structural approach. The estimation results allow us to rule out the labour markets with unskilled features complying with our directed search model. However, our theory model is much structured that delivers both rich predictions and welfare implications. We look into this issue in this section.

2.6.1 Past wage, current wage, and duration

The theory suggests a number of empirical relationships. The main idea is that wage is a signal of worker type and that the information contained in this signal should be reflected in the relationship between wages across transitions. We'll focus on two following connections here because they are a bit unexpected.

1. Past wage is an imperfect predictor of current wages;
2. Duration is negatively correlated with current wage. i.e, the higher the wage that a worker gets when leaving the unemployment, the shorter the duration that she has been searching.

[9]The direct comparison of these magnitudes in table 2.4 with our estimates is not valid. Our estimation uses workers gone through transition and unemployment during transition. For computation of table 2.4, we consider all terminations of jobs without worrying about unemployment. Therefore, we were looking for patterns across occupations only when investigating table 2.4.

To try to capture this, we first regressed the wage a worker's wage after a transition ($w_{i,t}$) on the wage at his last position ($w_{i,t-1}$) as in the following equation:

$$w_{i,t} = \alpha_0 + \alpha_1 w_{i,t-1} + \alpha_2 w_{i,t-1}^2 + \gamma_1' X_{i,t} + \mu_{i,t} \tag{2.21}$$

Here, X is the vector of observed characteristics of the worker. The squared term appears here because of the possibility that the strength of the correlation between wage ranks across transitions can vary with the wage the worker had before the transition. We repeat the same regression without other control variables.

The results are reported in table 2.5. Again, we focus on occupations with skilled sets of workers only. (PCS: 62, 63, 64 and 65). Consider regressions without covariates. If we evaluate the correlation between past wages and current wages at mean level, the correlations vary between 0.5 and 0.8. (.611, .781, .578 and .610 respectively). This correlation is strong enough to suggest that type information is important to firms, but is hardly strong enough to indicate either assortative matching, or a simple wage ladder. The interaction term w_{it-1}^2suggests that the relationship is concave for most of occupations, suggesting that more information is conveyed about type in low wage jobs.

Table 2.5: Regression Results between current and past wage

PCS	62		63		64	
w_{it-1}	0.853***	1.246***	0.966***	1.281***	0.768***	1.307***
	(9.043)	(17.929)	(12.175)	(17.993)	(5.649)	(12.350)
w_{it-1}^2	-0.020***	-0.030***	-0.025***	-0.034***	-0.023***	-0.039***
	(-5.330)	(-11.080)	(-7.793)	(-11.226)	(-3.559)	(-7.908)
Constant	4.276***	2.332***	5.362***	1.818***	3.642***	1.533***
	(4.717)	(5.649)	(7.150)	(4.588)	(3.546)	(2.836)
Covariates	Yes	No	Yes	No	Yes	No
Observations	6,745	10,321	9,360	12,554	5,130	7,100
R-squared	0.244	0.124	0.214	0.117	0.252	0.111
PCS	65		67		68	
w_{it-1}	0.543**	0.835***	0.614***	1.069***	0.399***	0.687***
	(2.206)	(5.329)	(4.357)	(10.744)	(3.563)	(7.550)
w_{it-1}^2	-0.006	-0.011*	-0.016**	-0.030***	-0.007	-0.018***
	(-0.578)	(-1.765)	(-2.435)	(-6.282)	(-1.184)	(-3.839)
Constant	6.061***	4.234***	5.529***	2.940***	6.416***	4.420***
	(2.835)	(4.730)	(5.392)	(5.956)	(4.214)	(10.410)
Covariates	Yes	No	Yes	No	Yes	No
Observations	1,580	2,357	3,837	5,744	7,858	11,474
R-squared	0.433	0.159	0.231	0.098	0.159	0.058

We next examine the second prediction of our interest, duration effect. For this purpose, we regress duration of a worker stayed in unemployment spell and the exit wage. The results are reported in table 2.6. It is apparent that except occupation 65, all other markets indicated strong evidence of negative relation between duration of unemployment and wage. The parameter for occupation 65 came out statistically insignificant, which perhaps is caused by the relatively small sample size. Therefore, again, the duration effect fits the story what our theory tries to tell - there is persistent information of type that wages pass on across transitions.

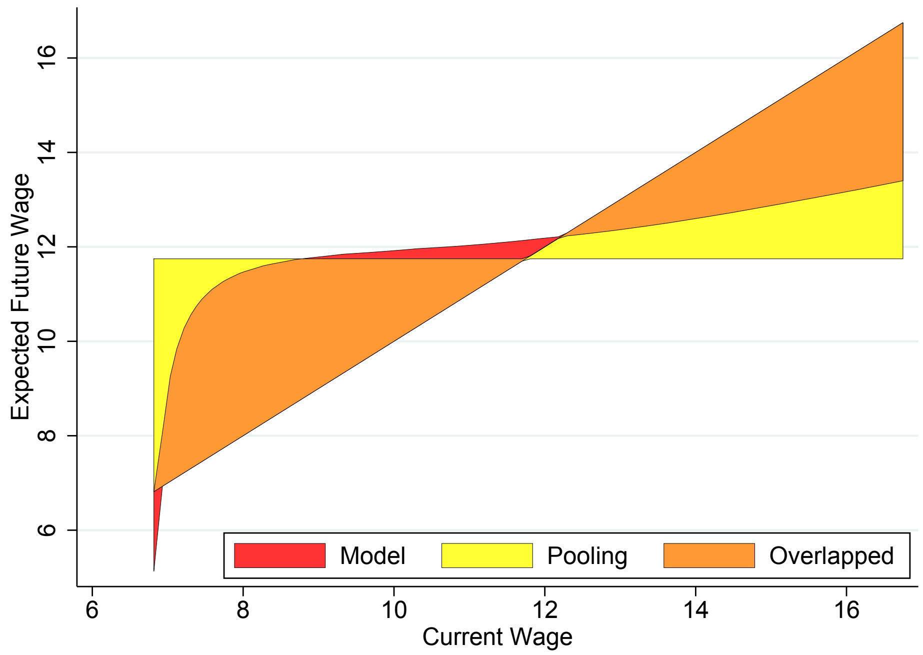
TABLE 2.6: Regression Results between duration and past wage

PCS	62		63		64	
w_{it-1}	-0.010***	-0.011***	-0.024***	-0.015***	-0.019***	-0.007*
	(-3.191)	(-5.385)	(-7.851)	(-6.028)	(-3.792)	(-1.939)
Constant	5.722***	5.770***	5.709***	5.797***	4.674***	5.654***
	(37.437)	(251.951)	(37.468)	(224.671)	(20.843)	(155.422)
Covariates	Yes	No	Yes	No	Yes	No
Observations	6,745	10,321	9,360	12,554	5,130	7,100
R-squared	0.138	0.003	0.132	0.003	0.104	0.001
PCS	65		67		68	
w_{it-1}	0.007	0.007*	-0.020***	-0.016***	-0.020***	-0.005
	(1.080)	(1.655)	(-2.951)	(-3.882)	(-4.111)	(-1.528)
Constant	5.430***	5.589***	5.347***	5.762***	5.581***	5.614***
	(16.254)	(124.641)	(24.858)	(156.433)	(24.391)	(185.091)
Covariates	Yes	No	Yes	No	Yes	No
Observations	1,580	2,357	3,837	5,744	7,858	11,474
R-squared	0.245	0.001	0.184	0.003	0.052	0.000

2.6.2 Welfare implication

Our structural approach allows us to make further implication on welfare. We recall one of the figures in estimation section to illustrate the idea. Consider the example as in figure 2.6.

FIGURE 2.6: Illustration: Welfare Implications



There are two directed search models we have compared in theory section. One is the assortative matching, and the other is random matching. Assortative matching as one extreme corresponds to the efficiency bound, which is 45-degree line in figure 2.6. The random matching outcome serves as another extreme where efficiency all lost. It is so as the channel that passes on the worker type information is completely shut down.

In figure 2.6, the horizontal line at mean level of future wage represents this model. Therefore, the complete loss of efficiency is the sum of yellow areas and orange areas.

Our directed search model stands in between of these extremes. We explicitly consider one particular type of loss incurred by search friction. The only channel induces the efficiency loss in our model arises from the mixing strategy of search in equilibrium. In figure 2.6, the loss of our model is areas of red and orange. We then compute the percentage of welfare loss due to equilibrium search strategy. Of course, if one is willing, this loss can also be interpreted as informational loss, due to incomplete knowledge of rivals behavior in market.

Table 2.7 reports the percentage of total loss on efficiency that can be explained by informational asymmetry as considered in our model. It varies between 45% to 100%. We read this result as evaluating significance of our directed search model. For some labour markets, the informational loss alone can outweigh most of other sources incurring efficiency loss.

Table 2.7: Welfare Implications

PCS	percentage of total loss
62	76.061
63	76.926
64	100.172
65	43.148

2.7 Conclusion

The results reported here are consistent with the theory of mismatch we have presented, but not consistent with either of the other three models of directed search. The model we use is about as simple as it could possibly be. Nonetheless it seems to get some of the basic empirical properties right.

On the theoretical side, there are at least three dimensions in which the model seems to be going too far. First, it assumes that workers' types don't change over the course of their life. This may be the least objectionable assumption. Workers will obviously acquire new skills as they age. Yet these skills are more often than not contractible. For example, a worker who acquires an MBA will probably be compensated for it. We don't interpret this as an improvement in the worker's type. In our regressions, we capture this by adding experience (measured by worker age) and assuming that this will explain much of the rise in income that workers experience as they move between jobs.

We also assume that match termination is independent of worker and firm type. Notice that this is different from the assumption that match termination doesn't depend on duration. It surely does. However, in our theory no one cares about duration per se, and wages represent expected income of the life of the match. The only important assumption is that matches are terminated in a way that maintains the distribution of worker and position types available on the market in any period.

Lastly, we do not allow firms to refuse to hire. If we did, high type firms would refuse to employ very low type workers making current wage a much better signal of worker quality. Pursuing this modification goes well beyond the scope of the current paper.

This brings us to one of the implications of the results here. The results are consistent with a model in which firms use uncontractible information to screen candidates when they hire. A natural empirical question this suggests is whether firms actually value this uncontractible information, or whether it is simply a way of coordinating search. These are independent questions. For example, it could well be that a firm hires a worker because the worker knows one of the bosses relatives. That isn't the same as saying that the firm is willing to pay more to hire someone who knows a relative, nor that the firm is more profitable when it hires someone who knows a relative instead of someone who doesn't. The model here seems to fit well enough to move to a structural approach which tries to estimate the distribution of firm types.

Chapter 3

Is Citation Behavior Biased? The Influence of Journal Editors

3.1 Introduction

Measures of scientific impact based on bibliometrics play an increasing role in the academic labor market: Tenure decisions and attribution of scientific grants are partly motivated by impact measures based on citations counts (see e.g. Ellison [2010] for an analysis of the use of the Hirsch index in economics; see also Starbuck [2005]). This is particularly true in economics and finance: Hamermesh et al. [1982], for instance, documents that citation counts is an important determinant of salary differences among academics in economics departments.

A "frictionless view" of scientific production would go as follows: Each paper has an intrinsic "fundamental value" that purely reflects its scientific contribution at time of submission and fully determines – up to random shocks, the probability that the paper gets published and the number of citations it will gather in the future. Being highly cited means that the paper embodies a piece of knowledge that is "useful for future science" (see e.g. Sikorav [1991] for a formalization of this view). Under that view, editors are disinterested experts who perform an objective evaluation of the "scientific contribution" of the papers they receive (which is identical to evaluating expected future impact).

We analyze the extent to which, by contrast with the frictionless view of the citation game, the propensity of researchers to quote a given paper can be biased by factors that are independent of the paper's scientific quality. When citing previous work, authors might not be simply listing the past scientific knowledge that has been useful to their own contribution, but might rather be referring to the most salient names or pandering to powerful people in the field, whose influence might matter for the fate of the paper in the editorial process.

Establishing the importance of social forces in the citation game requires finding a variable that affects citation levels without affecting a paper's scientific quality. Our paper looks at how citations of a researcher's paper are affected by her or one of her colleagues ceasing to be a journal editor. Because editors are in a key position in the publication process, it is likely that their work and their colleague's work might be cited with particular care. The end of the appointment of a colleague of the author(s), is a shock that can plausibly be assumed exogenous to the scientific quality of a paper: it's predictive power on citation rates provides a convincing example of how forces others than scientific value can affect the publication game. Our identification thus relies on a diff-in-diff approach: (1) we consider the end of an editorial appointment as a shock that is exogenous to the scientific quality of pre-existing papers (of the editor or his colleagues) (2) we compare citations in the journal where the author or her colleague is (or was) an editor to those where she is not an editor (this provides a control group).

We focus on the top three finance journals (Journal of Finance, Journal of Financial Economics, Review of Financial Studies), for which we collect publication and citation data using *Web of Science*. One advantage is that this is a subfield of economics where the turnout time is relatively low, which implies that publication time and production time are not too far away and that the time-period during which an editor can influence the journal is well approximated by his tenure time at the journal.

Our analysis proceeds in two main steps. First, using data on editorial appointments at the journal level, combined with citation data for all scientific authors, we establish a positive jump in citations when a scientist becomes an editor and a negative jump, of similar magnitude, when she steps down. Such jump does not occur in the journals where the person is not an editor. The effect is significant and large: When becoming an editor in a journal, citation rates in that journal go up by more than twenty percentage points. In a second step, we explore how the citation rate of a scientist's papers change when one of her colleagues is appointed (resp. leaves) as an editor in one of the top three journals.

Our paper is related to the literature studying social forces in the organization of science. Merton et al. [1968] defines the concept of cumulative advantage, whereby already established papers or scientists automatically keep getting higher peer recognition than those who have not yet made their mark (the so called "Matthew's effect"). A formalization of cumulative advantage in bibliometrics leads to models of citations where the probability of getting new citations is an increasing function of the stock of existing citations : Price

[1976] and Wang et al. [2013] develop such preferential attachment models where the dynamics of a paper's citation count exhibit path dependency because both the stock of already cumulated citations and the intrinsic quality of a paper matter for the flow of new citations. In the same vein, several papers uncover aspects of scientific careers that are independent from plain scientific productivity: Sandstrom et al. (2005) show that citations are related to social networks, as authors tend to cite more the articles by authors whom they know personally (see Bornmann and Daniel [2008] for a survey on citing behavior in science). Feenberg et al. [2015] establish that randomized placement of papers in NBER emails impact the citation rate of a paper over the next two years. Einav and Yariv [2006] finds that surname initials have an impact on academic success in economics, due to the social norm regarding alphabetic ordering of names in papers. King et al. find that men practice self-citation significantly more than women. Using data on academic promotion committees from Spain, Zinovyeva and Bagues [2015] find that connected candidates succeed better and that these candidates exhibit a significantly worse research record both before and after the evaluation. Ellison [2000] documents a rise in the time it takes to publish a paper in economics top journals. Card and DellaVigna [2014] and Card and DellaVigna [2013] uncover other time trends in economics journal publication that shape the profession, notably the increasing relative scarcity of space in the top five journals.

Most related to our work are two papers which reject the existence of editorial favoritism in economics and finance journals. Laband and Piette [1994] find that papers that are published in a journal edited by a colleague of one of the authors are cited more. They interpret this result as "empirical support for the contention that the editorial process is competitive and that editors use their connections to actively search out high-impact papers for publication in their journals." Brogaard et al. (2012) find a similar result. Using a larger dataset, they confirm that colleagues of an editor benefit from higher publication rates in the editor's journal but argue that such publication premium is not due to favoritism but to improved selection based on the editor's inside information regarding her colleagues' work. They test this hypothesis by confirming Laband and Piette (1994) result that "connected" papers are quoted significantly more, suggesting higher quality. Our analysis uses similar data but is different from Brogaard et al. (2012) since our topic is not publication favoritism but biases in citations.

Last, several papers study academic careers and the importance of knowledge spillovers for academic productivity. Oyer [2006] finds a causal effect of quality of an economist's initial placement job on her long-term career path. Kim et al. [2009] find that scientific productivity is less causally affected by being in a top department post 1990s than it was in the 1970s. Dubois et al. [2014] study mobility of mathematicians and its impact on productivity; they find that selection effects are much stronger than spillover effects.

While these papers show that ”luck can matter” in scientific outcomes, they do not challenge a frictionless view of the peer-reviewing process and citation game. Our paper abstracts from human capital complementarities or knowledge spillovers to focus on the impact of editorial power on publication and citation outcomes. We show that the citation game is directly affected by social influence forces: authors submitting to a journal cite the journal's editors and editors colleagues more often.

3.2 Data

Our analysis covers three finance journals, namely *The Journal of Finance*, *Journal of Financial Economics*, and *Review of Financial Studies*. This section explains how we construct our dataset.

3.2.1 Editor Turnover and Affiliations

We collect editors' names and their affiliations by manually checking the corresponding publications. There are various positions in the editorial board, and we only include *Editor*, *Managing Editor*, *Co-Editor*, and *Executive Editor* into our analysis. To this end, we construct a Journal-Issue-Editor dataset. And then we aggregate this dataset to Journal-Year-Editor level: We create a dummy $Editor_{i,j,t}$ equals 1 if author i has been an editor at journal j during year $t+2$. The 2-year shift is aimed at taking into account the lag between submission and publication of papers. We also create an institution-level dummy called editor which equals to one if there is at least one editor affliated with this school for journal j at year $t-2$. Table C.2 lists details of affiliations of journals from 1971 - 2014.

3.2.2 Publication and Citations

We collect publication, citation, and reference information from the Web of Science website[1] We first search items with publication title as *Journal of Finance*, *Journal of Financial Economics* and *Review of Financial Studies* and download the publication page which contains basic bibliographic information such as publication year, volume, issue, author(s), affiliations, abstract, document type and many others. We then retrieve the citation and reference pages from the publication page for each item. Reference page is a page that lists of all the references in the item's reference sections. It contains basic information such as title, author(s), journal, and publication year, volume, and issue

[1] We conduct the downloading process from Jan to April 2015.

for each reference listed. The citation page lists published citations of the publication if the citation journal is within the subscription. The page provides direct links to the citations, and we can retrieve their bibliographic information similar to publication page.[2]

Figure C.5 depicts the total number of publications and total number of references of publications cites across journal and year. There are a steady decreasing trend in total publications in *Journal of Finance* and increasing trends in both *Journal of Financial Economics* and *Review of Financial Studies*.

Table IV presents the summary statistics of our data. Panel A looks at the publication statistics at school-journal-year level. There are 543 schools included in our sample with average publication rate at 0.2 per journal per year. Within these school-journal-year observations, only 0.5% have an author who is an editor in the top finance journals.

Panel B looks at the article level information. To be included in the sample, the article should have document type as "Article "or "Article"; Paper and Proceedings". This restriction effectively remove all other types of publications such as "Corrections", "Note", "Editorial Report". We denote c_i^{Fina} for citations within the top finance journals for article i, and c_i for citations that could be found within *Web of Science*. The average articles in our data has about 62 overall citations and 11 citations within top three finance journals. For each article i, we compare the authors affiliation with editors' affiliations, and define a dummy variable $Connected_{i,J(i),0}$ which equals to 1 if authors of article i has the same affiliations as one of the editors of publication journal i two years before publication date, and 0 otherwise (the two year lag is in order to take publication lag into account, in line with the definition of the editor dummy above). We further define a dummy variable $Connected_{i,any,0}$ which equals to 1 if an author of article i has the same affiliations as at least an editor of the top three finance journals two years before publication, and 0 otherwise. On average, 10 percent of total publications are connected publications. And over 20% percent of the total publications are connected to any of the top finance journals.

We also construct control variables suggested by the literature. The first one is the *citation stock*$_{it}$, which is the total cumulative citations within the top three finance journals prior to year t for article i. $ArticleAge$ is the difference between citation year and publication year for article i. For connectedness, we explore the time series differences and define a dummy variable $Connected_{i,j,t}$ which is equal to 1 if article i's authors

[2]To test the accuracy of this data extraction method, we also download all publications from *Review of Financial Studies* from its publisher *Oxford Journals* website which allows for such download. We then compare the total number of publications across years, and number of connected articles across years in Figure C.10, respectively. We find that the two data extraction methods yield similar results.

share the same affiliation of journal j's editor at year t[3] Overall, there are 8% of articles that are connected.

3.3 The citation premium of editors in their own journal

A first question that we investigate is whether editors are cited more during their own tenure, in the journal that they edit. We restrict our analysis to the top three finance journals: *Journal of Finance*, *Journal of Financial Economics*, *Review of Financial Studies*. We use the sample of all researchers who were editors, managing editors, executive editors, or co-editors at some point for these three journals in the time-period 1971-2014. For each author i, we define $c_{i,j,t}$ the number of her citations in journal j during year t.

3.3.1 A simple graph

A first simple graphical pass at the question consists in plotting the editor's share in citations in a journal before, during, and after this editorship. Absent other controls, looking at citation shares rather than citation numbers is a natural way to account for the fact that numbers of citations vary substantially across time and journals (see Figures 2 and 3). Like in all the rest of the paper, editorial mandates dates are shifted two years forward, to account for the publication lag : a publication finally published at time t might quote an editor, based on a bibliographic choice made at date t-2.

In Figure C.1, we see that the citation share of a researcher jumps in the journal where she becomes editor, but drop back to its initial level soon after she steps down. To be able to compare, we also draw the average citation share in the two other journals than that where the person is an editor, and we observe no such jump. This gives a sort of diff-in-diff identification: Using other journals as a control group helps interpreting the jump we see as the direct effect of being an editor, since such effect is not visible in other top journals.

3.3.2 Regressions with controls

To control and quantify our effect more tightly, we next perform regressions that add important controls to the simple diff-in-diff graph provided in Figure C.1. Namely, we regress:

[3]We rely on the authors' affiliations at their publication year for proxy.

$$log(1 + c_{i,j,t}) = a + bEditor_{i,j,t} + \sum_{k=1}^{10} b_k PostEditor^k_{i,j,t} + controls_{i,j,t} + \epsilon_{i,j,t} \qquad (3.1)$$

$c_{i,j,t}$ is the number of citations of author i in journal j during year t. The dummy $Editor_{i,j,t}$ equals 1 if author i is an editor at journal j at year t. It gives a direct measure of the citation premium enjoyed by editors in their own journal during their appointment. The dummies $PostEditor^k_{i,j,t}$ are equal to 1 if author i was an editor at journal j during year $t - k$ but not from $t - k + 1$ to year t. The coefficients on these dummies measure the speed at which the editorial premium fades away once an editor steps down.

We control for characteristics of the author: $Yearssincefirstcitation$ is a variable counting the number of years since author i's first citation, which is a proxy for academic seniority. $log(CitationStock_{i,t})$ is the log of stock citations for author i before year t. This is a control

Last, we also control for various kinds of fixed effects. Particularly interesting is the author-year fixed effect in the last column, which identifies the $Editor$ and $PostEditor$ dummies based on the relative amount of publications in the journal where the individual is an editor relative to other journals (We apply a two year lag on these variables to account for the delay between publications and submissions, as explained earlier).

Equation 3.1 can be derived from the preferential attachment model of citation dynamics (Wang et al. [2013], Price [1976]). In that model, the probability that a paper i published at t_0 is cited during an infinitesimal interval $[t, t + dt]$ is proportional to

$$\eta_i f(t - t_0)(CitationStock_{i,t})^\alpha dt \qquad (3.2)$$

where f is a decreasing positive function and η_i is the "fitness" of the paper, which can be interpreted as its scientific value. f captures the aging of a paper. The existing stock of citations of a paper affects multiplicatively its current probability to be cited: This is the "Matthew's effect" uncovered by Merton et al. [1968], which states that papers that are already well cited keep getting noticed and cited again. This model, coming from the network formation literature, is known to fit the data well (Wang et al. [2013]): In particular, it generates flows of citations which have the shape observed in Figure 10. Taking logs of that probability, we see that, up to a journal-specific constant, the log number of citations per unit of time is: $log(c_{i,j,t}) = log(\eta_i) + \alpha log(CitationStock_{i,t})) +$

$log(f(t-t_0))$. In our own specification we log-linearize f and approximate $log(c_{i,j,t})$ by $log(c_{i,j,t}+1)$ which deals with potential zeros in citation numbers.[4]

Our estimations consist in evaluating whether events that are unrelated to scientific quality of papers (such as stepping down as an editor) have temporary impacts on the fitness parameter η_i. If these events are clearly exogenous to the scientific quality of the paper, this provides a counter-example to the view that, up to random noise, only scientific quality determines citation rates. Note that, using the specification of Equation 3.1 implies that we control for the path-dependency effect (or hysteresis) that arises from the stock of citations determining the future flow of citations: When we say that a certain event tends to make the paper be cited relatively more, this always mean net of the Mathew's effect, i.e. controlling for the pre-existing stock of citations.

The results are in Table C.5. All specifications yield a highly significant Editor dummy. As can be seen from the coefficients b_k, this citation premium fades away after the editor steps down: five years after leaving office, an editor does not benefit from any sort of citation premium (Figure C.2 plots the decline in excess citations using coefficients b and b_k from column 2 of Table C.5). Column 2 is the tightest identification, as it includes author-year and journal fixed effects. This means that we measure the citation premium in a diff-in-diff manner where the journals where the scientist is not serving as an editor serve as a control group. There are two sorts of events in this diff-and-diff: becoming an editor and stepping down as an editor. Clearly the second one is the most exogenous, so in our interpretation and measurement of an editorial citation premium, we focus on the effect of leaving an editorial position as an editor.

The economic size of the citation premium of editors is large: a coefficient b of .4 as in column 2 (the most conservative) means a log elasticity of $(c_{i,j,t}+1)$ equal to 40%; Given that the average editor in our data has $c_{i,j,t}=5.6$, this means that the average editor's citation in her journal go up (resp. down) by 23 percentage points[5] when she becomes an editor (resp. steps down). In robustness regressions we also run the same specifications as in Table C.5 using a Poisson regression (Table C.13) and also a regression without the two-year shift (Table C.14): We obtain highly similar results.

