Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE
Délivré par l'Université Toulouse 1 Capitole

Présentée et soutenue par

Paloma CARRILLO

Le 10 mai 2023

Essais en économie du développement : Prise de décision intra-familiale et gestion publique

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

Spécialité : **Sciences Economiques - Toulouse**

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

Thèse dirigée par
Stéphane STRAUB et Matteo BOBBA

Jury

Mme Zahra SIDDIQUE, Rapporteure
Mme Janneke PIETERS, Rapporteure
M. Daniel L. CHEN, Examinateur
Mme Roberta ZIPARO, Examinatrice
M. Stéphane STRAUB, Directeur de thèse
M. Matteo BOBBA, Co-directeur de thèse

Essays in Development Economics: Intra-household Decision-Making and Public Management

A dissertation Presented by

Paloma Carrillo

to the

Toulouse School of Economics

for the degree of
Doctor of Philosophy
in the subject of Economics

Toulouse, France
March 2023

Introduction - English

One of the main objectives of development economics is to produce findings that easily translate into policy recommendations. My thesis explores two topics that can influence policies aimed at empowering women and improving the functioning of public institutions in Latin America. In the first two chapters, I investigate the impact of gender norms and exposure to violence on women's decision-making power within households in Mexico. In the third chapter, I examine how reducing information frictions can enhance the use of public administrative data to improve the management and services of courts in Chile.

Although there have been significant advancements in women's empowerment in Latin American countries, particularly in Mexico, there is still a considerable gender gap in various indicators of gender equity. Since most decisions made by women are taken in collaboration with other household members, understanding the determinants of intra-household decision-making and women's bargaining power is crucial. The balance of power between men and women within households not only affects household outcomes but is also an important development goal in itself. Studies have shown that women have different preferences than men, and thus, the person making decisions can have a significant impact on household outcomes. For example, women with higher bargaining power have been shown to have a positive impact on prenatal care, the survival rates of girls, children's education levels, children's nutrition, family health, and reduced child labor. (Thomas et al., 1990; Thomas, 1993; Duflo, 2012; Beegle et al., 2001; Qian, 2008; Chakraborty and De, 2017; Majlesi, 2016; Reggio, 2011).

In the first chapter, I investigate why working mothers in Mexico dedicate an average of eighteen hours more to weekly paid and unpaid work than fathers. Particularly,

I examine the role gender norms play in determining this work time disparity. To do so, I extend a collective labor supply model with household production to include gender norms and estimate it using Mexican survey data from 2002, 2005, and 2009. The model predictions can replicate the changes in total work time disparity over time. I find that more egalitarian gender norms reduce the total work time disparity between spouses and that their impact is comparable with that of wages. For example, a 16 percentage-point increase in a gender norm index between 2002 and 2005 caused a 2.6-hour decrease in total work time disparity mainly through an increase in women's bargaining power. To achieve the same 2.6-hour reduction, women's wages would need to increase by 11% over 2005 levels.

In the second chapter, using longitudinal data on household decision-making in Mexico, I explore the impact of a violence environment, measured by homicides, on spouses' decisions and their bargaining power. I find that an increase in the homicide rate decreases the number of decisions taken by women and men, thus reducing the number of decisions taken jointly. For example, the average increase of 9.3 homicides in the twelve-month homicide rate during the War on Drugs caused couples to decrease the number of joint decisions by 6% from its baseline. The changes in joint decisions represent a reversal into more historical gender spheres of decisions, with men lowering their participation in decisions about children's education and clothing while women reducing their decisions on male private consumption goods and large expenditures. Suggesting that although the reduction in number of decisions was similar for both women and men, community violence might dis-empower women in the household.

The last topic I analyse in this thesis relates to public institutions. In Latin America, there is a wealth of administrative data available from public institutions, which presents both opportunities and challenges. However, a lack of data management and analytic skills is a significant barrier to utilizing this data for evidence-based decision-making in the public sector (OECD, 2014). To address this issue, One particularly promising application of administrative data in the public sector is the use of performance measures to evaluate the efficiency, effectiveness, procedural satisfaction, and productivity of public institutions. While performance measures have been shown to guide management and policy decisions effectively, their use is currently limited

(Hanson and Ostrom, 2014). The third chapter provides evidence that reducing information frictions could improve the use of available data, enabling public officials to make better-informed decisions.