The endogeneity of the appointment of editors is an obvious concern that can arise when interpreting the results: Researchers might be appointed as editors when they are rising stars, with an accelerating number of citations. We should therefore be worried about reverse causality and refrain from attributing the editorial premium to the power

[4] In a robustness regression we also run a Poisson regression to deal with discreteness and zeros and find highly similar results (Table C.13).

[5] $\frac{dc_i}{c_i}=(1+\frac{1}{c_i})\frac{dc_i}{c_i+1}=(1+\frac{1}{5.6})(exp(.4)-1)=.23$

given by the editorial position itself. A first answer to this concern is that author-year fixed effects allow us to identify our editor effect on citations by comparison to the citation performance of the editor in other journals. Thus, unless the researcher's recognition is confined to only one journal, the "rising-star" effect should be captured by these fixed effects. A second answer to the concern that editorial appointments are endogenous to citation dynamics, is that, fortunately, the stepping-down of editors (typically happening after their mandate expires), is not subject to the same problem. Editors are typically not "fired" because their research is seen as slowing down during their editorial appointment. Therefore, the fact that the coefficients b_k are fading away after the editor steps down is a convincing sign that the editorial job is actually causing the excess citation performance during the appointment.

The citation premium that editors benefit from in their own journal might be due to two, direct and indirect, effects that are hard to disentangle: (1) A direct influence on the bibliography (editors might strongly suggest that their own papers should be quoted, during the refereeing process) and (2) salience-driven or pandering-driven citations (authors might anticipate better treatment when quoting the editor). We do not distinguish between these two effects in our analysis.

3.4 Are editors' colleagues cited more?

We now investigate the following question: Do people who publish in journal j quote at an abnormally high level the papers of colleagues of the editors at journal j? This would be a symptom of bias in the citation game that would amplify the one we just found on editors' own citations. It is important because, in our sample, papers of scientists who have been or will be editors of one of the top three finance journals only represent only x%[6] of the stock of papers, whereas, at a given point in time, XX%[7] of the papers are currently connected to an editor of one of the top three journals. If a bias exists also for this population, this would represent a more widespread distortion of the citation game.

3.4.1 A simple graph

As in the previous section, we first produce a simple graph that captures in a transparent manner the effect we are after. For each school i, we define its citation share in journal j at year t. Each time a school has a faculty acting as an editor of a journal, we construct a window of time running from 4 years before editorship to 5 years after editorship and

[6] To be added.
[7] To be added.

we call "Same journal citation share", the citation share of the school in the edited journal and "Other journals citation share" the average citation share in other journals. We then draw the average of these quantities over all episodes of editorship. The graph shows that a school citation's share almost doubles at the journal where one of its faculty becomes an editor; this effect fades away after the editor steps down. The citation share in other journals also goes up and down , albeit to a lower extent. This might reflect that the greater salience of the school during the editorship spills over to the citations in other journals; it might also be that bibliographic choices made before submitting for a journal are not changed after the paper is rejected and submitted to another journal.

3.4.2 Regressions with controls

We next run regressions that add controls to the previous descriptive graph and evaluate statistical significance of the effect described in the simple graph:

$$log(1 + c^{Fina}_{i,j,t}) = a + b.Connected_{i,j,t} + c.Connected_{i,other,t} + controls_{i,j,t} + \epsilon_{i,j,t} \quad (3.3)$$

The variable $c^{Fina}_{i,j,t}$ is total citations for article i in Top finance journal j at year t. The dummy variable $Connected_{i,j,t}$ equals 1 if the article i's author's affiliations overlap with one of the editors of journal j at time $t - 2$. $Connected_{i,other,t}$ equals 1 if the article i's author's affiliations overlap with any journal $j' \neq j$ at time $t - 2$.

We use the following controls: $\log(CitationStock_{i,t})$ is the log of stock citations for article i before year t. $ArticleAge_{i,t}$ is the age of publication i at year t. These variables are known to determine the flow of future citations (Wang et al. [2013]). We control for several layers of fixed effects. For instance, Journal of Publication - Journal of Citation Fixed Effects Coefficients allows specific citation flows for each pair of journal. Coefficients are clustered at article level, to allow serial correlation in the flow of citations in a given paper.

Let's focus on column 5, which has article-year fixed effects, so that the coefficients are identified by comparing relative citation rates of an article in different journals depending on whether the article has an author who is a colleague of an editor of journal j at time $t - 2$. This is a pretty tight identification: to make its logic more intuitive, suppose that our data consists in one single article i, published in 2007, which has an author who is a colleague of Mr X; Mr X was an editor of the JF between 2011 and 2014; suppose further that the paper has no other connections to editors during its life-time. The regression uses article i in the estimation of the coefficient on $Connected_{i,j,t}$ by looking at whether there is a jump in citations in the JF during 2013-2015 vis-a-vis other journals. In other words, we can think about the identification as a quasi diff-in-diff where we use observation i by computing

$$\overline{[c_{i,JF,t} - (c_{i,JFE,t} + c_{i,RFS,t})/2]}^{[2013-2015]} - \overline{[c_{i,JF,t} - (c_{i,JFE,t} + c_{i,RFS,t})/2]}^{[2007-2012]}.$$

Of course in our data there are many articles, published in all journals, and with various connections to journals over time; Moreover, we add controls to take into account systematic effects, such as the fact that citations in one journal might be on average higher than in others. So the estimate of the coefficient on $Connected_{i,j,t}$ is obviously more complex than the previous formula, but it captures the spirit of the identification strategy nonetheless.

What we see, is that articles published in j at time t tend to "overcite" by roughly two percentage points the articles of colleagues of editors from journal j. Interestingly, no such effect is found for colleagues of editors from other journals than j. When it is identified (which implies that we do not control for article-year FE), the coefficient on $Connected_{i,other,t}$ shows up positive and statistically significant (albeit smaller than that on $Connected_{i,j,t}$), which is in line with the graph and likely to reflect the fact that editors' colleagues are "salient" or that people do not change bibliographies after a paper being rejected.

In panel B , we perform a robustness check, where we exclude papers that are connected to their publication journal at time of their publication. This is to mitigate the concern that our results might be polluted by the fact that colleagues' papers published by an editor might be of higher quality (an thus cited more) due to the editor having private information on her colleagues. We find that imposing this restriction on our sample does not affect our conclusions.

To summarize, we find clear evidence that papers cite the colleagues of their publication journal more than normal. This suggests that editors bias the citation game (deliberately or not) in a statistically significant manner. Having ties to an editor helps being cited more. This is a clear counterexample to the hypothesis that the number of citations provides an unbiased measure of an article's scientific value

3.5 Are colleagues' papers published by editors really of higher quality?

Brogaard et al. [2014] show that colleagues of an editor publish at higher rates than normal in the editor's journal. They also show that these papers with editorial ties at the publication journal are quoted more. This leads them to reject the hypothesis that the high publication rates of colleagues are partly due to some form of positive bias. Instead, they conclude that editors are likely to be making an efficient use of their private information on which of their colleagues' papers are really good. Their informational advantage would therefore explain the editors' colleagues abnormally high publication rate.

In this section we propose to revisit Brogaard et al. [2014]'s conclusion that editorial decisions regarding colleagues are purely driven by quality. As a starting point, we replicated the results from Brogaard et al. [2014] and we find like them that (1) colleagues of an editor publish relatively more in the editor's journal and (2) papers published by a colleague editor are cited relatively more (see appendix Tables C.10 and C.12). This second result holds whether we use Brogaard et al. [2014]'s methodology, which looks at total stock of citations of papers at a same final date T, or whether we use the panel specifications that we have been using throughout the paper (see appendix Table C.10, Panel B).

3.5.1 Two competing hypotheses: quality selection vs. citation bias

When an editor has some of her colleaues papers published in her journals, these papers are cited more. From this, Brogaard et al. [2014] conclude that favoritism can be rejected. They consider the higher citation rates of "connected" papers as indicative that that these papers are (efficiently) selected by editors based on the inside information that they have on their colleagues portfolio of working papers.

However, Brogaard et al. [2014] do not consider the possibility that citations of "connected papers" are inflated simply because other authors try to please the editor by citing these papers (or get advised to do so) or simply that the work of the editor's colleague are somehow made more salient (people have them in mind before submitting). Our goal is to perform a series of tests that help disentangle two competing hypotheses explaining the higher citation rates of "connected papers":

- Hypothesis 1 (quality): editors publish high quality papers from colleagues based on their informational advantage, thus these papers are cited more.
- Hypothesis 2 (influence/pandering/salience): because people try to please editors or follow their bibliographical advice or are more likely to have in mind work produced in the school of editors (salience), colleagues papers published by the editor are cited more.

3.5.2 What matters for citations is being colleague of an editor, not that the editor publishes the paper

A first test which is natural to disentangle between our two hypotheses is to check if excess citations of papers published at time t by a colleague of an editor are higher regardless of whether that paper is published in the editor's journal or in another: Indeed, the editor might be pleased by his colleagues being quoted independently of the journal where the paper is published. But under the "quality" hypothesis, only the papers of his colleagues that are published in the editor's journal should be cited more; not those published in other journals.

Thus we rerun Brogaard et al. [2014]'s excess citation regression by adding to $Connected_{i,J(i),0}$ (which states that a paper is connected to its own publication journal two years before publication) a second dummy, $Connected_{i,any,0}$, which states that paper i is connected to any publication journal two years before publication. $1_{j=J(i)}$ equals 1 if journal j is the publication journal of article i, and 0 otherwise.

We use the following controls: $\log(CitationStock_{i,t})$ is the log of stock citations for article i before year t. $ArticleAge_{i,t}$ is the age of publication i at year t. These variables are known to determine the flow of future citations (Wang et al. [2013]). We control for several layers of fixed effects. Journal of Publication - Journal of Citation Fixed Effects Coefficients allows specific citation flows for each pair of journal. Coefficients are clustered at article level, to allow serial correlation in the flow of citations in a given paper.

The results are in Table C.7. What we see is that all the information that predicts high future citations is in the second dummy: the coefficient on $Connected_{i,J(i),0}$ is insignificant and that on $Connected_{i,any,0}$ is significant. In other words, it makes no difference for future publications that an article be connected to its publication journal vs. to any other journal. This is in line with the view that what drives excess citations is simply pandering to editors by quoting their colleagues papers, whatever their publication outlet.

3.5.2.1 Are connected papers still cited more once the editor steps down?

We now explore the time-series of citations by asking the following question: Once the editor who published a colleague's paper steps down, is there still a citation premium for that paper?

This test helps disentangling the two competing hypotheses. Indeed, under the quality hypothesis, the answer should be yes. Under the pandering/influence hypothesis, the answer should be no, because the incentive to please the (former) editor vanishes once he has stepped down.

We run the following regression:

$$\begin{aligned} log(1 + c^{Fina}_{i,j,t}) = a + b.Connected_{i,J(i),0} + c.Connected_{i,J(i),t} + \\ d.Connected_{i,J(i),t}.Connected_{i,J(i),0} + controls_{i,j,t} + \epsilon_{i,j,t} \end{aligned} \tag{3.4}$$

The dummy variable $Connected_{i,J(i),0}$ equals 1 if the article i's author's affiliations overlap with its publication journal's editors affiliation at year of publication minus two.

The results are gathered in Table C.8. Let's focus on columns 4, which includes journal of Pub.-Journal of Cit. fixed effects and year fixed effects. What we find is that as soon as we control for whether a paper is currently connected to a journal, there is no more citation premium regarding that paper. More precisely, the citation premium that exists for papers published by a colleagues editor is only present when the paper is currently connected to the citing journal via an editor. As soon as the paper is not linked to the journal, there is no citation premium for initially connected papers. During the time where they are connected, papers published by a colleague editor enjoy an extra-level of citation compared to simply connected paper: this is measured by the interaction coefficient. This suggests that the editor is "very" pleased when a colleague paper that he published himself is cited. Column 5 shows that the excess citation premia on current connectedness is stable when adding a paper fixed-effect (but of course, in that case, the variable $Connected_{i,J(i),0}$ which is constant for each paper, is not identified). Overall, this regression strongly suggests that the excess citation of papers published in a journal edited by a colleague is driven by biased citation behavior that takes place while the editor is in office, rather than being caused by the intrinsic quality of the papers.

3.5.2.2 Citations in own vs. other journals

After having explored the time-series of citations of a given paper, we now explore the variation of these citations across journals: In this last test, we compare citations in the publication journal vs. in other journals for the papers that are published in a journal edited by a colleague.

If the higher citation effect is restricted to the journal where the article is published, it is likely to be driven by pandering to the editor or selection of articles that please the editor.

By contrast, if the excess citation result of Brogaard et al. [2014] holds for other journals than the publication journal, since these journals are a priori outside the editor's influence, it is legitimate to conclude that the articles have higher quality.

We perform the following regression:

$$log(1+c^{Fina}_{i,j,t}) = a+b.Connected_{i,J(i),0}+c.1_{j=J(i)}+d.1_{j=J(i)}.Connected_{i,J(i),0}+controls_{i,j,t}+\epsilon_{i,j,t} \quad (3.5)$$

The variable $c^{Fina}_{i,j,t}$ is total citations for article i in Top finance journal j at year t. The dummy variable $Connected_{i,J(i),0}$ equals 1 if the article i's author's affiliations overlap with its publication journal's editors affiliation when it publishes.

The results (Table C.9) show that, controlling for the effects we describe above, there is no observable citation premium for articles published at the journal where a colleague is an editor (b=0). All the citation premium is concentrated in citations from the *same* journal as the publication journal (d¿0), which suggests this is likely due to a form of bias (pandering to the editor, selection by the editor, bibliographical changes imposed by the editor). A similar test can actually be found for finance journals in Brogaard et al. [2014] who show in a robustness table (Table 7, panel B) that when excluding same journal citations, the citation premium of connected articles vanishes, as soon as one also controls for Journal-Year Fixed Effects and School Fixed Effects.

3.6 Conclusion

The evaluation of scientific output plays an increasingly important role in the organization of science: Grants and tenure decisions are partly based on measures of impact such as citation counts. This paper does not challenge the view that such transparent and objective measures of impact are important and desirable. But it sheds some light on the complexity of the citation game, which is not immune to the presence of extra-scientific psychological or social forces. We show that editors tend to be cited more in their own journal, and that papers published in their journal also cite their own colleagues more, introducing a potential bias in citation count as a measure of scientific quality. Thus, we reject the view that editors play the role of neutral gate-keepers, as their very presence affects the citation game. The exact source of the biased citation behavior might have several causes: it might result from pandering by researchers who try to submit papers which are likely to please the editor, it might be directly generated by the editor's policy, or it might result from a simple salience effect, whereby editor's colleagues are more "in the spotlight". Our data, which do not track changes in bibliographies during the editorial process, do not allow us to disentangle these effects.

Appendix A

Appendix for Chapter 1

A.1 General Identification

A.1.1 Derivation of Equation (1.13)

In this case, we have both dynamic and static inputs, together with exogeneous covariates in the productivity process. There are four equations at hand

$$r_{it} = \theta_l l_{it} + \theta_k k_{it} + \theta_m m_{it} + \omega_{it} \tag{A.1}$$

$$\omega_{it} = \mu + \gamma\omega_{it-1} + \alpha\tau_{it} + \pi x_{it-1} + \eta_{it} \tag{A.2}$$

$$\ln p_t^M = \ln\theta_m + (\theta_m - 1)m_{it} + \theta_k k_{it} + \theta_l l_{it} + \omega_{it} + \tau_{it} \tag{A.3}$$

$$\ln s_{it} = \ln\theta_m + \tau_{it} \tag{A.4}$$

Using the above four equations (A.1), (A.2), (A.3), and (A.4), I derive the reduced form representation of (1.13)

$$\begin{aligned}
r_{it} = \; & \theta_k k_{it} + \theta_l l_{it} + \theta_m m_{it} + \omega_{it} \\
= \; & \frac{\theta_k}{1-\theta_m} k_{it} + \frac{\theta_l}{1-\theta_m} l_{it} + \frac{\theta_m}{1-\theta_m}(\omega_{it} + \tau_{it}) + \omega_{it} + \text{Constant} \\
= \; & \frac{\theta_k}{1-\theta_m} k_{it} + \frac{\theta_l}{1-\theta_m} l_{it} + \frac{1}{1-\theta_m}\omega_{it} + \frac{\theta_m}{1-\theta_m}\tau_{it} + \text{Constant} \\
= \; & \frac{\theta_k}{1-\theta_m} k_{it} + \frac{\theta_l}{1-\theta_m} l_{it} + \frac{\pi}{1-\theta_m} x_{it-1} + \frac{\alpha+\theta_m}{1-\theta_m}\tau_{it} + \frac{\gamma}{1-\theta_m}\omega_{it-1} + \frac{1}{1-\theta_m}\eta_{it} + \text{Constant} \\
= \; & \frac{\theta_k}{1-\theta_m} k_{it} + \frac{\theta_l}{1-\theta_m} l_{it} - \frac{\gamma\theta_k}{1-\theta_m} k_{it-1} - \frac{\gamma\theta_l}{1-\theta_m} l_{it-1} + \frac{\pi}{1-\theta_m} x_{it-1} + \\
& \frac{\alpha+\theta_m}{1-\theta_m}\tau_{it} - \frac{\gamma}{1-\theta_m}\tau_{it-1} + \gamma m_{it-1} + \frac{1}{1-\theta_m}\eta_{it} + \text{Constant} \\
= \; & \frac{\theta_k}{1-\theta_m} k_{it} + \frac{\theta_l}{1-\theta_m} l_{it} - \frac{\gamma\theta_k}{1-\theta_m} k_{it-1} - \frac{\gamma\theta_l}{1-\theta_m} l_{it-1} + \frac{\pi}{1-\theta_m} x_{it-1} + \\
& \frac{\alpha+\theta_m}{1-\theta_m}\ln s_{it} - \frac{\gamma}{1-\theta_m}\ln s_{it-1} + \gamma m_{it-1} + \frac{1}{1-\theta_m}\eta_{it} + \text{Constant.}
\end{aligned}$$

The identification follows Proposition 1.2 in the main text.

A.1.2 Link to Structural Estimation

I produce a generalized method of moment estimator to estimate the production function with unobserved distortions. This links the reduced form representation with the general structural approach proposed by Olley and Pakes [1996], Levinsohn and Petrin [2003] and Ackerberg et al. [2006].

In the reduced form representation of revenue, equation (1.13), the error term is $\eta_{it}/(1-\theta_m)$, a scaled version of the productivity innovation. Since I assume that $\eta_{it} \perp \Gamma_{it}$, this leads to an alternative estimation strategy: First guess $\{\theta_k, \theta_l, \theta_m\}$, we can compute the productivity conditional on the parameters as

$$\omega_{it}(\theta_k, \theta_l, \theta_m) = r_{it} - \theta_k k_{it} - \theta_l l_{it} - \theta_m m_{it}$$

Second, we can then derive the implied productivity innovation $\eta_{it}(\theta_k, \theta_l, \theta_m)$ from a AR(1) regression.

$$\omega_{it}(\theta_k, \theta_l, \theta_m) = \mu(\theta_k, \theta_l, \theta_m) + \gamma(\theta_k, \theta_l, \theta_m)\omega_{it-1}(\theta_k, \theta_l, \theta_m) + \alpha \ln s_{it} + \eta_{it}(\theta_k, \theta_l, \theta_m)$$

and obtain the estimates of productivity innovation $\hat{\eta}_{it}(\theta_k, \theta_l, \theta_m)$.

Finally we can construct moments implied by the markovian assumption of productivity process, i.e.,

$$\mathbb{E}\left[\hat{\eta}_{it}\left(\theta_k, \theta_l, \theta_m\right) \middle| \begin{array}{c} k_{it} \\ l_{it} \\ m_{it-1} \\ k_{it-1} \\ l_{it-1} \\ \ln s_{it} \\ \ln s_{it-1} \end{array}\right] = 0$$

Notice that the last two moments are the keys to identify all parameters of interests. Without these two instruments, the collinearity problem addressed by Gandhi et al. [2012] arises. The structural estimation shares very much the same intuition with the reduced form approach, and yields quite similar statistical properties in the simulation.

A.1.3 Beyond Cobb-Douglas

So far the identification seems to be heavily relied on the Cobb-Douglas form of production function. Here I extend the identification to allow for higher order polynomials to approximate the production function that is similar to Gandhi et al. [2012]. I illustrate the identification using a second order approximation of the production function

$$r_{it} = \theta_m m_{it} + \theta_{mm} m_{it}^2 + \omega_{it}$$

and keep the same setting as in the illustration of identification section.

Using the structural form of estimation, given $\{\theta_m, \theta_{mm}, \mu, \gamma, \alpha\}$, the productivity innovation is

$$\eta_{it} = y_{it} - \theta_m m_{it} - \theta_{mm} m_{it}^2 - \mu - \gamma \left(y_{it-1} - \theta_m m_{it-1} - \theta_{mm} m_{it-1}^2 \right) - \alpha \tau_{it}$$

There are two sets of instruments: the observed instruments m_{it-1}, m_{it-1}^2 and unobserved instruments τ_{it}, τ_{it-1}, and their interaction terms with lagged material $\tau_{it} m_{it-1}$ and $\tau_{it-1} m_{it-1}$. The transformation of instrument is

$$\begin{aligned} s_{it} &= \tau_{it} + \ln(\theta_m + 2\theta_{mm} m_{it}) \\ s_{it-1} &= \tau_{it-1} + \ln(\theta_m + 2\theta_{mm} m_{it-1}) \end{aligned}$$

And the moment condition for both observed and unobserved instruments are

$$\begin{aligned} \mathbb{E}\left[\eta_{it}(\theta_m, \theta_{mm}, \mu, \gamma) \,\middle|\, \begin{matrix} m_{it-1} \\ m_{it-1}^2 \end{matrix} \right] &= 0 \\ \mathbb{E}\left[\eta_{it}(\theta_m, \theta_{mm}, \mu, \gamma) \,\middle|\, \begin{matrix} \tau_{it} \\ \tau_{it-1} \end{matrix} \right] &= 0 \end{aligned}$$

However, the unobserved instruments contain unknown parameters of interests. Since exp() is a strict monotone transformation of τ_{it}, $\exp(\tau_{it})$ is a valid instrument as well. The GMM with transformed unobserved instruments are

$$\mathbb{E}\left[\eta_{it}(\theta_m, \theta_{mm}, \mu, \gamma) \,\middle|\, \begin{matrix} \exp(\tau_{it}) \\ \exp(\tau_{it-1}) \end{matrix} \right] = 0$$

Notice that

$$\begin{aligned} \tau_{it} &= s_{it} - + \ln(\theta_m + 2\theta_{mm} m_{it}) \\ \tau_{it-1} &= s_{it-1} - + \ln(\theta_m + 2\theta_{mm} m_{it-1}) \end{aligned}$$

and

$$\begin{aligned} \exp(\tau_{it}) &= \frac{\exp(s_{it})}{\theta_m + 2\theta_{mm} m_{it}} \\ \exp(\tau_{it-1}) &= \frac{\exp(s_{it-1})}{\theta_m + 2\theta_{mm} m_{it-1}} \end{aligned}$$

We can transform the moment condition such that

$$\begin{aligned} \eta_{it} \exp(\tau_{it}) &= \frac{\eta_{it} \exp(s_{it})}{\theta_m + 2\theta_{mm} m_{it}} = \tilde{\eta}_{it} \exp(s_{it}) \\ \eta_{it} \exp(\tau_{it-1}) &= \frac{\eta_{it} \exp(s_{it-1})}{\theta_m + 2\theta_{mm} m_{it-1}} = \hat{\eta}_{it} \exp(s_{it-1}) \end{aligned}$$

with

$$\begin{aligned} \tilde{\eta}_{it} &= \frac{\eta_{it}}{\theta_m + 2\theta_{mm} m_{it}} \\ \hat{\eta}_{it} &= \frac{\eta_{it}}{\theta_m + 2\theta_{mm} m_{it-1}} \end{aligned}$$

We now have three moment conditions

$$\mathbb{E}\left[\eta_{it}(\theta_m, \theta_{mm}, \mu, \gamma) \middle| \begin{array}{c} m_{it-1} \\ m_{it-1}^2 \end{array} \right] = 0$$

and

$$\mathbb{E}\left[\tilde{\eta}_{it}(\theta_m, \theta_{mm}, \mu, \gamma) \middle| \exp(s_{it}) \right] = 0$$

and

$$\mathbb{E}\left[\hat{\eta}_{it}(\theta_m, \theta_{mm}, \mu, \gamma) \middle| \exp(s_{it-1}) \right] = 0$$

Therefore, by transforming the moment conditions, the identification is still valid when extending to higher order polynomials.

A.2 Matching Results for Productivity and Distortion Process

The estimated results in section **??** relies on cross year and provinces variations to identify the effects from privatization on productivity and distortions. The identification of the revenue function only requires that the unpredictable components of the productivity process is orthogonal to any elements in the firm's information set, which includes privatization. To alleviate potential selection issue, following Braguinsky et al. [2015] and Cooper et al. [2013], I rely on the propensity score matching for those privatized firms, and investigate the productivity and distortion effects in that sample.

I matched the privatized firms i which is privatized at year $t-1$ with SOEs that are not privatized on year $t-1$ or t from the same industry, region and exporting status, and with similar employment, productivity and distortion parameters. There are some privatized firms that the I cannot find matches. Finally I have a matched sample with 15279 observations. Table A.9 compare the pre-privatized and post-privatized characteristics. The matched sample is much closer to the privatized sample comparing to any SOEs.

Table A.9 is here.

This result suggests that privatization increases the productivity of the privatized firms comparing to a group of control firms that share similar firm level characteristics. The magnitude of the effect is about 4%, which is close to the estimated effect of privatization on productivity in section **??**. However, the effect of privatization on distortions are insignificant, which is also consistent with the results in section **??**. I further plot the mean productivity and mean absolute value of distortions around the privatization window in Figure A.13.