In the third chapter, co-authored with Daniel Chen, Manuel Ramos- Masqueda, and Bernardo Silveira, we examine the extent to which information frictions in management are a meaningful barrier to court productivity in Chile. To do so, first, we randomly promote the use of an online platform featuring court statistics through email campaigns directed towards court managers. Second, within this platform, we simplify the main homepage feedback containing the information on court statistics and randomize managers' access to it. We find that the email promotion and the simplification of the homepage feedback enhance court productivity across multiple indicators. For instance, the new simplified homepage increased case clearance by one standard deviation for those who log into the platform. Additionally, we find that the treatments are more effective for more experienced court managers, who have less accurate baseline beliefs. This suggests that reducing information frictions not only enhances overall court performance but also helps bridge the performance gap between less and more experienced court managers.

Introduction - Français

L'un des principaux objectifs de l'économie du développement est de produire des résultats facilement transposables en recommandations politiques. Ma thèse explore deux sujets qui peuvent influencer les politiques visant à autonomiser les femmes et à améliorer le fonctionnement des institutions publiques en Amérique latine. Dans les deux premiers chapitres, j'étudie l'impact des normes de genre et de l'exposition à la violence sur le pouvoir de décision des femmes au sein des ménages au Mexique. Dans le troisième chapitre, j'examine comment la réduction des frictions informationnelles peut améliorer l'utilisation des données administratives publiques pour améliorer la gestion et les services des tribunaux au Chili.

Bien qu'il y ait eu des avancées significatives dans l'autonomisation des femmes dans les pays d'Amérique latine, en particulier au Mexique, il reste un écart considérable entre les sexes dans divers indicateurs d'équité de genre. Étant donné que la plupart des décisions prises par les femmes sont prises en collaboration avec d'autres membres du ménage, comprendre les déterminants de la prise de décision intra-ménage et du pouvoir de négociation des femmes est crucial. L'équilibre des pouvoirs entre les hommes et les femmes au sein des ménages n'affecte pas seulement les résultats des ménages, mais constitue également un objectif de développement important en soi. Des études ont montré que les femmes ont des préférences différentes de celles des hommes et que la personne qui prend les décisions peut avoir un impact significatif sur les résultats des ménages. Par exemple, il a été démontré que les femmes ayant un pouvoir de négociation plus élevé ont un impact positif sur les soins prénatals, le taux de survie des filles, les niveaux d'éducation des enfants, la nutrition des enfants, la santé de la famille et la réduction du travail des enfants (Thomas et al., 1990 ;

Thomas, 1993 ; Duflo, 2012 ; Beegle et al., 2001 ; Qian, 2008 ; Chakraborty et De, 2017 ; Majlesi, 2016 ; Reggio, 2011).

Dans le premier chapitre, j'étudie pourquoi les mères qui travaillent au Mexique consacrent en moyenne dix-huit heures de plus par semaine au travail rémunéré et non rémunéré que les pères. En particulier, j'examine le rôle des normes de genre dans la détermination de cette disparité de temps de travail. Pour ce faire, j'étends un modèle collectif d'offre de travail avec production ménagère pour inclure les normes de genre et je l'estime en utilisant les données d'enquête mexicaines de 2002, 2005 et 2009. Les prévisions du modèle peuvent reproduire les changements dans la disparité totale du temps de travail au fil du temps. Je constate que des normes de genre plus égalitaires réduisent la disparité totale du temps de travail entre les conjoints et que leur impact est comparable à celui des salaires.Par exemple, une augmentation de 16 points de pourcentage dans un indice de normes de genre entre 2002 et 2005 a entraîné une réduction de 2,6 heures dans la disparité totale du temps de travail, principalement grâce à une augmentation du pouvoir de négociation des femmes. Pour obtenir la même réduction de 2,6 heures, les salaires des femmes devraient augmenter de 11% par rapport aux niveaux de 2005.