A.3 Decomposition with Ownership Switches

Olley and Pakes [1996] first decomposes the aggregate productivity, defined as a market share weighted productivity, into a industrial mean productivity and a covariance term between firm sizes and productivity. Bartelsman et al. [2013] uses this approach to highlight the usage of the covariance term in approximating distortions across countries. Following Olley and Pakes [1996] and Melitz and Polanec [2015], I consider group-wide aggregate productivity within an industry. Define the aggregate market share of a group G of firms

$$s_{Gt} = \sum_{i \in G} s_{it}$$

where s_{it} is the market share of firm i at year t, and firm i belongs to group G at year t. The group aggregate (average) productivity is defined as

$$\Phi_{Gt} = \sum_{i \in G} \frac{s_{it}}{s_{Gt}} \phi_{it}$$

In the original Melitz and Polanec [2015], there are only three groups in the two cross sectional data for year t and year t', and with out loss of generality I assume $t' > t$. The first group S is for survivors, which could be found in both year t and t' data. The second group X is for exiters, which could only be found in year t's data. And the third group E is for entrants, which could only be found in year t''s data. Therefore we can decompose the aggregate productivity into group aggregate productivity

$$\begin{aligned}
\Phi_t &= s_{St}\Phi_{St} + s_{Xt}\Phi_{Xt} = \Phi_{St} + s_{Xt}\left(\Phi_{Xt} - \Phi_{St}\right) \\
\Phi_{t'} &= s_{St'}\Phi_{St'} + s_{Et'}\Phi_{Et'} = \Phi_{St'} + s_{Et'}\left(\Phi_{Et'} - \Phi_{St'}\right)
\end{aligned}$$

Following Melitz and Polanec [2015], the decomposition uses the survivors group as the benchmark for entrants and exiters, which is the key differences comparing to other decomposition method like Baily et al. [1992], Griliches and Regev [1995] and Foster et al. [2001].

I then rewrite the aggregate productivity change between t' and t as the differences within the survivor group and entry and exit components. I can then further decompose the survivors components to an average productivity shift term and a difference in reallocation term within survivors using Olley and Pakes [1996] decomposition.

$$\begin{aligned}
\triangle\Phi &= \underbrace{(\Phi_{St'} - \Phi_{St})}_{\text{Survivors}} + \underbrace{s_{Et'}(\Phi_{Et'} - \Phi_{St'})}_{\text{Entrants}} + \underbrace{S_{Xt}(\Phi_{St} - \Phi_{Xt})}_{\text{Exiters}} \\
&= \underbrace{\triangle\bar{\phi}_S}_{\text{Productivity shift}} + \underbrace{\triangle cov_S}_{\text{Reallocation}} + \underbrace{s_{Et'}(\Phi_{Et'} - \Phi_{St'})}_{\text{Entrants}} + \underbrace{S_{Xt}(\Phi_{St} - \Phi_{Xt})}_{\text{Exiters}}
\end{aligned}$$

To address the transitions from State-owned firms to private firms, as well as the reallocation effects within each ownership during the transition period, I further build on Collard-Wexler and Loecker [2015] to decompose the reallocation effects among survivors to finer groups, state-owned firms and non state-owned firms. Following Collard-Wexler and Loecker [2015], for survivors with ownership $\phi \in (State, Non-State)$, I define the within-ownership aggregate productivity as

$$\begin{aligned}
\Omega_t^S &= \sum_{\phi} s_t(\phi)\left(\bar{\omega}_t(\phi) + \sum_{i \in \phi}(\omega_{it} - \bar{\omega}_t(\phi))(s_{it} - \bar{s}_t(\phi))\right) \\
&= \sum_{\phi} s_t(\phi)\left(\bar{\omega}_t(\phi) + \Gamma_t^{OP}(\phi)\right)
\end{aligned}$$

where $\bar{\omega}_t(\phi)$ is the average of productivity of group ϕ at year t, and $\Gamma_t^{OP}(\phi)$ is a OP covariance term within ownership ϕ at year t.

Similar, the Between-Ownership Decomposition for the survivors without transition in ownership is as follows: I focus on those firms which are survivors and do not have an ownership switch between t and t'. Denote $\bar{\Omega}_t = \frac{1}{2}\sum_\phi \Omega_t(\phi)$, which is the average productivity for those firms across ownerships. The aggregate productivity for those survivors (without ownership changes) is

$$\Omega_t^S = \bar{\Omega}_t + \sum_\phi (s_t(\phi) - 1/2)\left(\Omega_t(\phi) - \bar{\Omega}_t\right) = \bar{\Omega}_t + \Gamma_t^B$$

where Γ_t^B is the between-ownership term. Therefore, the aggregate productivity for survivors without a ownership change is defined as

$$\Omega_t^S = \frac{1}{2}\sum_\phi \left[\bar{\omega}_t(\phi) + \Gamma_t^{OP}(\phi)\right] + \Gamma_t^B$$

So far the exercise is quite similar to Collard-Wexler and Loecker [2015] except their focus is the technology and my focus is the ownership. The last piece is the contribution from those survivor firms which transit from state-owned to non-state-owned. Therefore, I decompose the survivor contribution into two components, first, the survivors that never change ownerships, and second, the survivors do change ownerships. To this end, the aggregate productivity growth is decomposed into eight components: Two within-ownership average productivity growth terms, two changes of within-ownership reallocation terms, one between ownership reallocation term, one between ownership transition term, and entry and exit components using survivor groups as benchmark.

$$\begin{aligned}\triangle\Omega =& \frac{1}{2}\sum_\phi \left[\underbrace{\triangle\bar{\omega}_t(\phi)}_{\text{Productivity Growth}} + \underbrace{\triangle\Gamma_t^{OP}(\phi)}_{\text{Reallocation Within Ownership}}\right] + \\ & \underbrace{\triangle\Gamma_t^B}_{\text{Reallocation Between Ownership}} + \underbrace{\triangle\Gamma_t^T}_{\text{Transition Between Ownership}} + \\ & \underbrace{s_{E2}(\Omega_{E2} - \Omega_{S2})}_{\text{Entry}} + \underbrace{s_{X1}(\Omega_{S1} - \Omega_{X1})}_{\text{Exit}}\end{aligned}$$

A.4 Tables

Table A.1: Privatization during 1999-2006, by industry.

This table shows the number of privatization events during 1999 to 2006 by two digit industrial classification. I record a firm i at year t with registration code 110 or 151 as a state-owned firm regardless of its capital composition. A firm i is privatized at year t if firm i is a state-owned firm at year t but a non-state-owned firm at year $t+1$. Industries are classified by 2-digit industrial classification code. I adopt the translation of industries from Bai et al. [2009].

Industry	ind	1999	2000	2001	2002	2003	2004	2005	2006	All Years
Food Processing	13	213	214	132	185	128	65	65	30	1032
Food Production	14	91	89	43	52	46	16	19	14	370
Beverage Production	15	100	89	55	61	54	17	17	11	404
Textile Industry	17	100	107	77	92	74	30	29	15	524
Garment and Other Fiber Products	18	13	21	5	12	11	11	6	4	83
Leather, Furs, Down and Related Products	19	14	11	6	5	4	5	2	1	48
Timber and Bamboo Processing	20	19	22	10	18	17	4	4	5	99
Furniture Manufacturing	21	9	3	2	5	2	0	2	1	24
Papermaking and Paper Products	22	38	36	20	48	43	10	11	5	211
Printing and Record Medium Reproduction	23	44	44	30	39	50	14	26	8	255
Cultural, Educational and Sports Goods	24	5	6	4	2	3	2	1	1	24
Petroleum Refining and Coking	25	11	23	14	11	16	9	10	6	100
Raw Chemical Materials and Chemical Products	26	180	172	112	154	156	45	57	31	907
Medical and Pharmaceutical Products	27	92	104	73	76	92	25	21	10	493
Chemical Fiber	28	5	8	8	5	7	3	3	1	40
Rubber Products	29	19	13	16	13	9	5	0	2	77
Plastic Products	30	29	20	18	22	19	5	5	8	126
Nonmetal Mineral Products	31	225	174	128	196	168	63	67	35	1056
Ferrous Metal Mining and Dressing	32	20	25	20	48	24	10	10	7	164
Nonferrous Metal Mining and Dressing	33	21	18	18	16	19	14	10	3	119
Metal Products	34	38	35	14	24	27	11	21	7	177
Ordinary Machinery	35	122	109	71	113	95	42	50	38	640
Special Purposes Equipment	36	94	96	67	88	84	24	43	16	512
Transport Equipment	37	78	88	57	83	92	30	59	31	518
Electric Equipment and Machinery	39	68	63	35	58	49	28	23	14	338
Electronic and Telecommunications	40	37	41	18	34	35	11	21	14	211
Instruments, meters, Cultural and Clerical Machinery	41	28	16	12	12	24	7	10	12	121
Other Manufacturing	42	9	7	7	4	3	2	4	0	36
Overall		1722	1654	1072	1476	1351	508	596	330	8709

Table A.2: Privatization during 1999-2006, by provinces.

This table shows the number of privatization events during 1999 to 2006 by provinces. I record a firm i at year t with registration code 110 or 151 as a state-owned firm regardless of its capital composition. A firm i is privatized at year t if firm i is a state-owned firm at year t but a non-state-owned firm at year $t+1$. I group provinces into four regions, namely eastern, central, northeastern, and western.

Provinces	Region	1999	2000	2001	2002	2003	2004	2005	2006	All Years
Anhui	Central	105	46	42	48	37	23	23	8	332
Beijing	Eastern	15	81	18	26	57	18	37	27	279
Chongqing	Western	14	30	3	19	20	8	6	1	101
Fujian	Eastern	25	20	14	25	15	12	6	6	123
Gansu	Western	83	115	0	36	16	15	6	6	277
Guangdong	Eastern	226	89	30	55	74	26	40	26	566
Guangxi	Western	32	24	31	55	39	20	23	10	234
Guizhou	Western	13	13	10	28	29	12	10	10	125
Hainan	Eastern	1	2	1	9	16	0	6	2	37
Hebei	Eastern	77	75	65	70	65	37	41	19	449
Heilongjiang	Northeastern	61	34	35	46	23	18	17	9	243
Henan	Central	1	56	43	113	137	27	27	16	420
Hubei	Central	91	118	132	142	73	35	30	14	635
Hunan	Central	63	57	20	73	85	27	19	15	359
Inner Monglia	Western	47	10	15	24	14	5	8	6	129
Jiangsu	Eastern	236	153	152	159	99	25	37	22	883
Jiangxi	Central	47	95	56	61	23	13	17	5	317
Jilin	Northeastern	50	54	31	50	39	33	36	13	306
Liaoning	Northeastern	41	33	72	68	42	23	25	17	321
Ningxia	Western	12	11	4	1	5	7	4	1	45
Qinghai	Western	13	8	8	4	9	0	1	2	45
Shandong	Eastern	122	206	58	124	146	41	46	25	768
Shanghai	Eastern	54	42	42	33	88	22	28	13	322
Shannxi	Western	30	16	29	39	23	17	20	6	180
Shanxi	Central	31	26	42	33	31	10	27	12	212
Sichuan	Western	51	80	40	49	34	14	19	21	308
Tianjing	Eastern	10	18	0	20	19	8	22	9	106
Tibet	Western	2	9	0	4	1	0	0	0	16
Xinjiang	Western	36	14	15	24	20	2	4	0	115
Yunnan	Western	29	29	27	21	28	10	5	3	152
Zhejiang	Eastern	108	92	39	22	47	2	9	9	328
Overall		1726	1656	1074	1481	1354	510	599	333	8733

Table A.3: Summary Statistics

Variable	Observations	Mean	Std	Min	Max
r_{it}	1377735	9.982	1.424	-0.317	19.135
k_{it}	1377735	8.657	1.591	-0.878	18.389
l_{it}	1377735	4.865	1.133	0.000	12.145
m_{it}	1377735	9.678	1.438	-1.005	18.942
r_{it-1}	1377735	9.863	1.349	-0.190	18.934
k_{it-1}	1377735	8.541	1.639	-0.935	18.389
l_{it-1}	1377735	4.858	1.130	0.000	12.025
m_{it-1}	1377735	9.581	1.372	-0.795	18.801
$\ln s_{it}$	1377735	-0.235	0.436	-12.841	11.645
$\ln s_{it-1}$	1377735	-0.228	0.432	-12.841	10.850
IMR_{it-1}	1377735	0.102	0.074	0.000	0.997
$\mathbb{1}_{i \in SOE,t}$	1377735	0.098	0.297	0.000	1.000
$\mathbb{1}_{i \in SOE,t-1}$	1377735	0.103	0.305	0.000	1.000
$\mathbb{1}_{i \in Exporting,t}$	1377735	0.264	0.441	0.000	1.000
$\mathbb{1}_{i \in Exporting,t-1}$	1377735	0.245	0.430	0.000	1.000
$\mathbb{1}_{Privatized,i,t-1}$	1377735	0.008	0.087	0.000	1.000

Table A.4: Estimation Results For Revenue Function and Productivity Process: Selected Industries

This table presents the structural estimation results for the revenue function for various industries: Food Processing (13), Textile (17), Raw Chemical and Chemical Products (26), Nonmetal Mineral Products (31), and Ordinary Machinery (35). The dependent variables are firm i's log revenue r_{it} at year t. I use the Nonlinear Least Square (NLS) approach outlined in the main text. Robust standard errors are reported in the parenthesis and are clustered at firm level. T tests of statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Industry	13	17	26	31	35	All
Coefficients/Column	(1)	(2)	(3)	(4)	(5)	(6)
θ_k	.0767***	.0688***	.0826***	.0784***	.0765***	.0785***
	(.0000153)	(.0000111)	(.0000202)	(.0000124)	(.0000237)	(1.80e-06)
θ_l	.121***	.13***	.132***	.125***	.139***	.135***
	(.0000339)	(.0000328)	(.0000377)	(.0000267)	(.0000571)	(3.90e-06)
θ_m	.65***	.663***	.634***	.652***	.623***	.652***
	(.000211)	(.000183)	(.00022)	(.000167)	(.000384)	(.0000223)
γ_1	.695***	.744***	.714***	.734***	.782***	.754***
	(.000144)	(.000139)	(.000367)	(.000118)	(.000222)	(.0000127)
γ_2	.0163***	.00657***	.0104***	.00805***	.00484***	.00672***
	(2.82e-06)	(3.80e-06)	(9.36e-06)	(3.85e-06)	(7.39e-06)	(4.52e-07)
γ_3	.00139***	.0021***	.00206***	.0021***	.00206***	.00189***
	(2.01e-08)	(2.12e-08)	(1.51e-08)	(1.81e-08)	(3.31e-08)	(2.06e-09)
ρ^{SOE}	-.12***	-.0838***	-.0796***	-.0742***	-.0774***	-.0925***
	(.0000386)	(.0000236)	(.0000168)	(.0000154)	(.0000263)	(2.59e-06)
$\rho^{Privatized}$	.108***	.0879***	.0935***	.0765***	.0871***	.093***
	(.000109)	(.000094)	(.0000659)	(.000051)	(.000107)	(8.60e-06)
ρ^{After}	-.0338***	-.0285***	-.0305***	-.0263***	-.0344***	-.0287***
	(.0000214)	(.0000147)	(.0000107)	(8.59e-06)	(.0000154)	(1.34e-06)
α	-.824***	-.835***	-.817***	-.855***	-.873***	-.851***
	(.0000815)	(.0000945)	(.000249)	(.0000678)	(.000124)	(6.99e-06)
ϕ	.612***	.627***	.61***	.635***	.628***	.629***
	(.000137)	(.000114)	(.000138)	(.0000905)	(.000231)	(.0000126)
π^{Export}	-.0279***	-.00454***	.00822***	-.00562***	.0097***	-.00105***
	(5.25e-06)	(1.28e-06)	(3.12e-06)	(2.29e-06)	(2.92e-06)	(1.74e-07)
π^{IMR}	-.355***	-.0376***	-.131***	-.0403***	.00366***	-.0814***
	(.00158)	(.000598)	(.00196)	(.000716)	(.0013)	(.0000619)
Number of Observations	87291	118140	103475	123919	102444	1439986
Adjusted R2	.822	.835	.858	.844	.868	.859

Table A.5: Estimation Results For Distortion Process: Selected Industries

This table presents the reduced form estimation results for distortion process for various industries: Food Processing (13), Textile (17), Raw Chemical and Chemical Products (26), Nonmetal Mineral Products (31), and Ordinary Machinery (35). The dependent variables are firm i's log revenue r_{it} at year t. I use the Nonlinear Least Square (NLS) approach outlined in the main text. Robust standard errors are reported in the parenthesis and are clustered at firm level. T tests of statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Industry Coefficients/Column	13 (1)	17 (2)	26 (3)	31 (4)	35 (5)	All (6)
	(1)	(2)	(3)	(4)	(5)	(6)
$\gamma^{\tau,c}\_1$	0.303***	0.318***	0.346***	0.305***	0.338***	0.330***
	(21.46)	(26.70)	(29.12)	(28.75)	(27.19)	(95.82)
$\gamma^{\tau,c}\_2$	0.024***	0.036***	0.018***	0.033***	0.006	0.028***
	(5.27)	(4.90)	(2.91)	(5.93)	(0.75)	(13.11)
$\gamma^{\tau,c}\_3$	-0.001	-0.000	-0.003***	-0.000	-0.002**	-0.001**
	(-1.32)	(-0.09)	(-4.02)	(-0.23)	(-2.33)	(-2.57)
$\gamma^{\tau,l}\_1$	0.122***	0.149***	0.143***	0.126***	0.144***	0.148***
	(8.90)	(13.23)	(14.20)	(11.54)	(15.55)	(50.84)
$\gamma^{\tau,l}\_2$	0.004	0.002	0.003	-0.003	0.001	0.001
	(1.39)	(0.34)	(0.58)	(-0.71)	(0.25)	(0.93)
$\gamma^{\tau,l}\_3$	-0.002***	-0.002***	-0.001**	-0.002***	-0.001**	-0.002***
	(-4.22)	(-3.08)	(-2.49)	(-3.22)	(-2.27)	(-7.84)
$\rho^{\tau,SOE}\_P$	0.038***	0.060***	0.031***	0.022***	0.032***	0.036***
	(4.57)	(5.54)	(5.41)	(3.68)	(4.91)	(16.08)
$\rho^{\tau,SOE}\_N$	0.024	-0.029	-0.018	-0.011	-0.007	-0.041***
	(0.94)	(-0.76)	(-0.72)	(-0.67)	(-0.38)	(-8.54)
$\rho^{\tau,Privatized}\_P$	0.028	-0.033	-0.035*	-0.032*	-0.034	-0.017**
	(0.93)	(-1.24)	(-1.85)	(-1.94)	(-1.44)	(-2.35)
$\rho^{\tau,Privatized}\_N$	0.122*	-0.045	0.008	0.039	-0.009	0.082***
	(1.86)	(-0.49)	(0.19)	(0.75)	(-0.14)	(4.77)
$\rho^{\tau,After}\_P$	0.002	0.012	0.023***	0.023***	0.002	0.014***
	(0.21)	(1.42)	(3.34)	(3.30)	(0.24)	(5.14)
$\rho^{\tau,After}\_N$	-0.040	-0.014	0.008	-0.009	-0.035	-0.036***
	(-1.60)	(-0.53)	(0.36)	(-0.48)	(-1.01)	(-4.59)
ϕ^{τ}	-0.001	0.013	0.011	0.006	-0.013*	0.015***
	(-0.10)	(1.36)	(1.19)	(0.80)	(-1.84)	(6.78)
$\pi^{\tau,Export}$	0.005	-0.015***	-0.007*	-0.019***	-0.017***	-0.011***
	(1.13)	(-5.39)	(-1.90)	(-5.36)	(-4.96)	(-11.29)
$\pi^{\tau,IMR}$	0.115*	0.033	0.069	0.084**	-0.026	0.082***
	(1.83)	(0.78)	(1.14)	(2.06)	(-0.60)	(7.36)
Constant	0.177***	0.107***	0.111***	0.118***	0.176***	0.104***
	(5.17)	(4.26)	(3.95)	(5.53)	(7.84)	(14.87)
Observation	60,724	83,172	73,868	90,161	70,242	960,433
Adjusted R2	0.157	0.183	0.150	0.159	0.168	0.198

TABLE A.6: Correlation between Distortions and Firm Characteristics

This table provides correlation between estimated log distortion and firm characteristics. $\mathbb{1}_{i,t}^{Exit}$ is a dummy variable equal to 1 if it is the last observation for firm i in the sample, and missing if $t = 2007$, and 0 otherwise. $\mathbb{1}_{i,t}^{State-Owned}$ is a dummy variable equal to 1 if firm i at year t's registration type is either 110 or 151, and 0 otherwise. $\mathbb{1}_{it}^{Exporting}$ is a dummy variable equal to 1 if firm i reports positive exporting value at year t, and 0 otherwise. $\mathbb{1}_{it}^{Subsidized}$ is a dummy variable equal to 1 if firm i reports positive subsidy received at year t, and 0 otherwise. All regressions include capital, age, and age square. Robust standard error are clustered both at firm and year level. T tests of statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dependent Variable: Log Distortions					
	(1)	(2)	(3)	(4)	(5)
$\mathbb{1}_{it}^{Exit}$		0.025*			
		(1.84)			
$\mathbb{1}_{it}^{SOE}$			0.017**		
			(2.05)		
$\mathbb{1}_{it}^{Exporting}$				-0.018***	
				(-3.87)	
$\mathbb{1}_{it}^{Subsidized}$					0.003*
					(1.92)
Constant	0.338***	0.330***	0.335***	0.334***	0.338***
	(23.64)	(24.03)	(22.76)	(23.98)	(23.63)
Covariates	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes
Observation	1,377,735	1,135,835	1,377,735	1,377,735	1,377,735
R^2	0.049	0.047	0.049	0.049	0.049
Dependent Variable: Absolute value of Log Distortions					
	(1)	(2)	(3)	(4)	(5)
$\mathbb{1}_{it}^{Exit}$		0.090***			
		(6.65)			
$\mathbb{1}_{it}^{SOE}$			0.143***		
			(18.10)		
$\mathbb{1}_{it}^{Exporting}$				-0.034***	
				(-2.65)	
$\mathbb{1}_{it}^{Subsidized}$					-0.001
					(-0.57)
Constant	0.408***	0.398***	0.367***	0.416***	0.408***
	(42.47)	(38.68)	(52.57)	(34.43)	(42.44)
Covariates	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes
Observation	1,377,735	1,135,835	1,377,735	1,377,735	1,377,735
R^2	0.032	0.034	0.043	0.033	0.032

Table A.7: Aggregate Productivity Growth Decomposition:1998-2007

This table presents the decomposition results for aggregate productivity growth between 1999 to 2007 using the estimated sample within the manufacturing sector.

Aggregate Productivity Growth	58.037
State	
Mean Productivity Growth	43.998
Reallocation Within State	1.916
Private	
Mean Productivity Growth	37.179
Reallocation Within Private	9.115
Between Ownership	-2.266
Transition	3.934
Entry	-21.473
Exit	27.598

Table A.8: Dispersion of MRPK, Volatility and Distortions

This table shows that there are increasing relationship between industrial-year level of dispersion in marginal revenue product of capital (MRPK) and dispersion of productivity innovation. Dependent variable is $Std_{jt}(MRPK_{it})$, the standard deviation of firm-level MRPK at industry year level. $Std_{jt}(\eta_{it})$ is the standard deviation of firm level log productivity innovation at industry year level. Column (1) and (3) present unweighted regression results. Column (2) and (4) present weighted regression results that observations for industry j at year t are weighted by their revenue shares of total revenue in year t. Robust standard error are clustered at industry level. T tests of statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	$Std_{jt}(MRPK)$			
	(1)	(2)	(3)	(4)
$Std_{jt}(\tau_{it})$	0.364***	0.351***		
	(3.73)	(3.44)		
$Std_{jt}(\eta_{it})$			0.765***	0.844***
			(4.99)	(4.94)
Constant	1.338***	1.311***	1.157***	1.089***
	(21.35)	(19.36)	(13.57)	(11.54)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Weighted	No	Yes	No	Yes
Observation	252	252	252	252
R^2	0.883	0.885	0.896	0.903

Table A.9: Propensity Score Matching: Sample

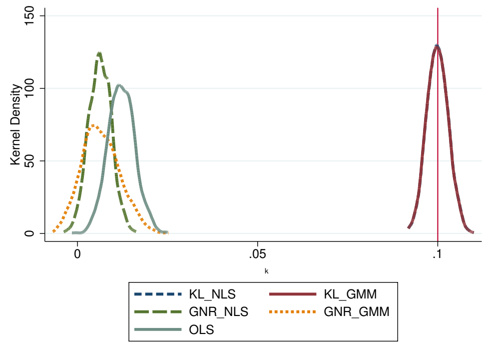
This table presents matching results. I matched the privatized firms i which is privatized at year $t-1$ with SOEs that are not privatized on year $t-1$ or t from the same industry, region and exporting status, and with similar employment, productivity and distortion parameters. T tests of statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Before Privatization	Privatized SOEs (A)	Non-Privatized SOEs (B)	Difference (B)-(A)	Test Statistics
ω_{it-1}	2.446	2.442	-0.004	-0.210
	(0.008)	(0.009)	(0.012)	
$\lvert\tau_{it-1}\rvert$	0.456	0.458	0.002	0.299
	(0.005)	(0.006)	(0.008)	
$InverseMillsRatio$	0.121	0.120	-0.001	-0.402
	(0.001)	(0.001)	(0.001)	
$\ln(Employment_{it-1})$	5.532	5.532	-0.001	-0.022
	(0.014)	(0.016)	(0.022)	
After Privatization				
ω_{it}	2.493	2.453	0.040***	-3.177
	(0.009)	(0.009)	(0.013)	
τ_{it}	0.445	0.465	0.019***	2.320
	(0.006)	(0.006)	(0.008)	

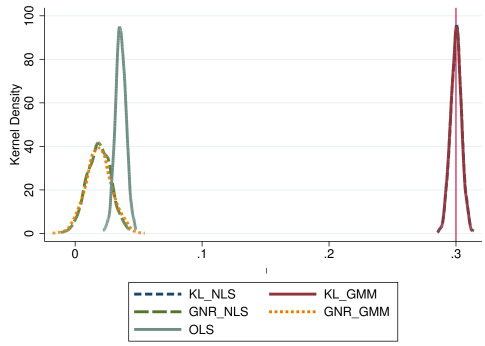
A.5 Figures

FIGURE A.1: Simulation Results for Technology Parameters

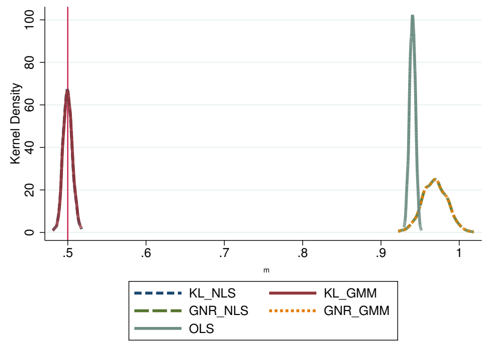
This set of figures presents the simulation results for estimated technology parameters as well as their asymptotic distributions. The true parameters are the redline within each graph. OLS denotes the method of directly regressing output on inputs. GNR_{NLS} denotes the method proposed by Gandhi et al. [2012] and is implemented using nonlinear least square, and GNR_{GMM} denotes the same method but is implemented using general method of moments. KL_{NLS} and KL_{GMM} denote the method proposed in section 1.3 via NLS and GMM.



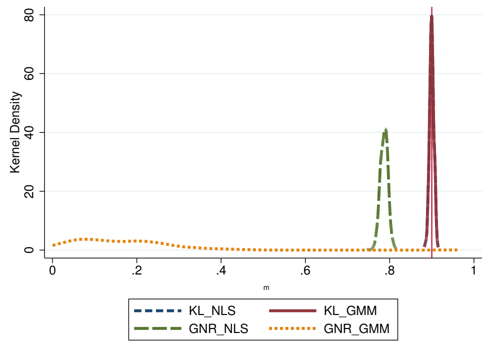
(a) Estimates of θ_k



(b) Estimates of θ_l



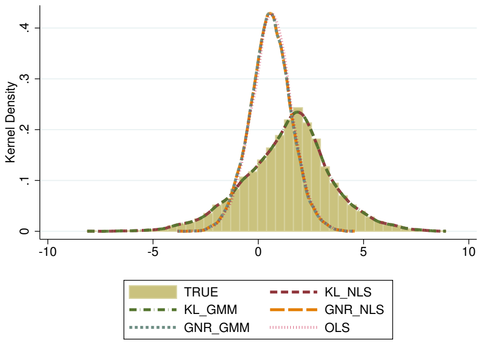
(c) Estimates of θ_m



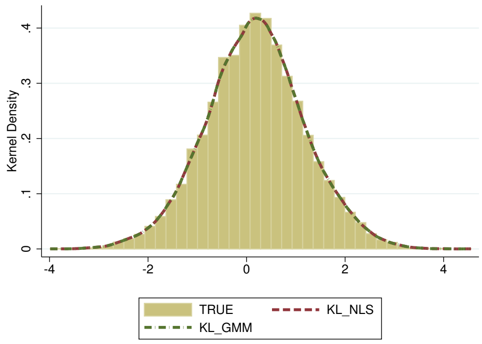
(d) Estimates of γ

FIGURE A.2: Simulation Results for Productivity and Distortions

This set of figures presents the simulation results for estimated productivity and distortions. OLS denotes the method of directly regressing output on inputs. GNR_{NLS} denotes the method proposed by Gandhi et al. [2012] and is implemented using nonlinear least square, and GNR_{GMM} denotes the same method but is implemented using general method of moments. KL_{NLS} and KL_{GMM} denote the method proposed in section 1.3 via NLS and GMM.



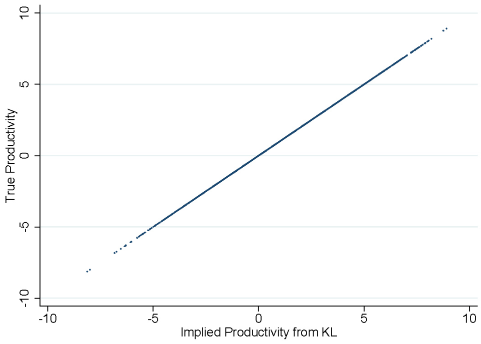
(a) Implied ω_{it}



(b) Implied τ_{it}

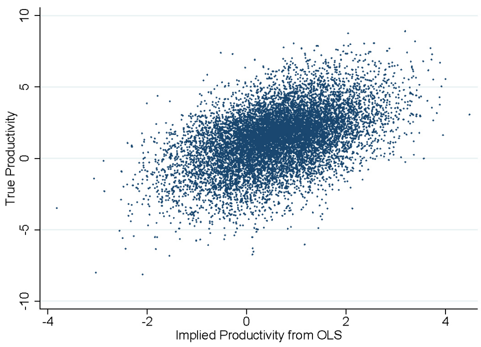
FIGURE A.3: Simulation Results for Correlation between Productivity Measures

This set of figures presents the simulation results for the correlation between estimated productivity and the true productivity. OLS denotes the method of directly regressing output on inputs. GNR_{NLS} denotes the method proposed by Gandhi et al. [2012] and is implemented using nonlinear least square, and GNR_{GMM} denotes the same method but is implemented using general method of moments. KL_{NLS} and KL_{GMM} denote the method proposed in section 1.3 via NLS and GMM.

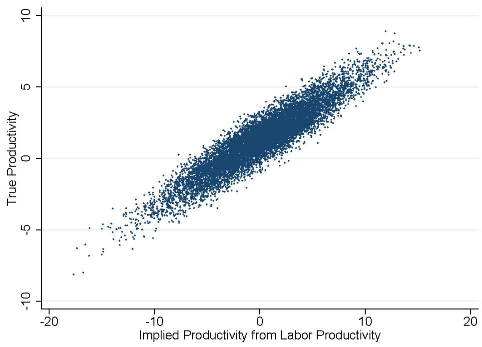


(a) Correlation between of ω_{it}^{KL} and ω_{it}

(b) Correlation between of ω_{it}^{GNR} and ω_{it}



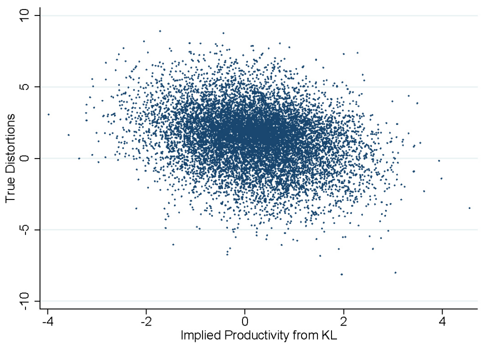
(c) Correlation between of ω_{it}^{OLS} and ω_{it}



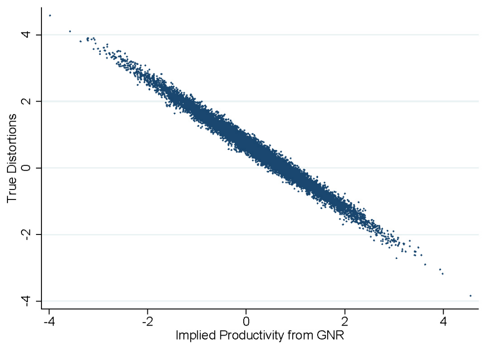
(d) Correlation between of ω_{it}^{Labor} and ω_{it}

FIGURE A.4: Simulation Results for Correlation between Productivity Measures and Distortions

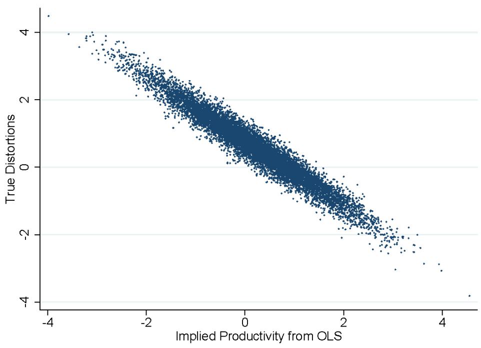
This set of figures presents the simulation results for the correlation between the estimated productivity and true distortion distribution. OLS denotes the method of directly regressing output on inputs. GNR_{NLS} denotes the method proposed by Gandhi et al. [2012] and is implemented using nonlinear least square, and GNR_{GMM} denotes the same method but is implemented using general method of moments. KL_{NLS} and KL_{GMM} denote the method proposed in section 1.3 via NLS and GMM.



(a) Correlation between of ω_{it}^{KL} and τ_{it}



(b) Correlation between of ω_{it}^{GNR} and τ_{it}



(c) Correlation between of ω_{it}^{OLS} and τ_{it}

(d) Correlation between of ω_{it}^{Labor} and τ_{it}

Correlation between of ω_{it} and τ_{it}

This set of figures presents the correlation between the estimated productivity and the estimated distortion across two digit industries, using the method proposed in section 1.3 via NLS.

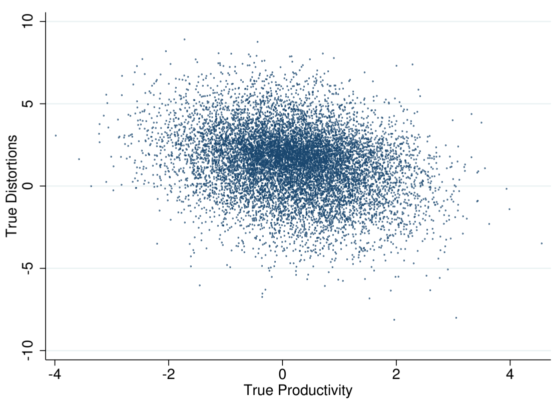


Figure A.5: Four Economic Regions of China

This figure depicts the four economic regions in China according to The Communist Party of China [2006].

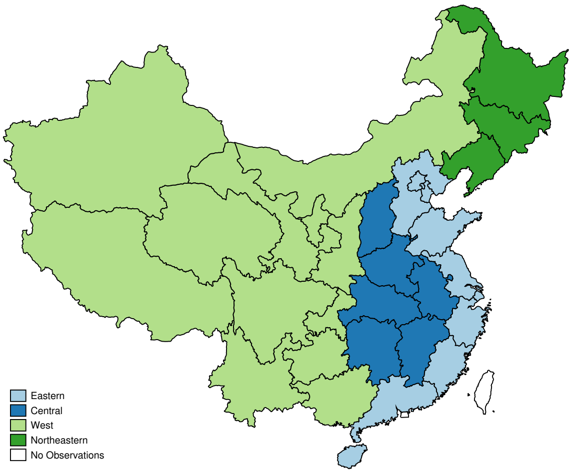


FIGURE A.6: Productivity by Ownership and Regions

These figures depict the estimated distribution of productivity (adjusted by industry-year mean) by ownership and economic regions using the method proposed in section 1.3.

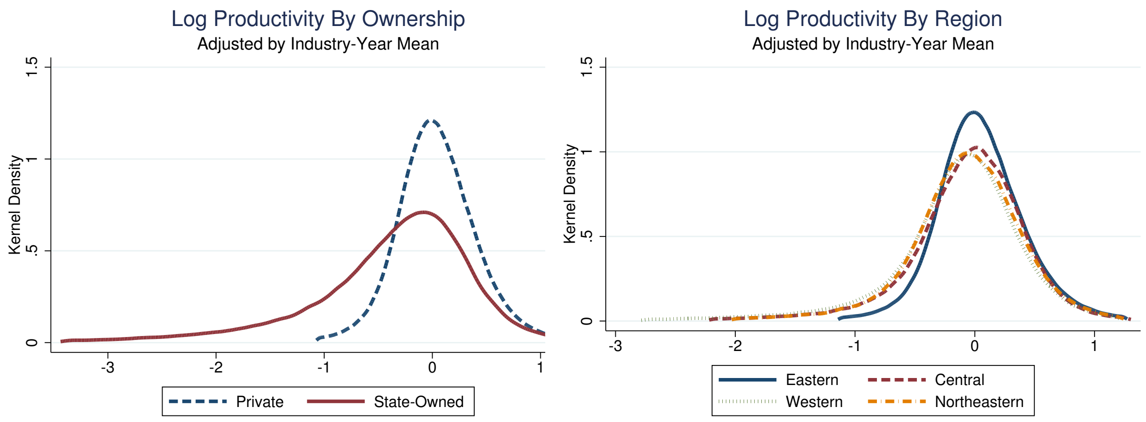


FIGURE A.7: Distortions by Ownership and Regions

These figures depict the estimated distribution of distortions by ownership and economic regions using the method proposed in section 1.3.

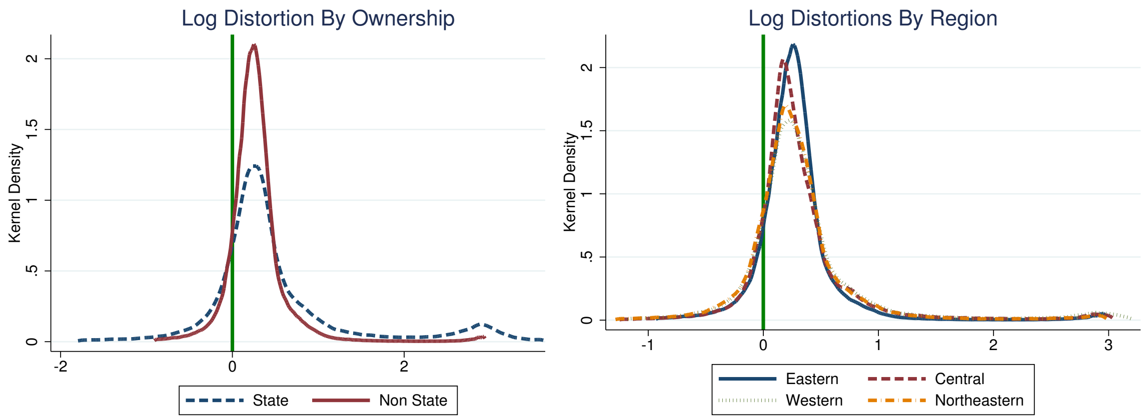


Figure A.8: Correlation Between Distortions and Productivity

This figure depicts the correlation between the estimated distribution of productivity and distortions across industries using the method proposed in section 1.3.

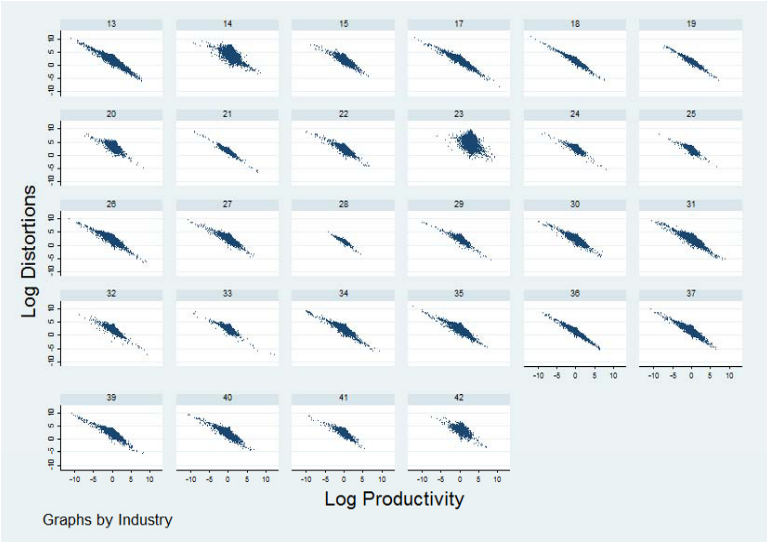


FIGURE A.9: Dispersion of MRPK and Dispersion of Distortions and Productivity

These figures depict the relationship between industry-year level dispersion of log marginal revenue product of capital and the dispersion of productivity innovation and distortion dispersion. The productivity and distortions are estimated following the method proposed in section 1.3. The size of the ball represents the revenue share for a particular industry-year's contribution to the total revenue for that year. The red line represents the unweighted fitting line, and the dashed line represents the weighted fitting line.

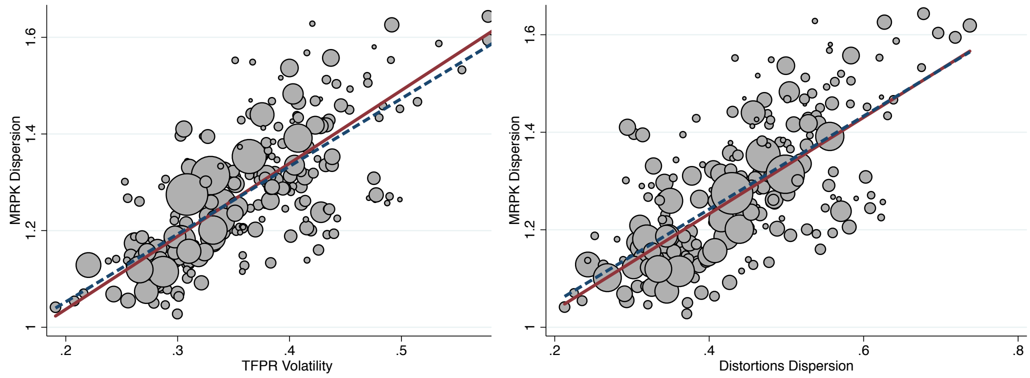


FIGURE A.10: Dispersion of Distortions and Productivity

This figure depicts the relationship between industry-year level dispersion of productivity innovation and distortion dispersion. The productivity and distortions are estimated following the method proposed in section 1.3. The size of the ball represents the revenue share for a particular industry-year's contribution to the total revenue for that year. The red line represents the unweighted fitting line, and the dashed line represents the weighted fitting line.

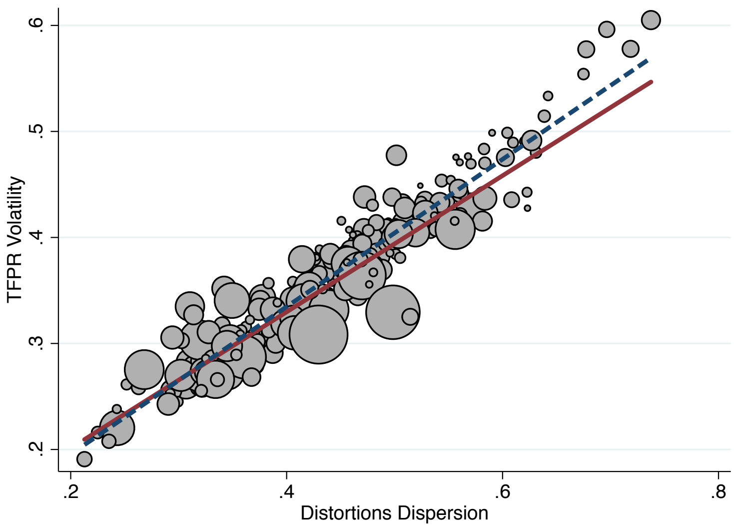
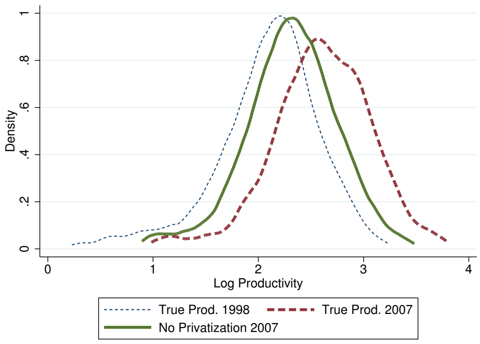
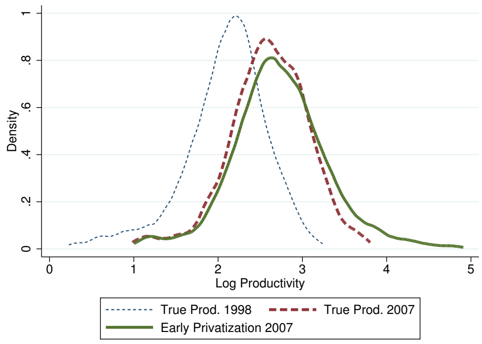


FIGURE A.11: Decomposition of the Productivity and Distortion Processes

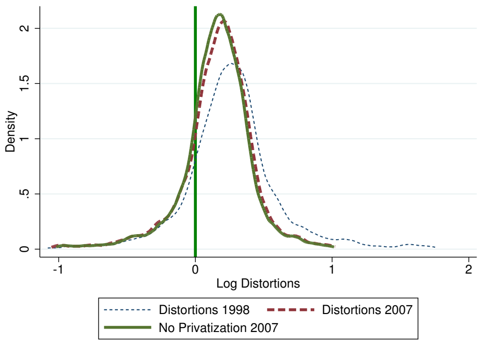
This set of figures presents the decomposition of the productivity and distortion processes with privatization. I take a balanced sample of SOEs firms which eventually are privatized within the sample period and conduct counterfactual experiments on their productivity process. The dashed line is the productivity distribution in 1998. And the short-dashed line is the productivity distribution in year 2007. The counterfactual distributions are depicted using solid line in each figure. The top left panel shows the counterfactual distortion distribution when there is no privatization. The top right panel shows the counterfactual distortion distribution if privatization occurs at 1998. The bottom-left panel shows the counterfactual productivity distribution by removing the transitory components of privatization from the productivity process. The bottom-right panel shows the counterfactual productivity distribution by removing the permanent components of privatization from the productivity process. The simulated distortion process using realized distortion innovations and assume other covariates are exogenous.



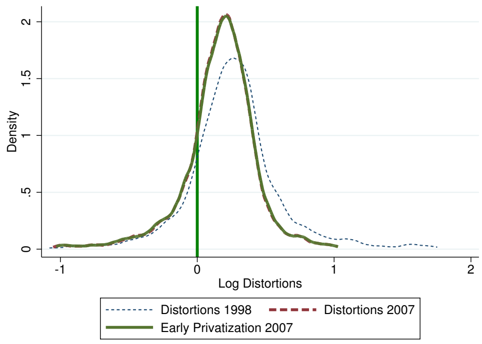
(a) No Privatization: Productivity



(b) Early Privatization: Productivity



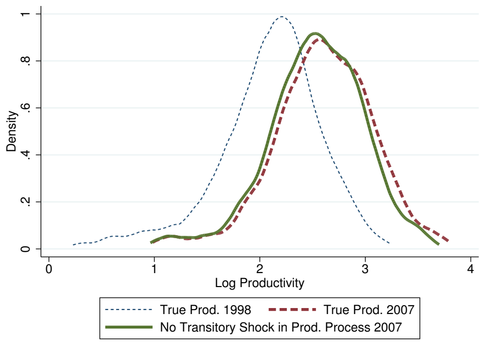
(c) No Privatization: Distortions



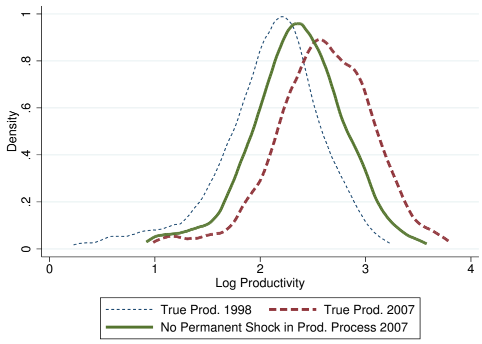
(d) Early Privatization: Distortions

FIGURE A.12: Decomposition of the Productivity Process

This set of figures presents the decomposition of the productivity process with privatization. I take a balanced sample of SOEs firms which eventually are privatized within the sample period and conduct counterfactual experiments on their productivity process. The dashed line is the productivity distribution in 1998. And the short-dashed line is the productivity distribution in year 2007. The counterfactual distributions are depicted using solid line in each figure. The top left panel shows the counterfactual distortion distribution when there is no privatization while fixing distortions. The top right panel shows the counterfactual distortion distribution if privatization occurs at 1998 while fixing distortions. The bottom-left panel shows the counterfactual productivity distribution by removing the transitory components of privatization from the productivity process while fixing distortions. The bottom-right panel shows the counterfactual productivity distribution by removing the permanent components of privatization from the productivity process while fixing distortions. The simulated distortion process using realized distortion innovations and assume other covariates, including distortions are exogenous.



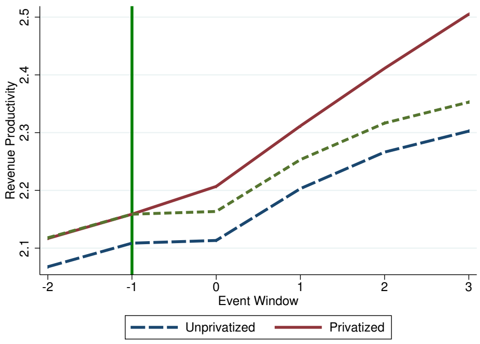
(a) No Transitory Shock in Productivity Process



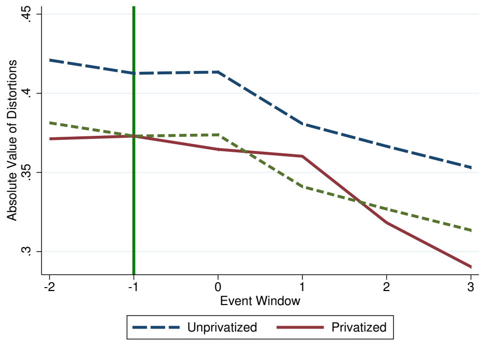
(b) No Permanent Shock in Productivity Process

Figure A.13: Matched Sample Productivity and Distortions Around Privatization Events

This set of figures presents mean of productivity and absolute distortions around the privatization window. It shows that productivity increases over time for privatized firm comparing to non privatized SOE firms. Absolute distortions decrease over time as well, but not differ significantly between the two groups.