Dans le deuxième chapitre, en utilisant des données longitudinales sur la prise de décision au sein du ménage au Mexique, j'explore l'impact d'un environnement violent, mesuré par les homicides, sur les décisions des conjoints et leur pouvoir de négociation. Je constate qu'une augmentation du taux d'homicides diminue le nombre de décisions prises par les femmes et les hommes, réduisant ainsi le nombre de décisions prises conjointement. Par exemple, l'augmentation moyenne de 9,3 homicides dans le taux d'homicides sur douze mois pendant la guerre contre la drogue a entraîné une diminution de 6% du nombre de décisions prises conjointement par les couples par rapport à la base. Les changements dans les décisions conjointes représentent un renversement dans des sphères de décision plus historiquement liées au genre, les hommes réduisant leur participation aux décisions concernant l'éducation et les vêtements des enfants, tandis que les femmes réduisent leurs décisions sur les biens de consommation privés masculins et les grandes dépenses. Cela suggère que bien que la réduction du nombre de décisions soit similaire pour les femmes et les hommes, la violence communautaire

pourrait désavantager les femmes dans le ménage.

Le dernier sujet que j'analyse dans cette thèse concerne les institutions publiques. En Amérique latine, il existe une richesse de données administratives disponibles provenant des institutions publiques, ce qui présente à la fois des opportunités et des défis. Cependant, le manque de compétences en gestion et en analyse de données constitue un obstacle significatif à l'utilisation de ces données pour la prise de décisions fondées sur des données probantes dans le secteur public (OCDE, 2014). Pour résoudre ce problème, une application particulièrement prometteuse des données administratives dans le secteur public est l'utilisation de mesures de performance pour évaluer l'efficacité, l'efficience, la satisfaction procédurale et la productivité des institutions publiques. Bien que les mesures de performance aient été démontrées comme étant efficaces pour orienter la gestion et les décisions politiques, leur utilisation est actuellement limitée (Hanson et Ostrom, 2014). Le troisième chapitre fournit des preuves que la réduction des frictions informationnelles pourrait améliorer l'utilisation des données disponibles, permettant aux fonctionnaires publics de prendre des décisions plus éclairées.

Dans le troisième chapitre, co-écrit avec Daniel Chen, Manuel Ramos-Masqueda et Bernardo Silveira, nous examinons dans quelle mesure les frictions informationnelles dans la gestion constituent un obstacle significatif à la productivité des tribunaux au Chili. Pour ce faire, tout d'abord, nous avons promu de manière aléatoire l'utilisation d'une plateforme en ligne présentant des statistiques de tribunal par le biais de campagnes de courrier électronique dirigées vers les gestionnaires de tribunaux. Ensuite, au sein de cette plateforme, nous avons simplifié la rétroaction principale de la page d'accueil contenant les informations sur les statistiques du tribunal et nous avons aléatoirement attribué aux gestionnaires l'accès à cette information. Nous constatons que la promotion par courrier électronique et la simplification de la rétroaction de la page d'accueil améliorent la productivité des tribunaux selon plusieurs indicateurs. Par exemple, la nouvelle page d'accueil simplifiée a augmenté la résolution des cas d'un écart-type pour ceux qui se connectent à la plateforme. De plus, nous constatons que les traitements sont plus efficaces pour les gestionnaires de tribunaux plus expérimentés, qui ont des croyances de base moins précises. Cela suggère que

la réduction des frictions informationnelles améliore non seulement la performance globale des tribunaux, mais aide également à combler l'écart de performance entre les gestionnaires de tribunaux moins et plus expérimentés.