(a) Productivity



(b) Absolute Distortions

Appendix B

Appendix for Chapter 2

B.1 (Proof of Theorem 2.1)

Proof. Fix a continuous non-decreasing rule $\omega : Y \to W$. Notice that ω is not required in this definition to have range contained in $\overline{G}$, so the proper interpretation is that $\omega(y)$ is the wage that yields the worker his market payoff if he is hired for sure at that wage. If all searching workers apply to all wages at or above their reservation wage, then

$$P(w, y) = \int_{\underline{y}}^{\min\left[\omega^{-1}(w), y\right]} \frac{G(w) - G(\omega(y'))}{G(\overline{w}) - G(\omega(y'))} dF(y').$$

The 'queue size' $p_w(y)$ has to satisfy (2.1), so

$$p_w(y) = \int_{\underline{y}}^{\min\left[\omega^{-1}(w), y\right]} \frac{1}{G(\overline{w}) - G(\omega(y'))} dF(y').$$

To see this observe that for any w,

$$\int_{\underline{w}}^{w} p_{\tilde{w}}(y)\, dG(w) = \int_{\underline{w}}^{w} \int_{\underline{y}}^{\min\left[\omega^{-1}(\tilde{w}), y\right]} \frac{1}{G(\overline{w}) - G(\omega(y'))} dF(y')\, dG(\tilde{w})$$

$$= \int_{\underline{y}}^{\min\left[\omega^{-1}(w), y\right]} \int_{\omega(y')}^{w} \frac{dG(\tilde{w})}{G(\overline{w}) - G(\omega(y'))} dF(y') =$$

$$\int_{\underline{y}}^{\min\left[\omega^{-1}(w), y\right]} \frac{G(w) - G(\omega(y'))}{G(\overline{w}) - G(\omega(y'))} dF(y').$$

This implies that

$$\int_{y}^{\overline{y}} dp_w(\tilde{y}) = \int_{y}^{\omega^{-1}(w)} \frac{1}{G(\overline{w}) - G(\omega(y'))} dF(y'). \tag{B.1}$$

So hiring probabilities will be given by (2.7) provided that workers all use the application strategy described. Given this matching probability we can now describe the condition that $\omega(y)$ has to satisfy in order for them to be willing to follow this strategy. In order for a searching worker of type $\omega(y) > \underline{w}$ to be indifferent between all wages above his reservation wage, it should be that for each $w' > \omega(y)$

$$\left(w' + \gamma U(y)\right) e^{-\int_y^{\overline{y}} dp_{w'}(\tilde{y})} + \left(1 - e^{-\int_y^{\overline{y}} dp_{w'}(\tilde{y})}\right) \gamma U(y)$$

$$= \omega(y) + \gamma U(y),$$

or

$$w' e^{-\int_y^{\overline{y}} dp_{w'}(\tilde{y})} = \omega(y).$$

Taking logs yields

$$\int_y^{\overline{y}} dp_{w'}(\tilde{y}) = \log\left(w'\right) - \log\left(\omega(y)\right).$$

By the fundamental theorem of calculus this implies

$$\int_{\omega(y)}^{w'} \frac{1}{\tilde{w}} d\tilde{w} = \int_y^{\overline{y}} dp_{w'}(\tilde{y}). \tag{B.2}$$

Substituting (B.1), then gives the identity

$$\int_{\omega(y)}^{w} \frac{1}{\tilde{w}} d\tilde{w} = \int_y^{\omega^{-1}(w)} \frac{1}{G(\overline{w}) - G(\omega(y'))} d\mathcal{F}\left(y'\right)$$

is satisfied for all y. Differentiating both sides with respect to w gives the differential equation

$$\omega'(y) = \frac{\omega(y) F'(y)}{G(\overline{w}) - G(\omega(y))}. \tag{B.3}$$

The reservation wage function ω will support the continuation equilibrium if it has a solution with $\omega(\overline{y}) = \overline{w}$. This is not immediate since the right hand side does not have a continuous derivative around the point $(\overline{y}, \overline{w})$.

However it does have a solution through the point $(\overline{y}, \overline{w} - \epsilon)$ for any $\epsilon > 0$. Denote the solution for $\epsilon > 0$ as $\omega^\epsilon(y)$. Observe that each ω^ϵ is strictly increasing and that ω^ϵ and $\omega^{\epsilon'}$ cannot cross, therefore the sequence $\{\omega^\epsilon\}_{\epsilon \to 0}$ is an increasing sequence of increasing functions. As the sequence $\omega^\epsilon(y)$ is a bounded increasing sequence of real numbers, ω^ϵ converges point-wise, therefore uniformly (Dini's Theorem) to some function ω. If (B.3) fails at some point y, then by uniform convergence, it must fail for small ϵ. So ω is a solution to (B.3).

The remaining bits of the theorem then follow by using (2.2) along with the reservation wage. $\square$

Analyzing the firms' problem

Though we pay little attention to the firms' problem in what follows, we can use the previous results to describe what firms do. Readers who are only interested in the implications for wage data can skip this section.

Using Theorem 2.1, we get the following characterization:

Lemma B.1. *In a symmetric steady state equilibrium, the firm's payoff function can be written as*

$$\tilde{V}(w) = \int_{\underline{y}}^{\omega^{-1}(w)} \left\{ \frac{\frac{v(w,x,y)}{w}\omega'(y)}{\gamma + (1-\gamma)\frac{\omega'(y)}{F'(y)}\int_w^{\overline{w}} \frac{dG(\tilde{w})}{\tilde{w}}} \right\} dy. \tag{B.4}$$

Proof. Substituting this into $\tilde{V}(w)$ and using (2.9) and Lemma 2.4, we can write the firm's payoff function as

$$\tilde{V}(w) = \int_{\underline{y}}^{\omega^{-1}(w)} \frac{v(w,x,y)e^{-\int_y^{\overline{y}} dp_w(\tilde{y})}}{Q(w,y)} dp_w(y) =$$

$$\int_{\underline{y}}^{\omega^{-1}(w)} \frac{v(w,x,y)\frac{\omega(y)}{w}}{Q(w,y)} \frac{\omega'(y)}{\omega(y)} dy =$$

$$\int_{\underline{y}}^{\omega^{-1}(w)} \left\{ \frac{\frac{v(w,x,y)}{w}\omega'(y)}{\gamma + (1-\gamma)\frac{\omega'(y)}{F'(y)}\int_w^{\overline{w}} \frac{dG(\tilde{w})}{\tilde{w}}} \right\} dy.$$

□

One way to view the formula in (B.4) is that the firm trades off the wage it pays against the highest quality worker who applies. With this interpretation the firm's maximization problem could be expressed as

$$\max_{w,y} \int_{\underline{y}}^{y} \left\{ \frac{\frac{v(w,x,\tilde{y})}{w}\omega'(\tilde{y})}{\gamma + (1-\gamma)\frac{\omega'(\tilde{y})}{F'(\tilde{y})}\int_w^{\overline{w}} \frac{dG(\tilde{w})}{\tilde{w}}} \right\} d\tilde{y}$$

subject to the constraint that $\omega(y) = w$.

This is a pretty standard directed search problem. The formula above is somewhat complex, but it illustrates a fundamental identification problem. Fix the (steady state) wage offer distribution G. Theorem 2.1 ensures the existence of a reservation wage strategy ω. Define

$$\phi(w,x,y) = \frac{\frac{v(w,x,\tilde{y})}{w}\omega'(\tilde{y})}{\gamma + (1-\gamma)\frac{\omega'(\tilde{y})}{F'(\tilde{y})}\int_w^{\overline{w}} \frac{dG(\tilde{w})}{\tilde{w}}}.$$

Now suppose that we change the distribution F so that it is uniform. Again using Theorem 2.1, there will be a new reservation wage rule, say $\tilde{w}(y)$. The equation

$$\frac{\frac{\tilde{v}}{w}\tilde{\omega}'(\tilde{y})}{\gamma + (1-\gamma)\tilde{\omega}'(\tilde{y})\int_w^{\overline{w}} \frac{dG(\tilde{w})}{\tilde{w}}} = \phi(w,x,y)$$

with unknown $\tilde{v}$ has a positive solution for each pair (w,y) given by

$$\tilde{v}(w,x,y) = \frac{w\phi(w,x,y)\left(\gamma + (1-\gamma)\tilde{\omega}'(\tilde{y})\int_w^{\overline{w}} \frac{dG(\tilde{w})}{\tilde{w}}\right)}{\tilde{\omega}'(y)}.$$

This means that when F is replaced with a uniform distribution function and the profit function is replace by $\tilde{v}$, the expected profit function for every firm type x is uniformly the same as the old one. As a consequence, the distribution of best replies G will remain unchanged.

As the result in the previous section shows, we might as well assume from now on that the distribution of worker types is uniform while imagining that the profit function describes the profit to the firm association with hiring workers at different quantiles of the type distribution.

B.2 Proof of Theorem 2.7

Proof. The steady state condition is

$$G(w) = \int_{\underline{w}}^{w} \frac{\int_{\underline{y}}^{\overline{y}} Q(\tilde{w}, y)\, d\psi(y|\tilde{w})}{\int_{\underline{y}}^{\overline{y}} e^{-\int_{x}^{\overline{y}} dp_{\tilde{w}}(\tilde{y})} dp_{\tilde{w}}(x)} dE(\tilde{w}).$$

By (2.2)

$$G(w) = \int_{\underline{w}}^{w} \frac{\int_{\underline{y}}^{\overline{y}} \left(\gamma + (1-\gamma)\, \omega'(y) \int_{\tilde{w}}^{\overline{w}} \frac{dG(w')}{w'}\right) d\psi(y|\tilde{w})}{\int_{\underline{y}}^{\omega^{-1}(\tilde{w})} e^{-\int_{x}^{\overline{y}} dp_{\tilde{w}}(\tilde{y})} dp_{\tilde{w}}(x)} dE(\tilde{w}).$$

By (2.1) and (2.4), the right hand side of this equation equals

$$\int_{\underline{w}}^{w} \frac{\int_{\underline{y}}^{\overline{y}} \left(\gamma + (1-\gamma)\, \omega'(y) \int_{\tilde{w}}^{\overline{w}} \frac{dG(w')}{w'}\right) d\psi(y|\tilde{w})}{\int_{\underline{y}}^{\omega^{-1}(\tilde{w})} \frac{\omega(y)}{\tilde{w}} \frac{\omega'(y)}{\omega(y)} dy} dE(\tilde{w}).$$

Simplifying the denominator gives

$$\int_{\underline{w}}^{w} \frac{\int_{\underline{y}}^{\omega^{-1}(\tilde{w})} \left(\gamma + (1-\gamma)\, \omega'(y) \int_{\tilde{w}}^{\overline{w}} \frac{dG(w')}{w'}\right) d\psi(y|\tilde{w})}{\frac{1}{\tilde{w}}(\tilde{w} - \underline{w})} dE(\tilde{w})$$

By (2.5), this is equal to

$$\int_{\underline{w}}^{w} \frac{\int_{\underline{y}}^{\omega^{-1}(\tilde{w})} \left(\gamma + (1-\gamma)\, \omega'(y) \int_{\tilde{w}}^{\overline{w}} \frac{dG(w')}{w'}\right) \frac{\omega'(y)}{(\tilde{w} - \underline{w})} dy}{\frac{1}{\tilde{w}} \int_{\underline{y}}^{\omega^{-1}(\tilde{w})} \omega'(y)\, dy} dE(\tilde{w})$$

Changing variable in the integration then gives the equality

$$\int_{\underline{w}}^{w} \frac{\int_{\underline{w}}^{\tilde{w}} \left(\gamma + (1-\gamma) \int_{\tilde{w}}^{\overline{w}} \frac{dG(w'')}{w''}\right) \frac{1}{(\tilde{w} - \underline{w})} dw'}{\frac{1}{\tilde{w}} \int_{\underline{y}}^{\omega^{-1}(\tilde{w})} \omega'(y)\, dy} dE(\tilde{w}).$$

$$G(w) = \int_{\underline{w}}^{w} \frac{\left(\gamma + (1-\gamma) \int_{\tilde{w}}^{\overline{w}} \frac{dG(w')}{w'}\right)}{\frac{1}{\tilde{w}}(\tilde{w} - \underline{w})} dE(\tilde{w})$$

$$\int_{\underline{w}}^{w}\left(\gamma+(1-\gamma)\int_{\tilde{w}}^{\overline{w}}\frac{dG\left(w'\right)}{w'}\right)\frac{\tilde{w}}{\left(\tilde{w}-\underline{w}\right)}dE\left(\tilde{w}\right).$$

This gives the fixed point stated in the Theorem. □

B.3 Algorithm of Defining Local Labor Markets

In this section, we explain how we partition occupational codes and geographic units into categories as of local labour markets. The technique we adopted is known as the community detection in statistics and computer science literature.

In graph theory, a graph is a representation of a set of vertices, and some pairs of vertices are connected by edges. In our application of labour data, each vertex is a pair of occupation (PCS-ECE) and geographic unit (Zone D'emploi). We consider only undirected graph, which means we do not trace the direction of transitions. Therefore, in our case, the edge is represented by the total number of transitions between the two vertices, (or to say, pairs of occupation-geographic units), which in the language of graph theory, naturally makes the graph weighted.[1]

The objective of partitioning graph is modularity, which measures the density of links within communities relative to the links between communities. Specifically, for a weighted graph, the modularity is defined as

$$Q=\frac{1}{2m}\sum_{i,j}\left[A_{i,j}-\frac{k_i k_j}{2m}\right]\delta(c_i,c_j)$$

where $A_{i,j}$ represents the weight of the edge between i and j. $k_i=\sum_j A_{i,j}$ is the sum of the weights of the edges attached to vertex i. c_i is the community to which vertex i is assigned. $\delta(u,v)=1$ if $u=v$ and 0 otherwise. $m=0.5\times\sum_{i,j}A_{i,j}$.

The task of partitioning graph amounts to searching for community assignment that maximizes value of modularity Q. A direct computation for the maximum of modularity however is a NP-Hard problem. We therefore adopted a more feasible approach, the "Louvain Method" proposed by Blondel et al. [2008], which is widely used for its fast speed and robust features.

There are two main phases to partition a weighted graph with N vertices. In the first phase, we experiment with different assignment of each vertex to another community. For each vertex i, we consider any other vertex $j\neq i$ and evaluate the gain of modularity by moving i from its current community to the one that j belongs. The vertex i is placed in j's community if the gain is maximum and positive. Otherwise, i stays in its current community. The experiment stops when local maxima of the modularity is attained, i.e., no individual move can further improve the modularity. In the second phase, we consolidate the identified communities from phase one into new vertices. The weights of the edges between the new vertices are then derived by adding the original weights of the edges from the corresponding communities.

We iterate the two-phase procedure until no detection of any further changes. Thus, a maximum of modularity is reached. In the end, we obtain a set of communities driven by

[1] In contrast, unweighted graph emphasizes the extensive margin of edges, ie, the existence of link.

transition data. These communities reflect the connection between cells of occupation-geographic unit. If there are sufficient mobility between any pair of these cells, we then treat them as a homogeneous community. In our context, we regard these communities as local labour markets.

B.3.1 Derivation of how to compute $h(w)$

Recall (2.18) states

$$-h'(w) = (\gamma + (1-\gamma)h(w))\frac{e(w)}{w - \underline{w}}$$

Rearrange terms we have

$$-\frac{h'(w)(w-\underline{w})}{e(w)} = \gamma + (1-\gamma)h(w)$$

We can always define $h(w) = \tilde{h}(w) + \frac{\gamma}{\gamma-1}$. We can then have the following.

$$\begin{aligned}
-\frac{\tilde{h}'(w)(w-\underline{w})}{e(w)} &= (1-\gamma)\tilde{h}(w) \\
\frac{\tilde{h}'(w)}{\tilde{h}(w)} &= -\frac{e(w)(1-\gamma)}{(w-\underline{w})} \\
\left(\ln \tilde{h}(w)\right)' &= -\frac{e(w)(1-\gamma)}{(w-\underline{w})}
\end{aligned}$$

Integrating both sides, we get

$$\begin{aligned}
\ln \tilde{h}(w) &= \int_w^{\overline{w}} \frac{e(\tilde{w})(1-\gamma)}{(\tilde{w}-\underline{w})} d\tilde{w} + \tilde{c} \\
\tilde{h}(w) &= \exp\left((1-\gamma)\int_w^{\overline{w}} \frac{e(\tilde{w})}{(\tilde{w}-\underline{w})} d\tilde{w} + \tilde{c}\right)
\end{aligned}$$

and therefore,

$$\begin{aligned}
h(w) &= \tilde{h}(w) + \frac{\gamma}{\gamma-1} \\
&= \exp\left((1-\gamma)\int_w^{\overline{w}} \frac{e(\tilde{w})}{(\tilde{w}-\underline{w})} d\tilde{w} + \tilde{c}\right) + \frac{\gamma}{\gamma-1}
\end{aligned}$$

with $\tilde{c}$ can be pin down by the boundary condition.

By definition of $h(w)$

$$h(w) \equiv \int_w^{\overline{w}} \frac{dG(w')}{w'}$$

we know that $h(\overline{w}) = 0$. Therefore $h(\overline{w}) = \exp(\tilde{c}) + \frac{\gamma}{\gamma-1} = 0$ which in turn implies

$$\tilde{c} = \log\left(\frac{\gamma}{1-\gamma}\right)$$

Substituting these and rearranging terms, we therefore get (2.19).

Appendix C

Appendix for Chapter 3

Tables

Table C.1: Variable Dictionary

Variable	Description
Panel A: School Level	
i	School.
j	Journal.
t	Year.
$Publication_{i,j,t}$	Number of publications from school i on journal j at year t.
$\overline{Publication}_{i,j,t}$	Average number of publications from school i on journals $j' \neq j$ at year t.
$During_{i,j,t}$	equals to 1 if school i has an editor at journal j at year $t-2$, 0 otherwise.
Panel B: Author Level	
i	Author.
j	Journal.
t	Year.
$Editor_{i,j,t}$	equals 1 if author i is an editor of journal j at year $t-2$, 0 otherwise.
$nYearPreEditor_{i,j,t}$	equals 1 if $Editor_{i,j,t-n} = 1$ and $Editor_{i,j,t-1} = Editor_{i,j,t-2} = \cdots = Editor_{i,j,t-n} = 0$.
$nYearPostEditor_{i,j,t}$	equals 1 if $Editor_{i,j,t+n} = 1$ and $Editor_{i,j,t+1} = Editor_{i,j,t+2} = \cdots = Editor_{i,j,t+n} = 0$.
$CitationStock_{i,t}$	Total citations for author i prior to year t.
$Yearssincefirstcitation_{i,t}$	Number of years between the year author i has first citation and year t.
$c_{i,j,t}$	Author i's citation in journal j at year t.
Panel C: Article Level	
i	Article.
j	Citation Journal.
t	Citation Year.
$J(i)$	Article i's publication journal.
$Pubyear_i$	Article i's publication year.
$Nauthor_i$	Number of authors for article i.
$CitationStock_{i,t}$	Total number of citations for article i prior to year t in top finance journals.
$ArticleAge_{i,t}$	Number of years between the publication year of article i and year t.
$Connected_{i,j,t}$	Equals 1 if article i shares the same author affiliations with journal j's editors' affiliation at year t, 0 otherwise.
$Connected_{i,J(i),t}$	Equals 1 if article i shares the same author affiliations with its publication journal $J(i)$'s editors' affiliation at year t, 0 otherwise.
$Connected_{i,J(i),0}$	Equals 1 if article i shares the same author affiliations with its publication journal $J(i)$'s editors' affiliation when it is published, 0 otherwise.
$Connected_{i,j,0}$	Equals 1 if article i shares the same author affiliations with journal j's editors' affiliation when it is published, 0 otherwise.
$Connected_{i,any,0}$	Equals 1 if article i shares the same author affiliations with any journal's editors' affiliation when it is published, 0 otherwise.
$Connectedtootherjournal_{i,j,t}$	Equals 1 if article i shares the same author affiliations with any journal $j' \neq j$'s editors' affiliation at year t, 0 otherwise.
$NotFirstConnected_i$	Equals 1 if article i does not share the affiliations with any journal j's editors' affiliations when it is published, 0 otherwise.
$Editor_{i,J(i),t}$	Equals 1 if the author of article i is an editor at its publication journal $J(i)$ at year t, 0 otherwise.
$Editor_{i,j\neq J(i),t}$	Equals 1 if the author of article i is an editor at journal $j \neq J(i)$ at year t, 0 otherwise.
c_i	Total citations of article i as of 2014.
c_i^{Fina}	Total citations of article i in top finance journals as of 2014.
$c_{i,j}^{Fina}$	Total citations of article i in journal j as of 2014.
$c_{i,j,t}$	Article i's citations in journal j at year t.

Table C.3: Editors of Finance Journals

Editors	JF	JFE	RFS
Michael Adler	1971-1978		
Yacine Ait-Sahalia			2003-2006
Franklin Allen			1994-1996
Kerry Back			1994-1997
Geert Bekaert			2009-2014
Bruno Biais	2012-2014		
Marshall Blume	1977-1980		
William Branson	1979-1980		
Michael Brennan	1980-1983		1988-1990
Stephen Brown			1988-1990
Stephen Buser	1988-1995		
Francesca Cornelli			2014
Dave Denis			2013-2014
Bernard Dumas			1997-2000
Philip Dybvig			1990-1993
Edwin Elton	1983-1988		
Eugene Fama		1974-1975	
Wayne Ferson			1997-1999
Michael Fishman			1998-2002
Paolo Fulghieri			2007-2010
Michael Gibbons			1988-1991
Lawrence Glosten			1997-1999
Itay Goldstein			2013-2014
Gary Gorton			1997-2000
John Graham	2006-2012		
Richard Green	2000-2003		1991-1992;1994-1995
Robin Greenwood			2014
Martin Gruber	1983-1988		
Jack Guttentag	1974-1977		
Campbell Harvey	2006-2012		2000-2005
Joel Hasbrouck			2004-2008
John Heaton			2000-2002
David Hirshleifer			1994-1997;2011-2014
Chi-Fu Huang			1993-1995
Jonathan Ingersoll			1988-1989
Ravi Jagannathan			1994-1999
Michael Jensen		1974-1975;1977-2014	
Andrew Karolyi			2011-2014
Leonid Kogan			2013-2014
Robert Korajczyk			1994-1996
Alexander Ljungqvist			2008-2014
Andrew Lo			1992-1995
John Long Jr		1983-1996	
David Mayers	1988-1995		
Robert Mcdonald			2002-2006
Robert Merton		1974-1975	
Wayne Mikkelson		1993-1996	
Tobias Moskowitz			2006-2008
Stefan Nagel			2014
Maureen O'Hara			2000-2005
Terrance Odean			2006
Michael Roberts	2012-2014		
Alexander Robichek	1971-1973		
Richard Ruback		1988-1996	
William Schwert		1979-1986;1989-2014	
Kenneth Singleton	2012-2014		2001-2003
Clifford Smith		1983-1996	
Chester Spatt			1988-1993
Matt Spiegel			2006-2011
Robert Stambaugh	2003-2006		1991-1993
Laura Starks			2009-2014
Philip Strahan			2014
Rene Stulz	1988-2009	1983-1987	
Sheridan Titman			1997-1998
Raman Uppal			2007-2010
Pietro Veronesi			2009-2013
Jerold Warner		1987-1996	

Continued on next page

Table C.3 – continued from previous page

Editors	JF	JFE	RFS
Michael Weisbach			2006-2013
Toni Whited		2014	

Table C.2: Affiliations of Finance Journals

This table lists affiliations with their finance journal editorship connections. We only focus on co-editors, editors, executive editor, and managing editors. Sources are from journals' preface.