Contents

Chapter 1

The Effects of Gender Norms on Work Time

1.1 Introduction

Mexico is facing major challenges in its endeavor to achieve greater gender equality. Less than half of women of working age in Mexico participate in the labor market, making it the country with the second-lowest rate of all OECD countries (Gurría, 2019). One of the causes of this low participation is women's excessive unpaid work burden, which prevents them from devoting time to market work. Women who do participate in the labor force work, on average, in paid and unpaid activities 12 hours more per week than their husbands and 18 hours more when the couple has children.[1][2] To what extent can gender norms explain this total work time disparity between spouses? Gender norms are the social norms, informal rules, or shared social expectations that distinguish expected behavior on the basis of gender (Cislaghi and Heise, 2020). Understanding the impact of gender norms on this large inequality within households and its changes over time is important for two main reasons. First, because the amount of work dictates the utility received from leisure and, consequently, gender disparities

[1] Based on data from 2002 to 2011 from the Mexican Family Life Survey.

[2] According to the OECD's Gender, Institutions and Development Database of 2014, Mexico's female to male ratio of time devoted to unpaid care work (excluding housework) was the 7th highest from the OECD countries, with women doing 3.32 times more unpaid care work than men.

can lead to welfare inequalities. Second, understanding total work time determinants would allow policymakers to infer how gender-specific work patterns will change as economies develop, education levels rise, gender norms equalize, or policies are implemented.

The literature on gender norms and total work time disparity has only commented on their correlation (Burda et al., 2013; Campaña et al., 2018). In this paper, I utilize the variation in perceptions of gender norms to study the causal impact of gender norms on total work time disparity between spouses. First, I examine the total work time disparity between spouses alongside multiple gender inequalities by extending a collective labor supply model with public goods to include gender norms and education levels. Most collective models considering public goods are estimated using datasets from developed countries. My extension of the collective model makes it more suitable for a developing country given the high dispersion in gender norms and education levels. Second, I compare the magnitude of the effects of gender norms and other gender differences on total work time disparity by estimating the model using a sample of Mexican couples and allowing for counterfactual analyses.

I begin the paper by building an index for gender norms, which will be used in the collective model, using multiple attitudes on gender roles by peers of spouses. I consider the spouses' peers to be those individuals who match with the wife in the following characteristics: region, level of education (low or high), rural or urban setting, cohort, a proxy for low or high income, and year of the survey. Gender norms are highly correlated with other socioeconomic factors that determine time allocation, such as wages, education, and children's characteristics. To disentangle gender norms' role in determining total work time from other factors, I use a collective model to explore how men and women jointly decide how to spend their time. I begin from the collective labor supply model of Cherchye, De Rock, and Vermeulen (2012), which allows for the home production of two public goods: children and a household good. I extend their model to allow an exogenously determined gender norm index to impact gender preferences and intrahousehold bargaining power. Furthermore, I also extend the model to consider the education level of each spouse. I include spouses' education level because it has been shown to impact child investments, it is highly correlated

with gender norms, and there is a large dispersion in attainment in Mexico. I estimate the model using detailed time usage and expenditure data for Mexican working couples with children from three survey waves in 2002, 2005/6, and 2009/11. The model fits the sample means well and replicates the changes in total work time over time.

The model shows that having peers with more egalitarian gender norms reduces the total work time disparity between spouses on the intensive margin. For example, the gender norm index increased by 16 percentage points during the period between 2002 and 2005/6, which caused the total work time disparity to decrease by 2.6 hours (14%). The changes in total work time arise from women increasing their leisure by reducing market and household hours. In contrast, men reduced their leisure by increasing their market work. More egalitarian gender norms impact total work time through three mechanisms: (1) increasing men's and (2) decreasing women's preference for the two public goods with respect to private consumption and leisure, and (3) increasing women's bargaining power. The impact of gender norms on total hours is primarily driven by the third mechanism regarding shifts in women's bargaining power.