Affiliations	JF	JFE	RFS
Boston College			1997,2014
Carnegie Mellon University	2000-2003		1988-1995
Columbia University	1971-1978		1997-1999,2009-2014
Cornell University			2000-2005,2011-2014
Dartmouth College		1993-1996	
Duke University	2006-2012		2000-2005
Federal Reserve System			1994-1997
Harvard University		1985-1996	2014
Hautes Etudes Commerciales			1997-1999
Hong Kong University of Science & Technology			1995
INSEAD Business School			1999-2000
London Business School			2007-2010,2014
Massachusetts Institute of Technology		1974-1975	1992-1995,2013-2014
New York University	1983-1988		1988-1990,2004-2014
Northwestern University			1994-2006
Ohio State University	1988-2009	1983-1987	2011-2013
Princeton University	1979-1980		2003-2006
Stanford University	1971-1973,2012-2014		1988-1989,2001-2003
Toulouse School of Economics	2012-2014		
University of California Berkeley			2006
University of California Irvine			2011-2014
University of California Los Angeles			1988-1990
University of Chicago		1974-1975	2000-2002,2006-2013
University of Illinois Urbana-Champaign			2006-2011
University of Michigan			1994-1997,2014
University of Minnesota Twin Cities			1994-1998
University of North Carolina Chapel Hill			2007-2010
University of Pennsylvania	1974-1980,2003-2006, 2012-2014		1990-2000,2013-2014
University of Pittsburgh			2013-2014
University of Rochester		1974-2014	
University of Texas Austin			1997-1998,2009-2014
University of Washington Seattle			1997-1999
Washington University in St. Louis			1990-1997
Yale University			1988-1989,2006-2011

Table C.4: Summary Statistics

Panel A: School-Journal-Year Level

	(1) . count	mean	sd	min	max
Year	73305	1992.000	12.987	1970	2014
$Publication\_i, j, t$	60273	0.218	0.788	0	15
$\overline{Publication\_i}, j, t$	57015	0.217	0.704	0	14
Has Editor(s)	60273	0.005	0.069	0	1
School ID	73305	272.000	156.751	1	543
Journal ID	73305	2.000	0.817	1	3

Panel B: Author-Journal-Year Level

	(1) . count	mean	sd	min	max
$StockCitation\_i, t$	5977	165.994	332.927	0	4249
Year	5977	1999.134	10.069	1970	2014
c_i, j, t	5977	5.662	10.213	0	164
$Editor\_i, j, t$	5977	0.065	0.246	0	1
Author ID	5977	35.557	20.335	1	71

Panel C: Article-Journal-Year Level

	(1) . count	mean	sd	min	max
c^{Fina}_i, j, t	362600	0.220	0.675	0	25
$Connected\_i, j, t$	362600	0.088	0.283	0	1
$Connectedtootherjournal\_i, j, t$	360617	0.168	0.374	0	1
Citation Stock	362600	7.872	15.651	1	528
$ArticleAge$	362600	13.490	10.479	0	44
$NotFirstConnected$	362600	0.819	0.385	0	1

Table C.5: Citations After the Office

This table shows the editor step-down effect. The dependent variable in column (1) and (2) are $\log(c_{i,j,t}+1)$ where $c_{i,j,t}$ is the total number of citation from author i at journal j at year t.The dependent variable in column (3), (4), (7), and (8) are $c_{i,j,t}$. The dependent variable in column (5) and (6) are $c_{i,j,t}/TotalCitation_{j,t}$, the share of citations within each journal-year.We run OLS for column (1) to (6) and poisson regression for column (7) and (8). We restrict journal $j \in \{$*Journal of Finance*, *Journal of Financial Economics*, *Review of Financial Studies*$\}$. The dummy $Editor_{i,j,t}$ equals 1 if author i is an editor at journal j at year t. The dummy n *Year Post Editor* equals 1 if $Editor_{i,j,t-n} = 1$ and $Editor_{i,j,t-1} = Editor_{i,j,t-2} = \cdots = Editor_{i,j,t-n} = 0$. The dummy n *Year Pre Editor* equals 1 if $Editor_{i,j,t+n} = 1$ and $Editor_{i,j,t+1} = Editor_{i,j,t+2} = \cdots = Editor_{i,j,t+n} = 0$. Robust standard error are clustered at author level. T tests of statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	$\log(c_{i,j,t}+1)$		$c_{i,j,t}$		$c_{i,j,t}/TotalCitations_{j,t}$		$c_{i,j,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
5 Year Pre Editor	-0.062	0.055	-0.345	0.729	0.088	0.123**	-0.098	0.052
	(-0.38)	(0.34)	(-0.41)	(0.79)	(1.56)	(2.01)	(-0.59)	(0.31)
4 Year Pre Editor	0.136	0.258*	0.509	1.598*	0.156***	0.190***	0.110	0.295**
	(1.06)	(1.90)	(0.76)	(1.95)	(2.65)	(2.98)	(0.92)	(2.50)
3 Year Pre Editor	0.047	0.166	0.025	1.097	0.112**	0.146**	-0.001	0.195
	(0.36)	(1.17)	(0.03)	(1.13)	(2.15)	(2.51)	(-0.01)	(1.28)
2 Year Pre Editor	0.030	0.146	0.221	1.250	0.097*	0.128**	0.040	0.226
	(0.24)	(1.12)	(0.25)	(1.31)	(1.81)	(2.27)	(0.30)	(1.54)
1 Year Pre Editor	0.035	0.149	0.506	1.502	0.143***	0.172***	0.085	0.272*
	(0.27)	(1.07)	(0.50)	(1.41)	(2.95)	(3.31)	(0.62)	(1.95)
Editor	0.371***	0.436***	2.784***	3.276***	0.272***	0.279***	0.324***	0.362***
	(5.33)	(6.46)	(2.91)	(3.53)	(2.98)	(3.15)	(5.94)	(4.83)
1 Year Post Editor	0.443***	0.549***	0.738	1.649	0.260**	0.284**	0.105	0.203
	(3.02)	(3.76)	(0.52)	(1.20)	(2.46)	(2.51)	(0.61)	(0.92)
2 Year Post Editor	0.173	0.277**	-0.588	0.317	0.124**	0.150**	-0.093	0.033
	(1.32)	(2.03)	(-0.46)	(0.26)	(2.35)	(2.49)	(-0.61)	(0.17)
3 Year Post Editor	0.164	0.264	-0.181	0.705	0.164*	0.191*	-0.031	0.100
	(1.05)	(1.66)	(-0.17)	(0.67)	(1.74)	(1.86)	(-0.21)	(0.60)
4 Year Post Editor	0.038	0.136	-0.942	-0.055	0.090	0.118	-0.177	-0.006
	(0.33)	(1.17)	(-0.96)	(-0.06)	(1.25)	(1.64)	(-1.22)	(-0.04)
5 Year Post Editor	0.106	0.199	-0.170	0.674	0.165*	0.192**	-0.033	0.143
	(0.74)	(1.42)	(-0.15)	(0.67)	(1.98)	(2.20)	(-0.22)	(0.99)
6 Year Post Editor	0.028	0.115	-0.957	-0.163	0.085**	0.111**	-0.148	0.027
	(0.24)	(1.02)	(-0.88)	(-0.17)	(2.02)	(2.15)	(-1.16)	(0.21)
7 Year Post Editor	0.157	0.242	-0.289	0.480	0.092*	0.117*	-0.058	0.148
	(0.99)	(1.57)	(-0.24)	(0.43)	(1.74)	(1.95)	(-0.32)	(0.84)
8 Year Post Editor	0.087	0.162	-0.170	0.514	0.122	0.145	-0.027	0.198
	(0.56)	(1.03)	(-0.19)	(0.55)	(1.20)	(1.32)	(-0.22)	(1.49)
9 Year Post Editor	0.170	0.238*	0.654	1.284	0.030	0.051	0.098	0.306**
	(1.26)	(1.84)	(0.54)	(1.01)	(0.70)	(1.06)	(0.72)	(1.98)
10 Year Post Editor	0.240*	0.311**	0.163	0.803	0.069	0.090*	0.038	0.263**
	(1.70)	(2.15)	(0.18)	(0.94)	(1.59)	(1.89)	(0.34)	(2.17)
Constant	1.236***	1.381***	5.486***	6.328***	0.234***	0.217***	-0.405***	-0.229***
	(94.12)	(61.39)	(52.86)	(18.77)	(24.53)	(13.59)	(-1.84e+10)	(-8.06)
Author-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Citation Jour. FE	No	Yes	No	Yes	No	Yes	No	Yes
Observation	5,977	5,977	5,977	5,977	5,977	5,977	5,977	5,977
R^2	0.788	0.802	0.790	0.798	0.823	0.827		

TABLE C.6: Editor's Colleagues Benefits from Citation Premium

This table shows that authors who publish in journal j do quote at an abnormally high level the papers of colleagues of the editors at journal j, even when those papers have not been published in the editor's journal. In panel A we use all the articles in our sample. In panel B we restrict to a subsample that contains all articles that are not first connected to their own journal when they are published. The dependent variable is $\log(c_{i,j,t}+1)$ where $c_{i,j,t}$ is total citations for article i in Top finance journal j at year t. The dummy variable $Connected_{i,j,t}$ equals 1 if the article i's author's affiliations at year t overlap with journal j's editors' affiliation at year $t-2$. The dummy variable $Connectedtootherjournal_{i,j,t}$ equals 1 if the article i's author's affiliations at year t overlap with journal $j' \neq j$'s editors' affiliation at year $t-2$. $\log(CitationStock_{i,t})$ is the log of stock citations for article i before year t. $ArticleAge_{i,t}$ is the age of publication i at year t. T statistics in the parenthesis, statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard error clustered at article level.

	Dependent variable: $\log(c^{Fina}_{i,j,t}+1)$					
Sample: All	(1)	(2)	(3)	(4)	(5)	(6)
$Connected\_i, j, t$	0.027***	0.028***	0.026***	0.026***	0.010***	0.018***
	(7.59)	(7.83)	(7.23)	(7.42)	(3.08)	(4.67)
$Connectedtootherjournal\_i, j, t$	0.008***	0.008***	0.007**	0.009***	-0.006**	
	(2.67)	(2.61)	(2.36)	(2.86)	(-2.41)	
$1\_j = J(i)$	0.031***	0.022***	0.031***	0.050***		0.031***
	(19.72)	(13.81)	(19.73)	(8.99)		(15.90)
$\log(CitationStock)$	0.098***	0.098***	0.097***	0.098***	0.022***	
	(36.67)	(36.66)	(36.55)	(36.68)	(9.91)	
$ArticleAge$	-0.008***	-0.008***	-0.008***	-0.008***	-0.001***	
	(-42.27)	(-42.27)	(-43.69)	(-42.35)	(-7.43)	
$Editor\_i, J(i), t$	0.070**	0.070**	0.056*	0.070**	0.030**	
	(2.36)	(2.36)	(1.91)	(2.36)	(2.39)	
$Editor\_i, j \neq J(i), t$	-0.061	-0.062	-0.031	-0.061	-0.141**	
	(-0.87)	(-0.87)	(-0.44)	(-0.86)	(-2.18)	
Constant	0.102***	0.103***	0.099***	0.076***	0.088***	0.116***
	(25.12)	(26.30)	(31.66)	(15.55)	(19.46)	(65.74)
Observations	362600	362600	362600	362600	362600	362600
Sample: Not First Connected to its own journal	(1)	(2)	(3)	(4)	(5)	(6)
$Connected\_i, j, t$	0.019***	0.019***	0.019***	0.020***	0.008**	0.006
	(5.32)	(5.30)	(5.40)	(5.62)	(2.42)	(1.60)
$Connectedtootherjournal\_i, j, t$	0.013***	0.013***	0.013***	0.013***	-0.005*	
	(3.83)	(3.91)	(3.88)	(3.69)	(-1.67)	
$1\_j = J(i)$	0.030***	0.020***	0.030***	0.054***		0.030***
	(18.01)	(12.04)	(18.01)	(8.97)		(14.49)
$\log(CitationStock)$	0.096***	0.096***	0.095***	0.096***	0.022***	
	(34.85)	(34.84)	(34.68)	(34.85)	(9.19)	
$ArticleAge$	-0.007***	-0.007***	-0.008***	-0.007***	-0.001***	
	(-42.67)	(-42.67)	(-43.89)	(-42.73)	(-6.75)	
$Editor\_i, J(i), t$	0.008	0.007	0.004	0.008	0.019*	
	(0.48)	(0.48)	(0.29)	(0.49)	(1.66)	
$Editor\_i, j \neq J(i), t$	-0.057	-0.057	-0.032	-0.057	-0.142**	
	(-0.80)	(-0.80)	(-0.45)	(-0.79)	(-2.18)	
Constant	0.102***	0.104***	0.099***	0.073***	0.084***	0.111***
	(26.73)	(29.04)	(33.48)	(14.31)	(18.65)	(59.07)
Journal of Pub. FE	Yes	Yes	No	No	No	No
Journal of Cit. FE	Yes	No	Yes	No	No	Yes
Year of Cit. FE	Yes	No	No	Yes	Yes	No
Journal of Cit.-Year of Cit. FE	No	Yes	No	No	No	No
Journal of Pub.-Year of Cit. FE	No	No	Yes	No	No	No
Journal of Pub.-Journal of Cit. FE	No	No	No	Yes	No	No
Article-Jou of Cit. FE	No	No	No	No	Yes	No
Article-Year of Cit. FE	No	No	No	No	No	Yes
Observation	331,123	331,123	331,123	331,123	331,123	331,123
R^2	0.168	0.178	0.175	0.169	0.371	0.577

Table C.7: Test I

This table shows that citations of colleagues of the editor increase in the journal of editor both for the papers these colleagues publish in the journal of the editor and in other journals. The dependent variable is $\log(c_{i,j,t}+1)$ where $c_{i,j,t}$ is total citations for article i at journal j in year t. The dummy variable $Connected_{i,J(i),0}$ equals 1 if the article i's author's affiliations when it is published at year t overlap with its publication journal $J(i)$'s editors' affiliation at year $t-2$. The dummy variable $Connected_{i,any,0}$ equals 1 if the article i's author's affiliations when it is published at year t overlap with any journal j's editors' affiliation at year $t-2$. $\log(CitationStock_{i,t})$ is the log of stock citations for article i before year t. $ArticleAge_{i,t}$ is the age of publication i at year t. T statistics in the parenthesis, statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard error clustered at article level.

Depedent Variable: $\log(c_{i,j,t}+1)$					
	(1)	(2)	(3)	(4)	(5)
$Connected\_i, J(i), 0$	-0.016*	-0.016*	-0.018**	-0.016*	-0.018**
	(-1.78)	(-1.78)	(-2.10)	(-1.81)	(-2.10)
$Connected\_i, any, 0$	0.029***	0.029***	0.028***	0.029***	0.028***
	(4.57)	(4.57)	(4.40)	(4.57)	(4.40)
$1\_j = J(i)$	0.031***	0.022***	0.031***		
	(19.84)	(13.98)	(19.84)		
$\log(CitationStock)$	0.098***	0.098***	0.097***	0.098***	0.097***
	(36.73)	(36.73)	(36.59)	(36.74)	(36.58)
$ArticleAge$	-0.007***	-0.007***	-0.008***	-0.007***	-0.008***
	(-41.10)	(-41.10)	(-42.47)	(-41.18)	(-42.46)
$Editor\_i, J(i), t$	0.074**	0.074**	0.061**	0.074**	0.061**
	(2.55)	(2.55)	(2.11)	(2.54)	(2.11)
$Editor\_i, j \neq J(i), t$	-0.057	-0.057	-0.028	-0.057	-0.028
	(-0.84)	(-0.84)	(-0.41)	(-0.83)	(-0.41)
Constant	0.098***	0.100***	0.096***	0.123***	0.106***
	(24.11)	(25.53)	(29.73)	(30.33)	(34.75)
Journal of Pub. FE	Yes	Yes	No	No	No
Journal of Cit. FE	Yes	No	Yes	No	No
Year of Cit. FE	Yes	No	No	Yes	No
Journal of Cit.-Year of Cit. FE	No	Yes	No	No	No
Journal of Pub.-Year of Cit. FE	No	No	Yes	No	No
Journal of Pub.-Journal of Cit. FE	No	No	No	Yes	No
Journal of Pub.-Journal of Cit-Year of Cit. FE	No	No	No	No	Yes
Observation	362,600	362,600	362,600	362,600	362,600
R^2	0.170	0.180	0.179	0.172	0.192

TABLE C.8: Test II: Time Series Evidence

This table shows that citations of colleagues of the editor decrease when editor steps down. The dependent variable is $\log(c_{i,j,t}+1)$ where $c_{i,j,t}$ is total citations for article i in journal j at year t. The dummy variable $Connected_{i,J(i),0}$ equals 1 if the article i's author's affiliations when it is published at year t overlap with its publication journal $J(i)$'s editors' affiliation at year $t-2$. The dummy variable $Connected_{i,J(i),t}$ equals 1 if the article i's author's affiliations at year t overlap with its own journal $J(i)$'s editors' affiliation at year $t-2$. $\log(CitationStock_{i,t})$ is the log of stock citations for article i before year t. $ArticleAge_{i,t}$ is the age of publication i at year t. T statistics in the parenthesis, statistical significance: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Robust standard error clustered at article level.

Dependent variable: $\log(c^{Fina}_{i,j,t}+1)$					
	(1)	(2)	(3)	(4)	(5)
$Connected\_i, J(i), 0$	0.010	0.003	-0.006	-0.002	
	(1.54)	(0.40)	(-0.94)	(-0.24)	
$Connected\_i, J(i), t$		0.055***	0.008	-0.008	0.023***
		(7.97)	(1.04)	(-1.05)	(4.00)
$Connected\_i, J(i), 0 \times Connected\_i, J(i), t$			0.092***	0.077***	0.059***
			(5.93)	(4.99)	(5.15)
$\log(CitationStock)$	0.099***	0.098***	0.099***	0.099***	0.020***
	(37.25)	(37.14)	(37.18)	(37.21)	(8.86)
$ArticleAge$	-0.008***	-0.008***	-0.008***	-0.008***	-0.001***
	(-42.81)	(-42.73)	(-42.59)	(-42.71)	(-9.16)
$Editor\_i, J(i), t$	0.076***	0.065**	0.066**	0.070**	0.021*
	(2.63)	(2.26)	(2.30)	(2.46)	(1.76)
$Editor\_i, j \neq J(i), t$	-0.043	-0.042	-0.042	-0.042	-0.138**
	(-0.59)	(-0.58)	(-0.58)	(-0.58)	(-2.15)
Constant	0.124***	0.124***	0.124***	0.128***	0.092***
	(31.71)	(31.49)	(31.79)	(31.43)	(24.22)
Year of Cit FE	Yes	Yes	Yes	Yes	Yes
Journal of Cit. FE	Yes	Yes	Yes	No	No
Journal of Pub. FE	Yes	Yes	Yes	No	No
Journal of Pub.-Journal of Cit. FE	No	No	No	Yes	Yes
Article-Year of Cit. FE	No	No	No	No	No
Article FE	No	No	No	No	Yes
Observation	362,600	362,600	362,600	362,600	362,600
R^2	0.168	0.169	0.169	0.172	0.323

Table C.9: Test III: Cross Sectional Evidence

This table show that citations of colleague's papers published by an editor are not higher in other journals. The dependent variable is $\log(c_{i,j,t}+1)$ where $c_{i,j,t}$ is total citations for article i in finance journal j at year t. The dummy variable $Connected_{i,J(i),0}$ equals 1 if the article i's author's affiliations when it is published at year t overlap with its publication journal $J(i)$'s editors' affiliation at year $t-2$. $1_{j=J(i)}$ equals 1 if journal j is the publicaiton journal of article i, and 0 otherwise. $\log(CitationStock_{i,t})$ is the log of stock citations for article i before year t. $ArticleAge_{i,t}$ is the age of publication i at year t. T statistics in the parenthesis, statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard error clustered at article level.

Dependent Variable: $\log(c_{i,j,t}+1)$				
	(1)	(2)	(3)	(4)
$Connected\_i, J(i), 0$	-0.001	-0.003	-0.003	0.002
	(-0.11)	(-0.51)	(-0.43)	(0.26)
$1\_j = J(i)$	0.028***	0.028***	0.018***	
	(17.23)	(17.48)	(11.09)	
$Connected\_i, J(i), 0 \times 1\_j = J(i)$	0.032***	0.029***	0.038***	0.024***
	(4.89)	(4.51)	(5.73)	(3.79)
$\log(CitationStock)$	0.099***	0.098***	0.099***	0.099***
	(37.24)	(37.10)	(37.24)	(37.25)
$ArticleAge$	-0.008***	-0.008***	-0.008***	-0.008***
	(-41.72)	(-42.97)	(-41.72)	(-41.79)
Number of Authors	-0.001	-0.001	-0.001	-0.001
	(-0.37)	(-0.26)	(-0.37)	(-0.36)
Constant	0.108***	0.104***	0.109***	0.129***
	(18.80)	(20.06)	(19.37)	(22.48)
Year of Cit. FE	Yes	No	No	Yes
Journal of Cit. FE	Yes	Yes	No	No
Journal of Pub. FE	Yes	No	Yes	No
Journal of Pub.-Journal of Cit. FE	No	No	No	Yes
Journal of Pub.-Year of Cit. FE	No	Yes	No	No
Journal of Cit.-Year of Cit. FE	No	No	Yes	No
Observation	362,600	362,600	362,600	362,600
R^2	0.170	0.178	0.180	0.171

Figures

Figure C.1: Editors' Citations Share

This figure shows the citations share dynamics for editors from 2 years before editorship and 7 years after editorship. We define editorship starts from 2 years after assuming the office and ends after 2 years stepping down the office. Citations share is defined as number of citations for editor i at year t over the total number of references in editor's jouranl j at year t. Same journal citations are citations in the editors' own journal. Other journal citations are average citations from the other journals.

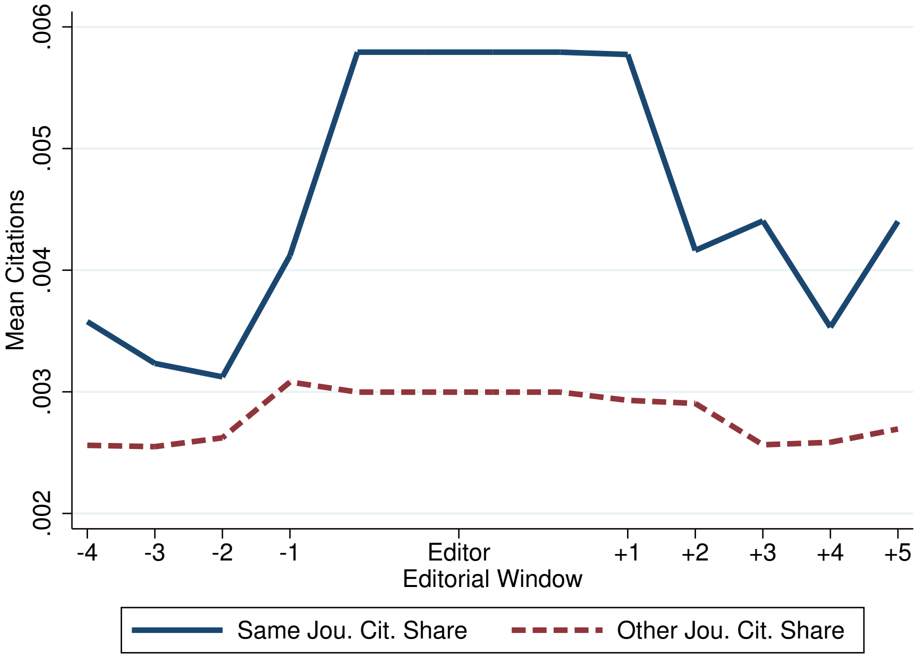


FIGURE C.2: Coefficients of Editor's Citations Before and After Office

This table shows the editor step-down effect in column 1 of table III. The dependent variable is $\log(c_{i,j,t}+1)$ where $c_{i,j,t}$ is the total number of citation from author i at journal j at year t. We restrict journal $j \in \{$*Journal of Finance*, *Journal of Financial Economics*, *Review of Financial Studies*$\}$. The dummy variable $Editor_{i,j,t} = 1$ if author i is an editor of journal j at year $t-2$. The dummy n *Year Post Editor* equals 1 if $Editor_{i,j,t-n} = 1$ and $Editor_{i,j,t-1} = Editor_{i,j,t-2} = \cdots = Editor_{i,j,t-n} = 0$. The dummy n *Year Pre Editor* equals 1 if $Editor_{i,j,t+n} = 1$ and $Editor_{i,j,t+1} = Editor_{i,j,t+2} = \cdots = Editor_{i,j,t+n} = 0$.

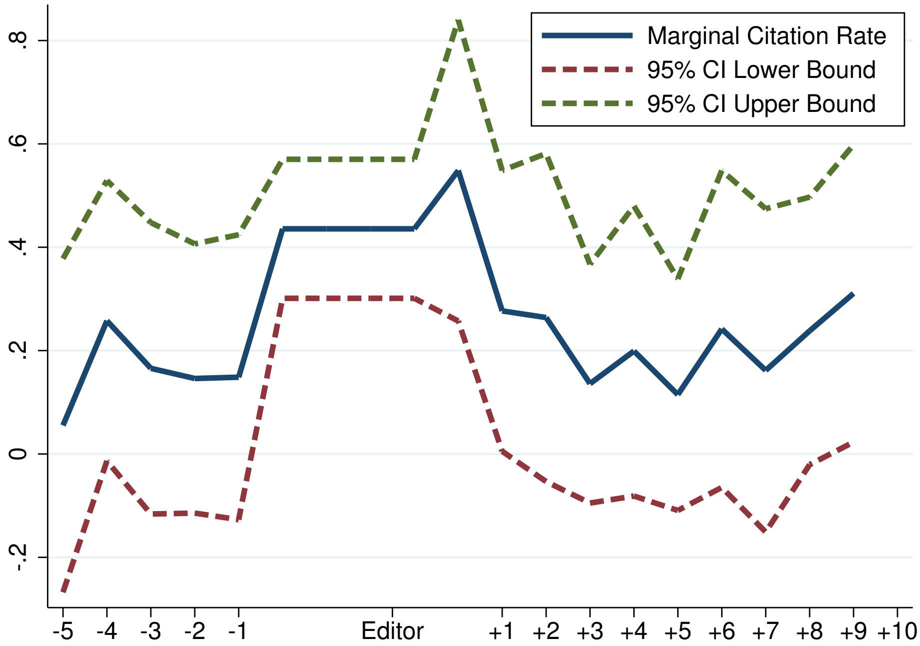


FIGURE C.3: Depedent Variable is $\log(c_{i,j,t}+1)$ with author-year and journal fixed effects

FIGURE C.4: Schools' Citations Share

This graph shows the citations share dynamics for Schools that have editors from 2 years before editorship and 7 years after editorship. We define editorship starts from 2 years after assuming the office and ends after 2 years stepping down the office. Citations share is defined as number of citations for school i with editor(s) in journal j at year t over the total number of references in jouranl j at year t. Same journal citations are citations in the editors' own journal. Other journal citations are average citations from the other journals.