To better understand the effects of gender norms and their magnitude, it is useful to compare their impacts to other factors with gender differences. I use the model to study the impacts of (1) relative wages, (2) gender preferences for leisure, and (3) social stigma costs on the work time disparity of spouses. First, I find that the impact of gender norms is comparable in magnitude or size to changes in relative wages between spouses. For example, to achieve the same 2.6-hour reduction in total work time disparity from an increase in gender norms index (16 pp), women's wages would have to increase by 11 percent. Furthermore, while the impact of gender norms remains sizable across distinct relative wages, the impact of wage changes diminishes as the wages of spouses become more equal. Second, gender differences in preferences have a limited impact on total work time disparity. Gender differences in preferences for leisure are small, with women slightly preferring private consumption to leisure in comparison to men. In a counterfactual analysis, where women and men within a household have the same exogenous variables such as wages and education, and equal bargaining power, the model predicts the total work time disparity would be close to zero. This finding means there are no large discrepancies in how women and men

value private consumption, leisure, and the two public goods. Finally, another gender disparity concerns the social stigma incurred by the household when husbands dedicate time to household work and child care. The model estimates show considerable social stigma costs for both public goods. However, even if there were no social stigma costs and women and men were equally productive in the production of public goods, the large total work time disparity would remain but with different work arrangements. This result indicates that social stigma costs for men's participation in public goods are not a driving factor of total work time disparity in the model but rather of how the hours of work are divided across activities.

This paper falls within the intersection of three fields: total work time disparity, gender norms and household choices, and collective models with public goods or gender norms. Some of the existing research on total work time includes Burda et al. (2013), who find from a 27-country comparison that the gender total work time disparity appears to be significant in developing countries, particularly Mexico, and in developed countries where the predominant religion is Catholicism, such as Ireland and Spain. Campaña et al. (2018) focus on three distinct Latin American countries and find that those with more equal gender norms have similarly more equal total work time distributions. This literature comments on the correlation between gender norms and total work time disparity but does not consider the mechanism through which gender norms affect the disparity or how it may compare with other determinants. In this paper, I distinguish the effect of gender norms through changes in preferences and bargaining power and find that the magnitude of the effects is sizable and, for certain couples, even more impactful than gender wage differences.

Another literature has focused on gender norms and household choices. Studies have shown that gender norms can mediate the role of bargaining power between spouses (Oreffice, 2014), women's labor force participation (Fortin, 2005; Fernández and Fogli, 2009; Cavapozzi et al., 2021), and the number of public good work hours (Hwang et al., 2019). However, the literature on gender norms and time allocation has focused on one type of hour allocation at a time. My model allows for the analysis of the impact of gender norms on the final allocation of spouses' time across different activities and, consequently, total work time disparity.

This paper applies the standard approach of studying joint household decisions through collective models.[3] Chiappori (1992) spearheaded the characterization of the household as a group of agents making joint decisions and modeling it to allow for any type of efficient decision process. Cherchye et al. (2012) extended Chiappori's initial paper to include the decision of hours and expenditures on two domestic goods. They develop a novel estimation strategy using production shifters and distributional factors while building upon the familiar two-stage allocation representation of the collective model. Most collective models considering public goods are estimated using datasets from developed countries.[4] The total work time disparity between spouses in these countries is relatively small (Burda et al., 2013). This may be a reason why this subfield of collective models has not focused on the question of total work time disparity between spouses or included gender norms in the model. More recently, a sub-field of collective models has considered how perceptions of gender roles might affect the joint decision of couples. For example, Field et al. (2019) and Afridi et al. (2019) include gender norms in their collective models as an additional utility cost whenever the wife has positive hours of market work, which limits women's labor force participation but does not consider public good hours. This paper extends collective models to consider public goods and gender norms simultaneously to understand total work time disparity and its determinants. To the best of my knowledge, this is a new contribution to this literature. The inclusion of gender norms better adapts collective models to a developing country such as Mexico.