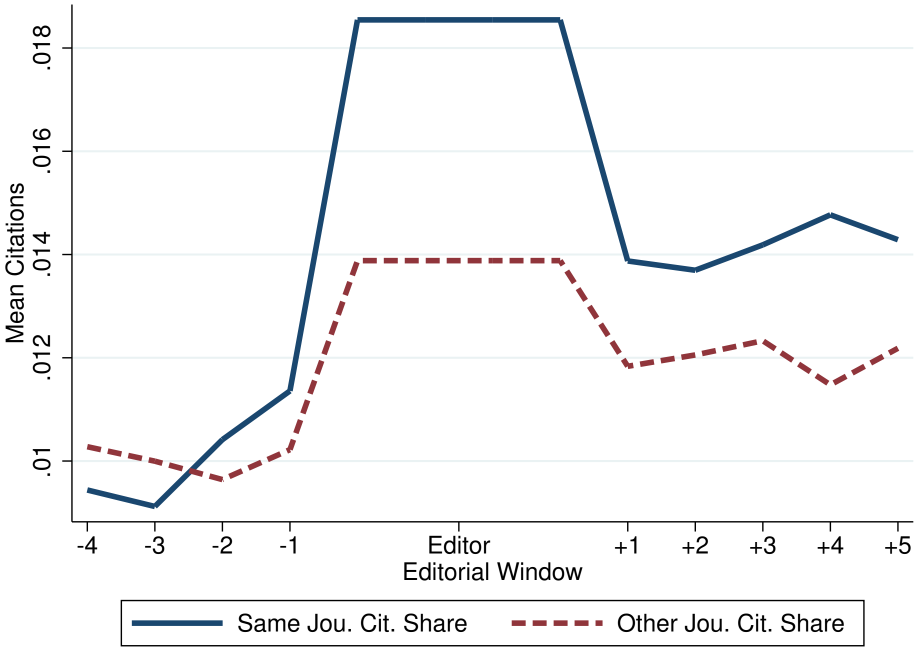
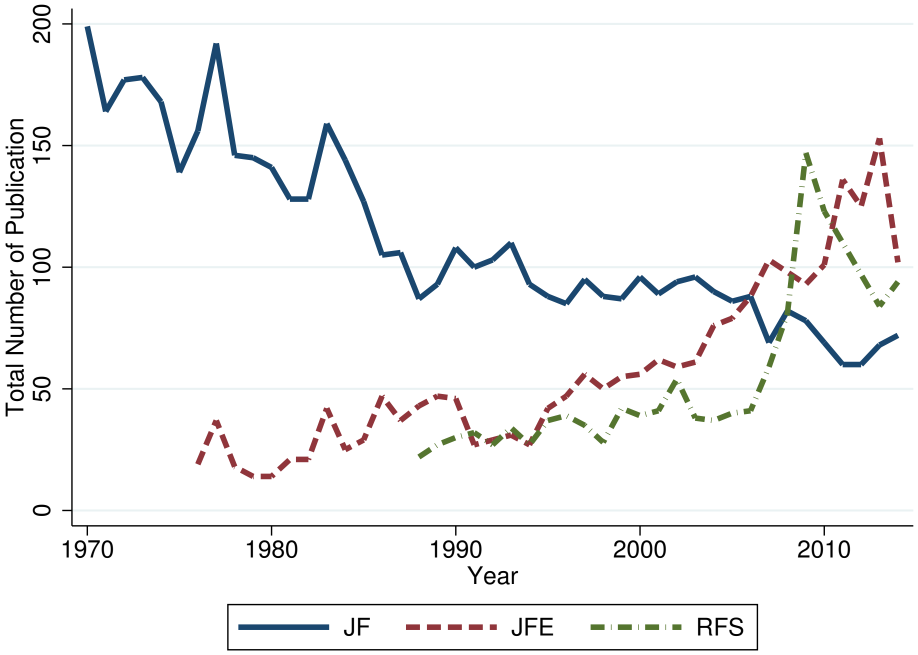


Figure C.5: Number of Publications and References

This graph shows the total number of publications and references across journal and year.



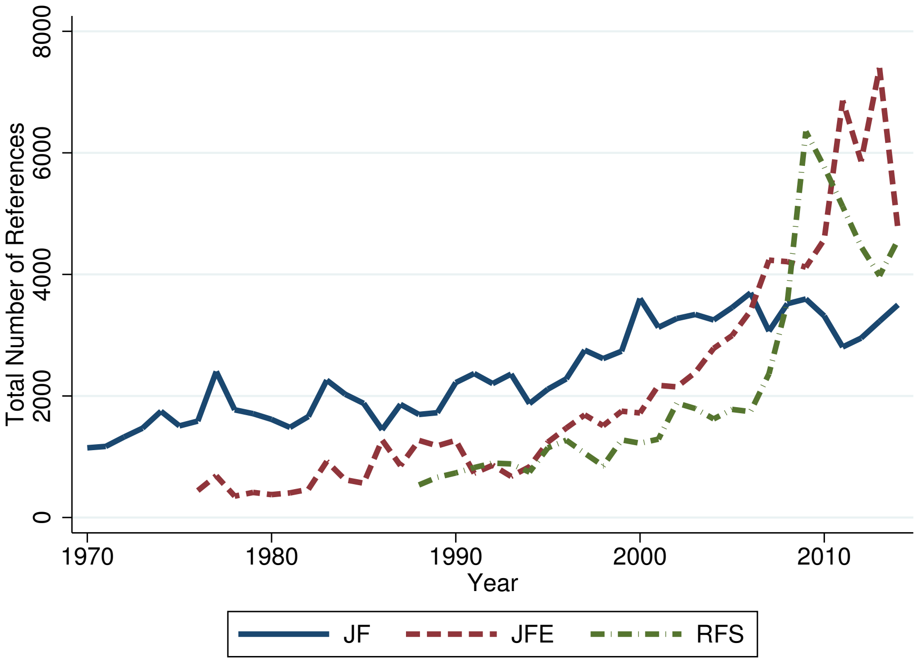


Figure C.6: Share of Connected Articles

This graph shows the share of connected articles across journals and years.

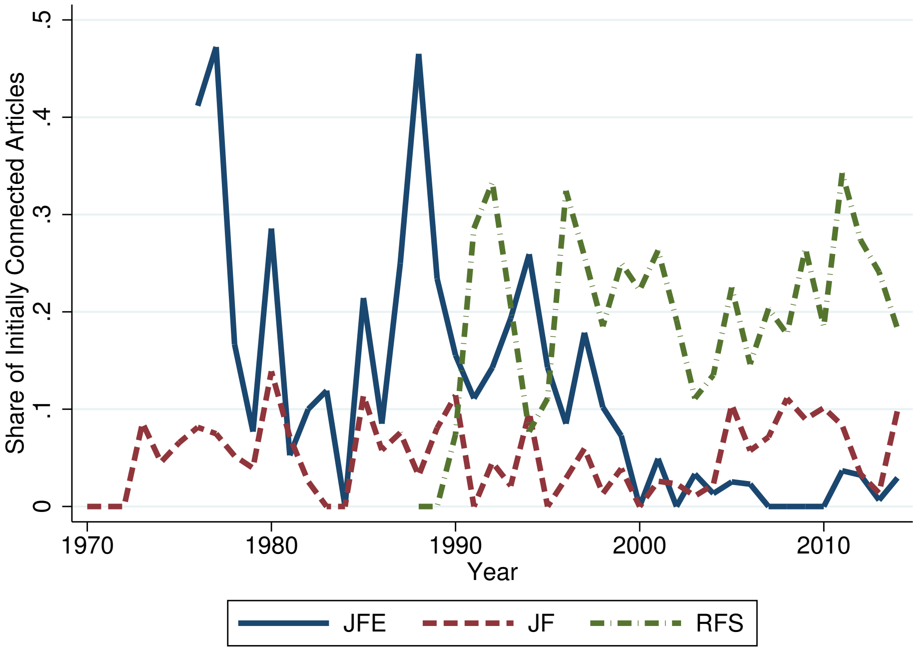


Figure C.7: Number of Publications

This graph shows the total number of publications for each journal each year.

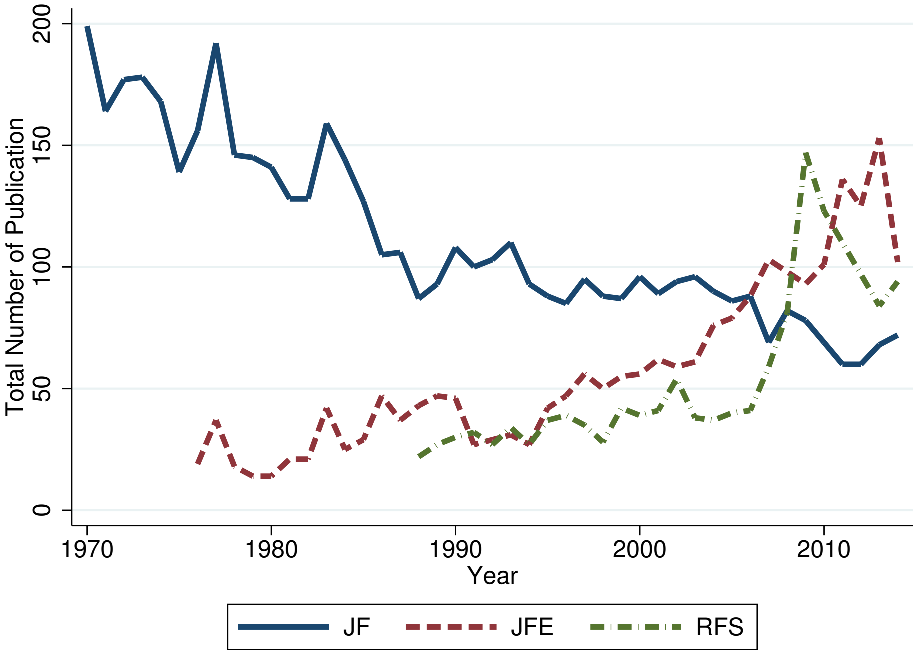


Figure C.8: Number of References

This graph shows the total number of references for each journal each year.

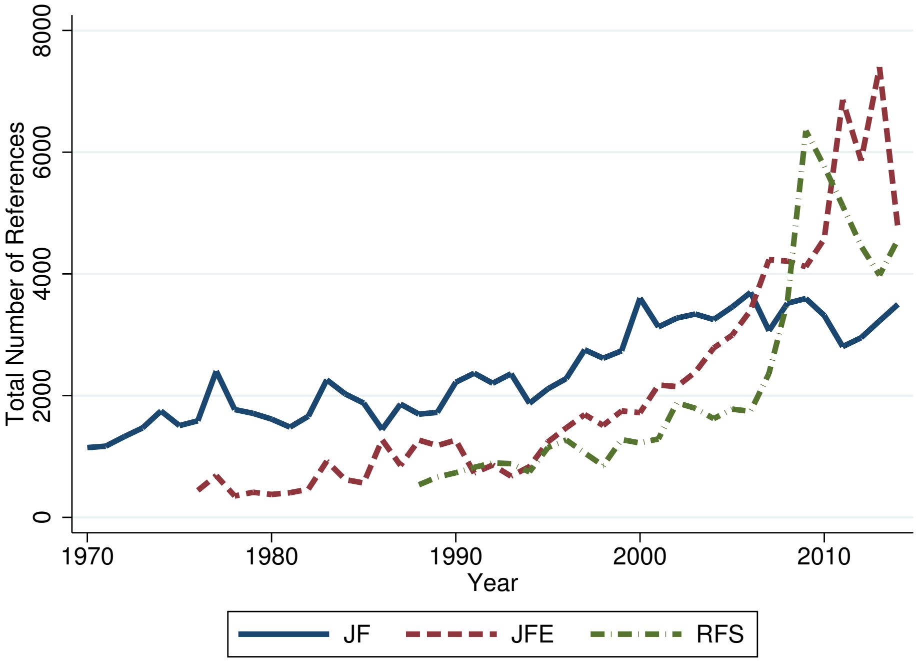


FIGURE C.9: Life Cycle of Article Citations

This graph shows the life cycle citation shares for connected and not connected articles. Connected articles are those articles that their author's affiliations when it is published at year t overlap with its own journal;s editors' affiliation at year $t-2$. Citations share is defined by total number of citations for article i at journal j at year t devided by the total number of citations at journal j at year t. We construct mean citations over articles' age by connectedness.

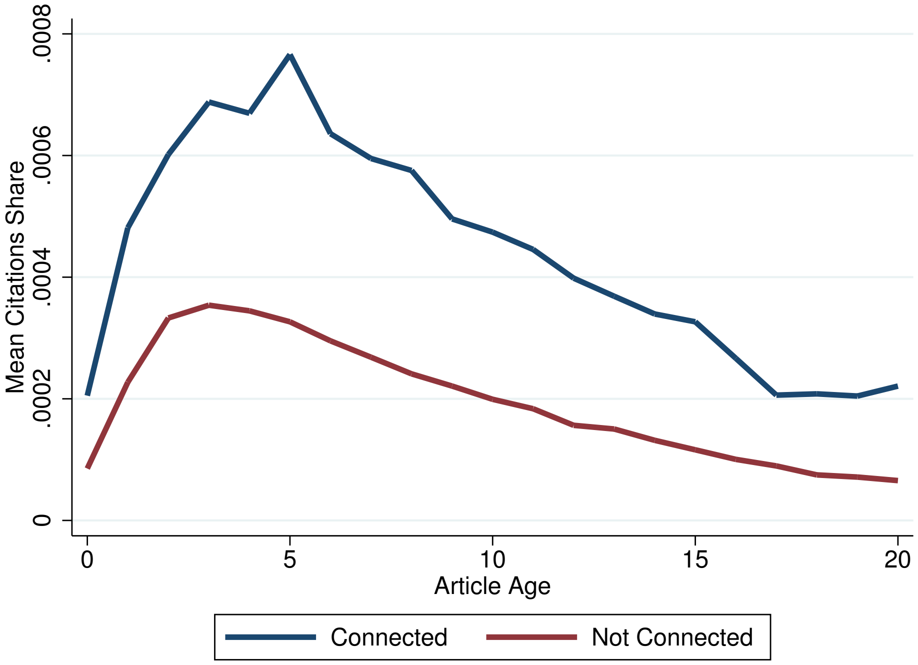


TABLE C.10: Schools Connected to a Journal Publish More in that Journal

This table shows during editorship premium. The dependent variable is logarithm of $Publication_{i,j,t}$ which is the total number of publication from school i and journal j in year t. $During_{i,j,t}$ is a dummy variable, equals to 1 when school i has an editor at journal j at year $t-2$. We estimate the following regressions $\log(Publication_{i,j,t}) = \beta During_{i,j,t} + FE + \epsilon_{i,j,t}$ where FE are various fixed effects. We include all schools who ever published at *Journal of Finance*, *Journal of Financial Economics*, and *Review of Financial Studies* in the sample. Robust standard error are clustered at school level. T tests of statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Dependent variable: $\log(Publication_{i,j,t}+1)$					
	(1)	(2)	(3)	(4)	(5)	(6)
$During\_i, j, t$	0.935***	0.131**	0.132**	0.094*	0.102	0.453***
	(13.18)	(2.40)	(2.44)	(1.78)	(1.16)	(7.58)
$\log \overline{Publication}\_i, j, t$						0.550***
						(21.03)
Constant	0.108***	0.131***	0.131***	0.107***	0.136***	0.042***
	(12.20)	(11.87)	(11.86)	(9.69)	(40.99)	(14.10)
Journal FE	No	Yes	No	No	Yes	Yes
Year FE	No	Yes	No	Yes	No	No
School FE	No	Yes	Yes	No	No	No
Journal-Year FE	No	No	Yes	No	No	No
Journal-School FE	No	No	No	Yes	No	No
School-Year FE	No	No	No	No	Yes	No
Observation	60,273	60,273	60,273	60,273	60,273	57,015
R^2	0.037	0.462	0.471	0.490	0.696	0.304

Table C.11: Schools Connected to a Journal Do Not Publish More in Other Journals

This table shows indirect effect of editor premium. The dependent variable is logarithm of $\overline{Publication}_{i,j,t}$ which is defined as average number of publication from school i and journal $j' \neq j$ in year t. Formally $\overline{Publication}_{i,j,t} = \sum_{j' \neq j} Publication_{i,j',t}/N$. $During_{i,j,t}$ is a dummy variable, equals to 1 when school i has an editor at journal j at year t. $During_{i,j,t}$ is a dummy variable that equals to 1 if school i has an editor at journal j at year $t-2$, and 0 otherwise. We estimate the following regressions $\log(\overline{Publication}_{i,j,t}) = \beta Dur_{i,j,t} + FE + \epsilon_{i,j,t}$ where FE are various fixed effects. We include all schools who ever published at *Journal of Finance*, *Journal of Financial Economics*, and *Review of Financial Studies* in the sample. Robust standard error are clustered at school level. T tests of statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Dependent variable: $\log(\overline{Publication}_{i,j,t} + 1)$					
	(1)	(2)	(3)	(4)	(5)	(6)
$During\_i, j, t$	0.878***	-0.014	-0.015	-0.002	-0.075	0.405***
	(12.11)	(-0.27)	(-0.29)	(-0.03)	(-1.34)	(6.67)
logpub						0.500***
						(17.63)
Constant	0.119***	0.075***	0.153***	0.095***	0.103***	0.065***
	(12.45)	(11.35)	(17.83)	(14.55)	(40.55)	(15.21)
Journal FE	No	Yes	No	No	Yes	Yes
Year FE	No	Yes	No	Yes	No	No
School FE	No	Yes	Yes	No	No	No
Journal-Year FE	No	No	Yes	No	No	No
Journal-School FE	No	No	No	Yes	No	No
School-Year FE	No	No	No	No	Yes	No
Observation	57,015	57,015	57,015	57,015	57,015	57,015
R^2	0.037	0.580	0.585	0.601	0.847	0.304

TABLE C.12: Citations of connected articles

This table establishes that connected articles have more citations. The dependent variable in Panel A is $\log(C_{i,T}^{Fina} + 1)$ where $C_{i,T}^{Fina}$ is the total stock of citations from top finance journals for article i as of year $T = 2014$. The dependent variable in Panel B is $\log(c_{i,j,t} + 1)$ where $c_{i,j,t}$ is the citations for article i in journal j at year t. The dummy variable $Connected_{i,J(i),0}$ equals 1 if the article i's author's affiliations when it is published at year t overlap with its publication journal $J(i)$'s editors' affiliation at year $t-2$, 0 otherwise. $\log(CitationStock_{i,t})$ is the log of stock citations for article i before year t. $ArticleAge_{i,t}$ is the age of publication i at year t. $NumberofAuthor$ is the number of authors for article i, and $Citationcountsforthepast5years$ is the citations counts for the past 5 years for the best author, defined by the citation-publication ratio. School Fe is measured at article i level, which is the best school among within authors' affliations, defined by the cumulative citation and publication ratio at the publication year of article i. T statistics in the parenthesis, statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard error clustered at article level.

Panel A: Dependent Variable: $\log(C_{i,T}^{Fina} + 1)$	(1)	(2)	(3)	(4)	(5)
$Connected\_i, J(i), 0$	0.308***	0.248***	0.042	0.165***	0.073
	(5.92)	(5.89)	(0.90)	(3.08)	(1.25)
Times Cited Count: Last 5 Years	0.002***	0.001***	0.001***		
	(21.66)	(18.86)	(14.03)		
Number of Authors	-0.096***	-0.018	-0.031**	-0.025	-0.046**
	(-5.70)	(-1.22)	(-2.11)	(-1.27)	(-2.21)
$Editor\_i, J(i), 0$	0.151	0.012	0.044	-0.070	-0.067
	(0.96)	(0.09)	(0.34)	(-0.42)	(-0.41)
Constant	1.588***	0.484***	0.775***	1.144***	1.078***
	(43.32)	(6.42)	(9.91)	(7.12)	(5.97)
Author FE	No	No	No	Yes	Yes
School FE	No	No	Yes	No	Yes
Journal of Pub.-Year of Pub. FE	No	Yes	Yes	Yes	Yes
Observation	7,092	7,092	7,092	7,097	7,097
R^2	0.087	0.425	0.485	0.636	0.666

Panel B: Dependent Variable: $\log(c_{i,j,t}^{Fina} + 1)$	(1)	(2)	(3)	(4)	(5)
$Connected\_i, J(i), 0$	0.007	0.007	0.003	0.007	0.003
	(1.05)	(1.05)	(0.45)	(1.02)	(0.45)
$1\_j = J(i)$	0.031***	0.022***	0.031***		
	(19.84)	(13.98)	(19.84)		
$\log(CitationStock)$	0.099***	0.099***	0.098***	0.099***	0.098***
	(37.15)	(37.15)	(37.02)	(37.16)	(37.01)
$ArticleAge$	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***
	(-43.18)	(-43.17)	(-44.63)	(-43.25)	(-44.62)
$Editor\_i, J(i), t$	0.065**	0.065**	0.061*	0.064**	0.061*
	(1.99)	(1.99)	(1.91)	(1.99)	(1.91)
Constant	0.104***	0.106***	0.101***	0.129***	0.111***
	(25.70)	(27.23)	(32.09)	(31.92)	(37.44)
Year of Cit. FE	Yes	No	No	Yes	No
Journal of Cit. FE	Yes	Yes	No	No	No
Journal of Cit. FE	Yes	No	Yes	No	No
Journal of Cit.-Year of Cit. FE	No	Yes	No	No	No
Journal of Pub.-Year of Cit. FE	No	No	Yes	No	No
Journal of Cit.-Journal of Pub. FE	No	No	No	Yes	No
Journal of Cit.-Journal of Pub.-Year of Cit. FE	No	No	No	No	Yes
Observation	362,600	362,600	362,600	362,600	362,600
R^2	0.169	0.179	0.178	0.171	0.192

TABLE C.13: Citations After the Office: Poisson Regression

This table shows the editor step-down effect. The dependent variable is $c_{i,j,t}$ where $c_{i,j,t}$ is the total number of citation from author i at journal j at year t. We restrict journal $j \in \{$*Journal of Finance*, *Journal of Financial Economics*, *Review of Financial Studies*$\}$. The dummy $Editor_{i,j,t}$ equals 1 if author i is an editor at journal j at year $t-2$. The dummy n *Year Post Editor* equals 1 if $Editor_{i,j,t-n} = 1$ and $Editor_{i,j,t-1} = Editor_{i,j,t-2} = \cdots = Editor_{i,j,t-n} = 0$. The dummy n *Year Pre Editor* equals 1 if $Editor_{i,j,t+n} = 1$ and $Editor_{i,j,t+1} = Editor_{i,j,t+2} = \cdots = Editor_{i,j,t+n} = 0$. $Yearsincefirstcitation$ is a variable count the number of years since author i's first citation, which approximates the academic age. $\log(CitationStock_{i,t})$ is the log of stock citations for article i before year t. Robust standard error are clustered at author level. T tests of statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Dependent variable: $c_{i,j,t}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
main							
5 Year Pre Editor	-0.053	0.070	0.213**	0.064	0.100	0.101	-0.098
	(-0.28)	(0.49)	(2.38)	(0.55)	(0.68)	(0.74)	(-0.59)
4 Year Pre Editor	0.063	0.204*	0.340***	0.160	0.190	0.230	0.110
	(0.38)	(1.65)	(4.57)	(1.50)	(1.35)	(1.59)	(0.92)
3 Year Pre Editor	0.023	0.211	0.334***	0.107	0.113	0.230	-0.001
	(0.12)	(1.31)	(3.68)	(0.77)	(0.65)	(1.18)	(-0.01)
2 Year Pre Editor	0.085	0.278*	0.310***	0.154	0.160	0.315	0.040
	(0.45)	(1.71)	(3.60)	(1.12)	(1.15)	(1.58)	(0.30)
1 Year Pre Editor	0.175	0.389***	0.372***	0.304**	0.333***	0.427**	0.085
	(0.97)	(2.80)	(4.82)	(2.51)	(2.66)	(2.45)	(0.62)
Editor	0.434**	0.607***	0.432***	0.481***	0.350***	0.588***	0.324***
	(2.22)	(6.90)	(7.25)	(6.31)	(4.02)	(4.22)	(5.94)
1 Year Post Editor	0.262**	0.451***	0.503***	0.357***	0.434***	0.461***	0.105
	(1.97)	(3.84)	(6.92)	(2.76)	(4.15)	(3.03)	(0.61)
2 Year Post Editor	-0.007	0.210*	0.353***	0.159	0.196*	0.217	-0.093
	(-0.06)	(1.80)	(4.31)	(1.40)	(1.75)	(1.26)	(-0.61)
3 Year Post Editor	-0.090	0.110	0.292***	0.117	0.135	0.119	-0.031
	(-0.62)	(0.83)	(3.32)	(0.97)	(1.42)	(0.73)	(-0.21)
4 Year Post Editor	-0.270	-0.047	0.246***	-0.022	0.008	-0.038	-0.177
	(-1.48)	(-0.32)	(3.27)	(-0.19)	(0.08)	(-0.21)	(-1.22)
5 Year Post Editor	-0.066	0.135	0.300***	0.181	0.201*	0.133	-0.033
	(-0.46)	(1.17)	(4.01)	(1.63)	(1.75)	(0.87)	(-0.22)
6 Year Post Editor	-0.096	0.113	0.263***	0.152	0.192	0.108	-0.148
	(-0.95)	(1.09)	(3.59)	(1.39)	(1.36)	(0.78)	(-1.16)
7 Year Post Editor	-0.195	0.009	0.275***	0.061	0.043	0.004	-0.058
	(-1.34)	(0.07)	(3.50)	(0.50)	(0.34)	(0.03)	(-0.32)
8 Year Post Editor	-0.059	0.135	0.225**	0.157	0.043	0.129	-0.027
	(-0.47)	(1.16)	(2.49)	(1.50)	(0.37)	(0.79)	(-0.22)
9 Year Post Editor	0.107	0.304**	0.356***	0.341***	0.236*	0.295	0.098
	(0.63)	(2.06)	(4.62)	(3.10)	(1.95)	(1.56)	(0.72)
10 Year Post Editor	0.007	0.170	0.334***	0.170*	0.092	0.170	0.038
	(0.05)	(1.49)	(5.29)	(1.85)	(0.81)	(1.14)	(0.34)
$Editor\_i, \bar{j}, t$	0.036	0.245***	0.116***	0.137*	-0.088	0.252***	
	(0.21)	(2.69)	(2.73)	(1.79)	(-1.47)	(3.14)	
$\log(CitationStock)$	0.452***	0.049	0.112***	0.547***	1.061***	0.075	
	(6.20)	(0.95)	(3.19)	(5.01)	(27.22)	(1.51)	
$Yearssincefirstcitation$	-0.011	0.020**	-0.003	0.075***	-0.032***	0.018*	
	(-1.09)	(1.98)	(-0.48)	(2.87)	(-4.74)	(1.94)	
Constant	-0.071	0.799***	-0.121	4.155***	1.850***	0.975***	-0.405***
	(-0.26)	(6.66)	(-1.32)	(6.61)	(3.43)	(8.54)	(-1.37e+09)
Author FE	No	Yes	Yes	Yes	No	No	No
Year FE	No	No	No	Yes	No	No	No
Citation Jou. FE	No	No	Yes	No	No	No	No
Citation Jou.-Year FE	No	No	No	No	Yes	No	No
Author-Citation Jou. FE	No	No	No	No	No	Yes	No
Author-Year FE	No	No	No	No	No	No	Yes
Observation	5,977	5,977	5,977	5,977	5,977	5,977	5,977
R^2							

TABLE C.14: Citations After the Office: No Two-Year Time Shift

This table shows the editor step-down effect. The dependent variable is $\log(c_{i,j,t}+1)$ where $c_{i,j,t}$ is the total number of citation from author i at journal j at year t. We restrict journal $j \in \{$*Journal of Finance*, *Journal of Financial Economics*, *Review of Financial Studies*$\}$. The dummy $Editor_{i,j,t}$ equals 1 if author i is an editor at journal j at year t. The dummy $n\,Year\ Post\ Editor$ equals 1 if $Editor_{i,j,t-n}=1$ and $Editor_{i,j,t-1}=Editor_{i,j,t-2}=\cdots=Editor_{i,j,t-n}=0$. The dummy $n\,Year\ Pre\ Editor$ equals 1 if $Editor_{i,j,t+n}=1$ and $Editor_{i,j,t+1}=Editor_{i,j,t+2}=\cdots=Editor_{i,j,t+n}=0$. $Yearsincefirstcitation$ is a variable count the number of years since author i's first citation, which approximates the academic age. $\log(CitationStock_{i,t})$ is the log of stock citations for article i before year t. Robust standard error are clustered at author level. T tests of statistical significance: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

	Dependent variable: $\log(c_{i,j,t}+1)$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
5 Year Pre Editor	0.123	0.204*	0.352***	0.167*	0.225*	0.290**	0.163
	(0.99)	(1.92)	(3.30)	(1.69)	(1.99)	(2.53)	(1.24)
4 Year Pre Editor	0.034	0.129	0.271***	0.070	0.119	0.207*	-0.002
	(0.29)	(1.33)	(3.02)	(0.76)	(1.16)	(1.95)	(-0.01)
3 Year Pre Editor	0.107	0.211*	0.358***	0.171*	0.224**	0.294**	0.038
	(0.81)	(1.89)	(3.30)	(1.70)	(2.04)	(2.45)	(0.25)
2 Year Pre Editor	0.207*	0.291***	0.433***	0.198**	0.222**	0.375***	0.109
	(1.85)	(3.25)	(4.53)	(2.30)	(2.21)	(3.30)	(0.95)
1 Year Pre Editor	0.196	0.275**	0.408***	0.176*	0.195*	0.350***	0.048
	(1.57)	(2.55)	(3.50)	(1.82)	(1.73)	(2.66)	(0.40)
Editor	0.481***	0.451***	0.539***	0.381***	0.379***	0.493***	0.299***
	(3.60)	(7.09)	(7.52)	(6.42)	(3.60)	(4.61)	(4.16)
1 Year Post Editor	0.302**	0.423***	0.540***	0.379***	0.298***	0.454***	0.400***
	(2.52)	(4.54)	(5.77)	(4.50)	(2.72)	(4.00)	(3.15)
2 Year Post Editor	0.124	0.251***	0.361***	0.212**	0.108	0.263**	0.153
	(1.10)	(2.74)	(3.67)	(2.46)	(0.95)	(2.51)	(1.16)
3 Year Post Editor	0.376***	0.523***	0.644***	0.461***	0.344***	0.537***	0.473***
	(3.21)	(5.51)	(6.78)	(5.05)	(3.19)	(4.65)	(3.11)
4 Year Post Editor	0.152	0.292***	0.409***	0.218**	0.113	0.306**	0.235*
	(1.33)	(3.06)	(4.05)	(2.48)	(1.04)	(2.53)	(1.85)
5 Year Post Editor	0.052	0.179	0.288**	0.103	-0.042	0.196	0.191
	(0.37)	(1.50)	(2.48)	(0.99)	(-0.36)	(1.48)	(1.20)
6 Year Post Editor	-0.031	0.112	0.225**	0.035	-0.083	0.132	0.064
	(-0.25)	(1.23)	(2.42)	(0.44)	(-0.78)	(1.15)	(0.56)
7 Year Post Editor	0.052	0.195*	0.300***	0.095	-0.037	0.204*	0.118
	(0.41)	(1.97)	(3.13)	(1.02)	(-0.31)	(1.87)	(0.80)
8 Year Post Editor	-0.003	0.149	0.251***	0.082	-0.060	0.155	0.053
	(-0.02)	(1.58)	(2.67)	(0.90)	(-0.49)	(1.49)	(0.46)
9 Year Post Editor	-0.010	0.148	0.242**	0.085	-0.114	0.146	0.178
	(-0.08)	(1.42)	(2.38)	(0.84)	(-0.95)	(1.30)	(1.08)
10 Year Post Editor	-0.086	0.063	0.147	0.011	-0.219*	0.051	0.108
	(-0.62)	(0.54)	(1.28)	(0.10)	(-1.77)	(0.40)	(0.68)
$\log(CitationStock)$	0.297***	0.166***	0.174***	0.329***	0.508***	0.195***	
	(13.86)	(4.57)	(5.06)	(6.45)	(12.53)	(5.60)	
$Yearssincefirstcitation$	-0.014***	-0.007	-0.005	-0.033***	-0.007	-0.008	
	(-3.66)	(-1.29)	(-0.92)	(-4.23)	(-1.30)	(-1.50)	
Constant	0.429***	0.743***	0.848***	0.850***	-0.458***	0.646***	1.233***
	(4.69)	(9.51)	(10.71)	(6.43)	(-3.43)	(8.17)	(96.76)
Author FE	No	Yes	Yes	Yes	No	No	No
Year FE	No	No	No	Yes	No	No	No
Citation Jou. FE	No	No	Yes	No	No	No	No
Citation Jou.-Year FE	No	No	No	No	Yes	No	No
Author-Citation Jou. FE	No	No	No	No	No	Yes	No
Author-Year FE	No	No	No	No	No	No	Yes
Observation	5,979	5,979	5,979	5,979	5,979	5,979	5,979
R^2	0.307	0.589	0.615	0.636	0.545	0.646	0.788

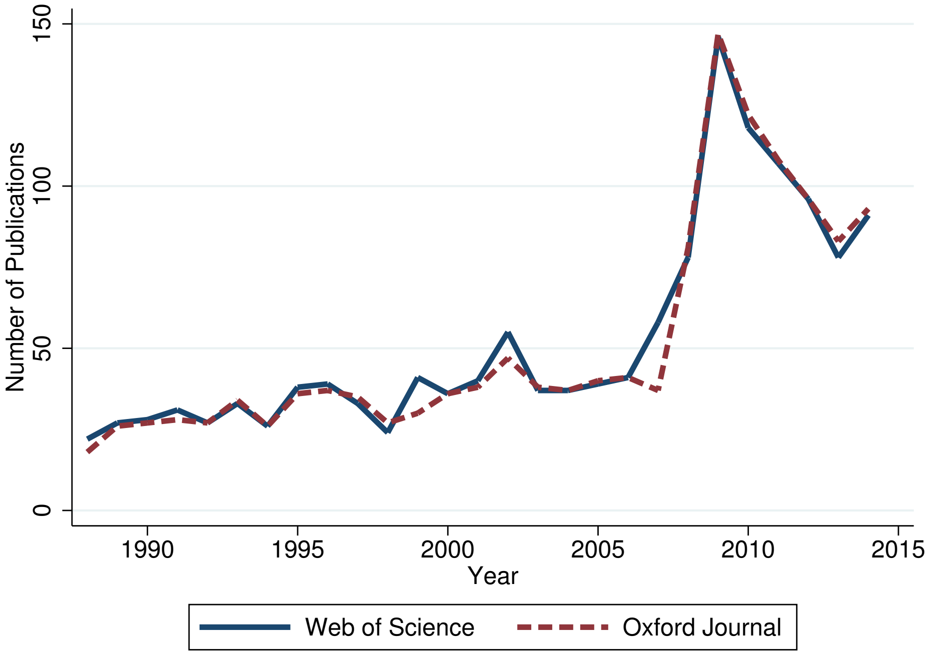
Table C.15: Citations After the Office: Other Specifications

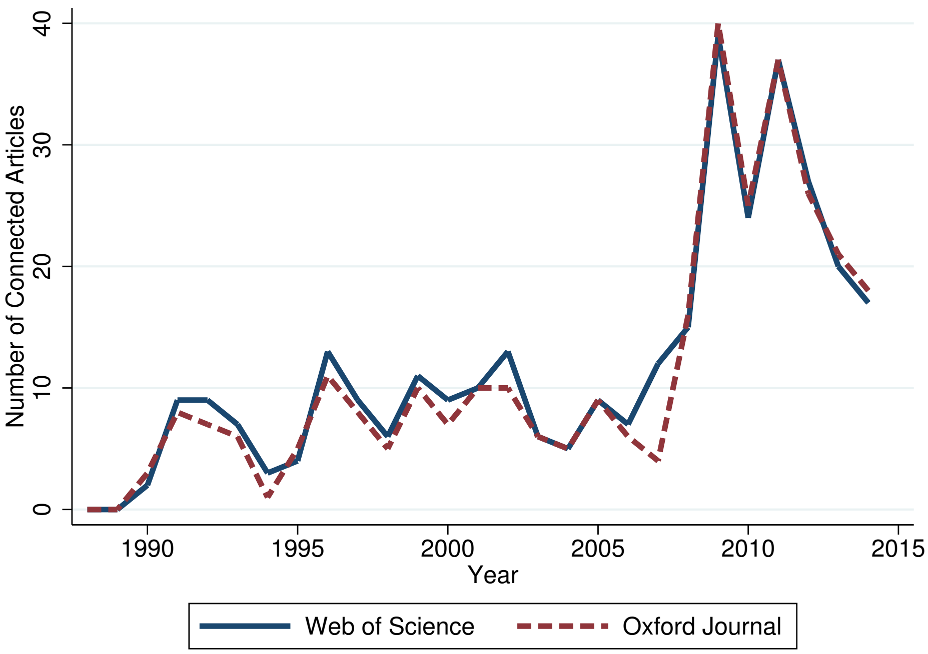
This table shows the editor step-down effect. The dependent variable is $\log(c_{i,j,t}+1)$ where $c_{i,j,t}$ is the total number of citation from author i at journal j at year t. We restrict journal $j \in \{$*Journal of Finance*, *Journal of Financial Economics*, *Review of Financial Studies*$\}$. The dummy $Editor_{i,j,t}$ equals 1 if author i is an editor at journal j at year $t-2$. The dummy $n\,Year\ Post\ Editor$ equals 1 if author i was an editor at journal j at year $t-n-2$ but not from $t-n-1$ to year t. The dummy $n\,Year\ Pre\ Editor$ equals 1 if author i was an editor at journal j at year $t+n-2$ but not from $t+n-3$ to year t. $Yearsincefirstcitation$ is a variable count the number of years since author i's first citation, which approximates the academic age. $\log(CitationStock_{i,t})$ is the log of stock citations for article i before year t. Robust standard error are clustered at author level. T tests of statistical significance: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

	Dependent variable: $\log(c_{i,j,t}+1)$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
5 Year Pre Editor	0.074	0.125	0.246**	0.118	0.166	0.158	-0.062
	(0.54)	(1.11)	(2.30)	(1.12)	(1.42)	(1.49)	(-0.38)
4 Year Pre Editor	0.238**	0.278***	0.399***	0.218**	0.227**	0.312***	0.136
	(2.00)	(2.96)	(3.99)	(2.41)	(2.14)	(2.79)	(1.06)
3 Year Pre Editor	0.214*	0.270**	0.381***	0.179*	0.165	0.291**	0.047
	(1.67)	(2.48)	(3.30)	(1.78)	(1.43)	(2.30)	(0.36)
2 Year Pre Editor	0.180	0.233**	0.348***	0.158	0.152	0.274**	0.030
	(1.54)	(2.22)	(3.16)	(1.61)	(1.42)	(2.17)	(0.24)
1 Year Pre Editor	0.271**	0.327***	0.437***	0.259***	0.215*	0.367***	0.035
	(2.28)	(3.31)	(4.08)	(2.82)	(1.94)	(3.06)	(0.27)
Editor	0.568***	0.514***	0.583***	0.455***	0.420***	0.500***	0.371***
	(3.91)	(7.13)	(7.74)	(7.14)	(4.04)	(4.80)	(5.33)
1 Year Post Editor	0.495***	0.522***	0.627***	0.475***	0.430***	0.528***	0.443***
	(5.11)	(6.01)	(7.26)	(5.65)	(5.11)	(4.93)	(3.02)
2 Year Post Editor	0.242**	0.284***	0.386***	0.219**	0.169*	0.285**	0.173
	(2.58)	(3.05)	(3.91)	(2.54)	(1.98)	(2.37)	(1.32)
3 Year Post Editor	0.177	0.204*	0.302***	0.138	0.053	0.209	0.164
	(1.41)	(1.77)	(2.67)	(1.36)	(0.55)	(1.59)	(1.05)
4 Year Post Editor	0.088	0.128	0.229**	0.063	0.005	0.136	0.038
	(0.83)	(1.40)	(2.49)	(0.79)	(0.06)	(1.16)	(0.33)
5 Year Post Editor	0.185*	0.221**	0.316***	0.135	0.067	0.220*	0.106
	(1.71)	(2.27)	(3.34)	(1.43)	(0.66)	(1.97)	(0.74)
6 Year Post Editor	0.128	0.172*	0.264***	0.106	0.035	0.169	0.028
	(1.36)	(1.88)	(2.87)	(1.19)	(0.34)	(1.61)	(0.24)
7 Year Post Editor	0.136	0.180*	0.268***	0.120	0.007	0.173	0.157
	(1.23)	(1.77)	(2.69)	(1.21)	(0.07)	(1.50)	(0.99)
8 Year Post Editor	0.079	0.109	0.188	0.056	-0.087	0.095	0.087
	(0.64)	(0.96)	(1.65)	(0.52)	(-0.81)	(0.72)	(0.56)
9 Year Post Editor	0.275**	0.308***	0.383***	0.258**	0.119	0.292**	0.170
	(2.22)	(2.85)	(3.57)	(2.47)	(1.12)	(2.35)	(1.26)
10 Year Post Editor	0.277***	0.286***	0.361***	0.212**	0.077	0.274***	0.240*
	(2.93)	(3.42)	(3.92)	(2.31)	(0.75)	(2.73)	(1.70)
$Editor\_i, \bar{j}, t$	0.101	0.099	0.091	0.071	-0.047	0.106**	
	(0.81)	(1.60)	(1.58)	(1.40)	(-0.56)	(2.08)	
$\log(CitationStock)$	0.294***	0.164***	0.174***	0.328***	0.508***	0.195***	
	(13.17)	(4.53)	(5.14)	(6.47)	(12.29)	(5.56)	
$Yearsincefirstcitation$	-0.015***	-0.007	-0.005	-0.037***	-0.008	-0.008	
	(-3.69)	(-1.28)	(-1.05)	(-5.95)	(-1.50)	(-1.55)	
Constant	0.420***	0.738***	0.850***	0.983***	-0.436***	0.650***	1.236***
	(4.80)	(9.59)	(10.88)	(9.24)	(-3.19)	(8.37)	(94.12)
Author FE	No	Yes	Yes	Yes	No	No	No
Year FE	No	No	No	Yes	No	No	No
Citation Jou. FE	No	No	Yes	No	No	No	No
Citation Jou.-Year FE	No	No	No	No	Yes	No	No
Author-Citation Jou. FE	No	No	No	No	No	Yes	No
Author-Year FE	No	No	No	No	No	No	Yes
Observation	5,977	5,977	5,977	5,977	5,977	5,977	5,977
R^2	0.312	0.590	0.614	0.638	0.546	0.646	0.788

Figure C.10: Alternative Data Extraction Method

This graph compare the differences of total number of publications and total number of connected articles across year between two data extraction process for publications in *Review of Financial Studies*. The solid line is from *Web of Science* website and the dash line is from *Oxford Journal* website.





Bibliography

Daniel Ackerberg, Kevin Caves, and Garth Frazer. Structural identification of production functions. MPRA Paper 38349, University Library of Munich, Germany, December 2006.

John Asker, Allan Collard-Wexler, and Jan De Loecker. Dynamic inputs and resource (mis)allocation. *Journal of Political Economy*, 122(5):1013 – 1063, 2014.

Chong-En Bai, Jiangyong Lu, and Zhigang Tao. How does privatization work in China? *Journal of Comparative Economics*, 37(3):453–470, September 2009.

Martin N. Baily, Charles Hulten, and David Campbell. Productivity dynamics in manufacturing plants. *Brookings Papers onEconomic Activity: Microeconomics*, 4:182–267, 1992.

Eric Bartelsman, John Haltiwanger, and Stefano Scarpetta. Cross-country differences in productivity: The role of allocation and selection. *American Economic Review*, 103 (1):305–34, February 2013.

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 10:"p10008", 2008.

Lutz Bornmann and Hans-Dieter Daniel. What do citation counts measure? a review of studies on citing behavior. *Journal of Documentation*, 64(1):45–80, 2008.

Serguey Braguinsky, Atsushi Ohyama, Tetsuji Okazaki, and Chad Syverson. Acquisitions, Productivity, and Profitability: Evidence from the Japanese Cotton Spinning Industry. *American Economic Review*, 105(7):2086–2119, July 2015.

Loren Brandt, Johannes Van Biesebroeck, and Yifan Zhang. Creative accounting or creative destruction? firm-level productivity growth in chinese manufacturing. *Journal of Development Economics*, 97(2):339–351, 2012.

Loren Brandt, Trevor Tombe, and Xiadong Zhu. Factor Market Distortions Across Time, Space, and Sectors in China. *Review of Economic Dynamics*, 16(1):39–58, January 2013.

Jonathan Brogaard, Joseph Engelberg, and Christopher A Parsons. Networks and productivity: Causal evidence from editor rotations. *Journal of Financial Economics*, 111(1):251–270, 2014.

J. David Brown, John S. Earle, and Almos Telegdy. The Productivity Effects of Privatization: Longitudinal Estimates from Hungary, Romania, Russia, and Ukraine. *Journal of Political Economy*, 114(1):61–99, February 2006.

K. Burdett and D. Mortensen. Wage differentials, employer size, and unemployment. *International Economic Review*, 39(2):257–273, 1998.

David Card and Stefano DellaVigna. Nine facts about top journals in economics. Technical report, National Bureau of Economic Research, 2013.

David Card and Stefano DellaVigna. Page limits on economics articles: Evidence from two journals. *The Journal of Economic Perspectives*, pages 149–167, 2014.

Allan Collard-Wexler and Jan De Loecker. Reallocation and Technology: Evidence from the US Steel Industry. *American Economic Review*, 105(1):131–71, January 2015.

Russell Cooper, Guan Gong, and Ping Yan. Costly labor adjustment: General equilibrium effects of china's employment regulations. Working Paper 19324, National Bureau of Economic Research, August 2013.

Russell Cooper, Guan Gong, and Ping Yan. Dynamic labor demand in china: Public and private objectives. *Rand Journal of Economics*, Fall 2015.

Joel M. David, Hugo A. Hopenhayn, and Venky Venkateswaran. Information, Misallocation and Aggregate Productivity. *Quarterly Journal of Economics*, 2015.

Jan De Loecker. Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity. *Econometrica*, 79(5):1407–1451, 09 2011.

Jan De Loecker. Detecting Learning by Exporting. *American Economic Journal: Microeconomics*, 5(3):1–21, August 2013.

Alain Delacroix and Shouyong Shi. Directed search on the job and the wage ladder. *International Economic Review*, 47:651–699, 2006.

Pierre Dubois, Jean-Charles Rochet, and Jean-Marc Schlenker. Productivity and mobility in academic research: Evidence from mathematicians. *Scientometrics*, 98(3): 1669–1701, 2014.

Jan Eeckhout and Philipp Kircher. Sorting and decentralized price competion. *Econometrica*, 78(2):539–574, March 2010.

Liran Einav and Leeat Yariv. What's in a surname? the effects of surname initials on academic success. *The Journal of Economic Perspectives*, pages 175–188, 2006.

Glenn Ellison. The slowdown of the economics publishing process. Technical report, National Bureau of Economic Research, 2000.

Glenn Ellison. How does the market use citation data? the hirsch index in economics. Technical report, National Bureau of Economic Research, 2010.

Daniel R Feenberg, Ina Ganguli, Patrick Gaule, and Jonathan Gruber. It's good to be first: Order bias in reading and citing nber working papers. Technical report, National Bureau of Economic Research, 2015.

Lucia Foster, John C. Haltiwanger, and C. J. Krizan. Aggregate Productivity Growth. Lessons from Microeconomic Evidence. In *New Developments in Productivity Analysis*, NBER Chapters, pages 303–372. National Bureau of Economic Research, Inc, March 2001.

Amit Gandhi, Salvador Navarro, and David Rivers. On the identification of production functions: How heterogeneous is productivity? 2012 Meeting Papers 105, Society for Economic Dynamics, 2012.

Daniel Garcia-Macia, Chang-Tai Hsieh, and Peter J. Klenow. How destructive is innovation? Working paper, 2015.

Zvi Griliches and Haim Regev. Firm productivity in Israeli industry 1979-1988. *Journal of Econometrics*, 65(1):175–203, January 1995.

Daniel S Hamermesh, George E Johnson, and Burton A Weisbrod. Scholarship, citations and salaries: Economic rewards in economics. *Southern Economic Journal*, pages 472–481, 1982.

Thomas J. Holmes, Ellen R. McGrattan, and Edward C. Prescott. Quid pro quo: Technology capital transfers for market access in china. *The Review of Economic Studies*, 2015. doi: 10.1093/restud/rdv008.

Chang-Tai Hsieh and Peter J. Klenow. Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics*, 124(4):1403–1448, November 2009.

Myrto Kalouptsidi. Detection and impact of industrial subsidies: The case of world shipbuilding. Working Paper 20119, National Bureau of Economic Research, May 2014.

Amit K. Khandelwal, Peter K. Schott, and Shang-Jin Wei. Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters. *American Economic Review*, 103(6):2169–95, October 2013.

E Han Kim, Adair Morse, and Luigi Zingales. Are elite universities losing their competitive edge? *Journal of Financial Economics*, 93(3):353–381, 2009.

Molly M King, Shelley J Correll, Jennifer Jacquet, Carl T Bergstrom, and Jevin D West. Men set their own cites high: Gender and self-citation across fields and over time. *women*, 2:1.

David N Laband and Michael J Piette. Favoritism versus search for good papers: Empirical evidence regarding the behavior of journal editors. *Journal of Political Economy*, pages 194–203, 1994.

James Levinsohn and Amil Petrin. Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies*, 70(2):317–341, April 2003.

Dan Lu. Exceptional exporter performance? evidence from chinese manufacturing firms. Working paper, 2010.

Jacob Marschak and Jr. Andrews, William H. Random simultaneous equations and the theory of production. *Econometrica*, 12(3/4):143–205, 1944.

Marc Melitz and Saso Polanec. Dynamic olley-pakes productivity decomposition with entry and exit. *Rand Journal of Economics*, 46(2):362–375, Summer 2015.

Robert K Merton et al. The matthew effect in science. *Science*, 159(3810):56–63, 1968.

G Steven Olley and Ariel Pakes. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64(6):1263–97, November 1996.

Paul Oyer. The macro-foundations of microeconomics: Initial labor market conditions and long-term outcomes for economists. 2006.

Michael Peters. Limits of exact equilibria for capacity constrained sellers with costly search. *Journal of Economic Theory*, 95(2):139–168, December 2000.

Michael Peters. Unobservable heterogeneity in directed search. *Econometrica*, 78(4): 1173–1200, July 2010.

F. Postel-Vinay and J.M. Robin. Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica*, 70(6):2295–2350, 2002.

Derek De Solla Price. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American society for Information science*, page 293, 1976.

Diego Restuccia and Richard Rogerson. Policy distortions and aggregate productivity with heterogeneous plants. *Review of Economic Dynamics*, 11(4):707–720, October 2008.

Diego Restuccia and Richard Rogerson. Misallocation and productivity. *Review of Economic Dynamics*, 16(1):1–10, January 2013.

Shouyong Shi. Frictional assignment. 1. efficiency. *Journal of Economic Theory*, 98: 232–260, 2001.

J Sikorav. The utility of scientific papers. *Scientometrics*, 21(1):49–68, 1991.

Zheng Song and Chang-Tai Hsieh. Grasp the large, let go of the small: The transformation of the state sector in china. *Brookings Papers on Economic Activity*, Forthcoming.

Zheng Song, Kjetil Storesletten, and Fabrizio Zilibotti. Growing Like China. *American Economic Review*, 101(1):196–233, February 2011.

William H Starbuck. How much better are the most-prestigious journals? the statistics of academic publication. *Organization Science*, 16(2):180–200, 2005.

The Communist Party of China. *The Eleventh Five-Year Plan for National Economic and Social Development of the People's Republic of China*. 2006.

Dashun Wang, Chaoming Song, and Albert-László Barabási. Quantifying long-term scientific impact. *Science*, 342(6154):127–132, 2013.

Shing-Yi Wang. State misallocation and housing prices: Theory and evidence from china. *The American Economic Review*, 101(5):2081–2107, 2011.

Natalia Zinovyeva and Manuel Bagues. The role of connections in academic promotions. *American Economic Journal: Applied Economics*, 7(2):264–292, 2015.