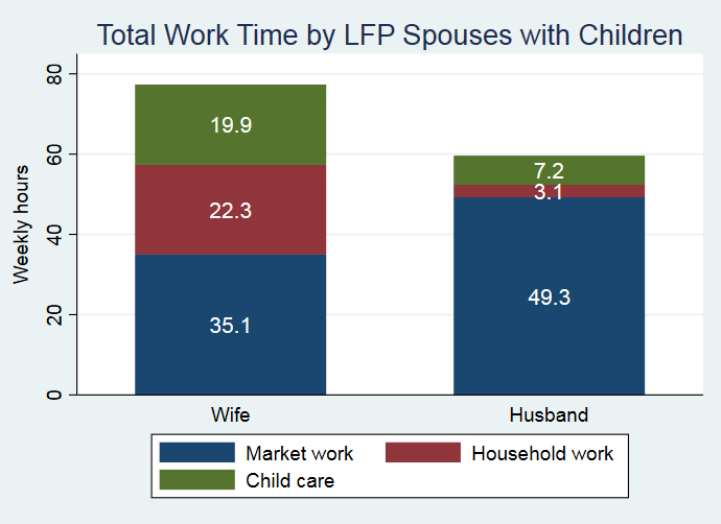
The next section describes the dataset for Mexican couples. In section 3, I build a gender norm index using the gender role perceptions of the couple's peers. Sections 4 and 5 present the collective model and the parametric specifications. Section 6 presents the model estimates, model fit, and counterfactual analyses. Section 7 concludes.

[3] A common framework for studying household behavior, which replaced the unitary model, is the collective approach, whereby household decisions are the results of a bargaining process among its members with specific preferences. For a review of the literature on the collective approach of the past three decades, refer to Donni and Molina (2018).

[4] For example, there are collective modela estimated using datasets from Japan (Lise and Yamada, 2019), the Netherlands (Cherchye et al., 2012), the U.S. (Donni and Matteazzi, 2012), the UK (Van Klaveren et al., 2008), and France (Rapoport et al., 2011).

1.2 Mexican Family Life Survey (MxFLS)

Figure 1.1: Total Work Time Decomposition by LFP Couples with Children



I analyze the question of total work time disparity in Mexico through a model and estimate it with data from the Mexican Family Life Survey (MxFLS). This survey is a longitudinal multithematic survey of households developed and managed by researchers from the Iberoamerican University, the Center for Economic Research and Teaching in Mexico, and Duke University. The first wave of the survey begins with 8,440 households representative of the Mexican population. The data contains information for a 10-year period, collected in three rounds: 2002, 2005–2006, and 2009–2012.

Using the distinct waves, I find that women, on average, work more total hours of paid and unpaid work than their spouses if they participate in the labor market. Particularly, married women participating in the labor market work on average 12 hours more per week than their male partners, and a total of 18 hours more per week if there are children in the household.[5] Given that the largest differences in total work time are seen in couples where both spouses participate in the labor force and have children, besides being the most common category, I chose to focus the paper on these couples.[6] Figure 1.1 shows that the disparity in total work time arises from

[5]For couples where the wife does not participate in the labor force, wives work on average 1 hour less per week than their husbands.

[6]In the MxFLS for those married couples with children living in the household, 71% of them are both in the labor force, in 27% only the husband participates, and in 2% only the wife participates.

women dividing a large number of working hours between market work, child care, and household chores, while men primarily work in the market only.

To focus on labor force participating couples with at least one child in the household, I drop from the sample those headed by an unmarried or widowed individual, those where both spouses are not in the labor force, or those without a child below the age of 18. I also drop any household head that is still a student as they will be dividing their time with studying. Finally, I restrict the sample to households with no other adults living in the household that could be helping with household work or child care. The final sample obtained by pooling the three waves of the survey contains 782 household observations.[7]

For the collective model I will be using, I require information on how time is distributed and expenditures are allocated by the spouses. The surveys include questions about the detailed usage of time in the past week. There were 14 different activities asked about that could be easily grouped into those concerning children, household work, market work, and leisure activities. There are multiple questions about expenditures. For expenditure on private goods, there are questions regarding purchased gender-specific personal products (such as perfume or deodorant), clothing, and footwear. The questions regarding public good expenditures for the house include rent, general hygiene products, cleaning products, utilities, services, and general products (such as pots, plates, and sheets). For the child public good expenditures, there are children's educational costs, clothing, and toy expenditures.

Table 1.1 summarizes expenditures, time use, and other household characteristics over the period 2002-2011 for the 782 household-year observations used for the estimation of the collective model. As seen in Table 1, husbands spend, on average, slightly more on private expenditures than their wives. Expenditures on children are, on average, 350 pesos per month ($\approx$ 33 dollars).[8] Most of the household's expenditures are spent on other public goods averaging 2,097 pesos ($\approx$ 198 dollars). Next, when looking

[7]The sample is small in comparison to the initial sample, particularly because numerous couples were dropped for having at least one missing piece of information for either individuals' time usage, expenditures, income, education, age, or other key variables. Furthermore, given the small sample, only a few of the households are repeated throughout the waves. For this reason, I pool the data for estimation.

[8]All expenditures are deflated to 2005 levels

Table 1.1: Summary Statistics for Labor Force Participating Couples with Children

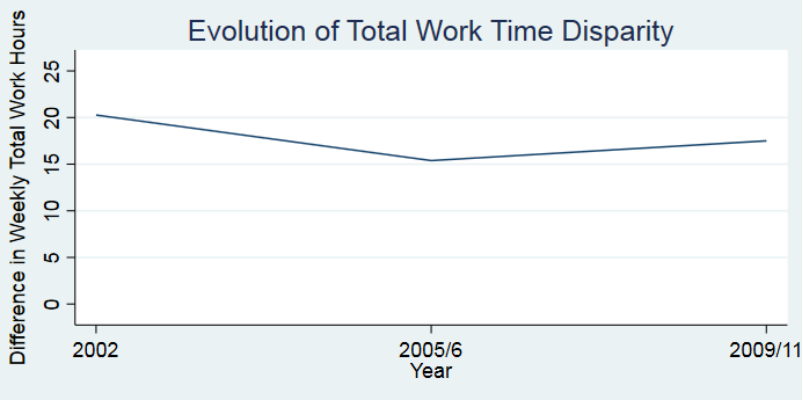
	All years	2002	2005/6	2009/11
Mean Expenditures (Monthly)				
Men expenditures	347	570	262	217
Women expenditures	300	379	262	261
Children expenditures	350	394	230	413
Hhld expenditures	2097	2297	1927	2059
Mean Time use (hrs per week)				
Men market labor	49.3	48.4	47.6	51.5
Women market labor	35.1	34.2	35.0	35.9
Men child care	7.2	8.9	6.1	6.6
Women child care	19.9	23.5	15.3	20.6
Men domestic work	3.1	3.9	2.3	3.0
Women domestic work	22.4	23.7	21.1	22.2
Socioeconomic variables				
Men's age	36.7	37.1	37.3	35.7
Women's age	34.2	33.9	34.8	34.0
Men's wage rate	22.1	22.2	21.6	22.4
Women's wage rate	20.8	19.2	22.2	20.9
Number of children	2.3	2.3	2.4	2.1
Mean age children	8.7	8.5	9.3	8.4
Men highly educated	0.39	0.40	0.36	0.39
Women highly educated	0.36	0.29	0.34	0.46

Note: 782 household-year observations from the 2002, 2005-2006, and 2009-2011 MXFLS surveys. Wages and expenditures are in Mexican pesos and normalized to year 2005.

at spouses' time use, husbands spend substantially more hours on market work than their wives: on average, husbands work 49.3 hours per week in the market (including commuting) and wives work 35.1 hours. In contrast, women spend on average 19.9 hours on child care and 22.4 hours on domestic work, while their husbands only spend on average 7.2 hours on child care and 3.1 hours on domestic work. This results in a total work time disparity across all waves of 17.7 hours.

Figure 1.2 presents the average total work time disparity for each wave. In 2002, the total work time disparity in the sample was at its highest, with an average disparity of 20.2 hours per week; then, for 2005/6 it fell to 15.4 hours and in 2009/11 to 17.6 hours per week. The drop of 4.9 hours from 2002 to 2005/6, although a 24% sizable reduction in the total work time disparity, still leaves a large disparity between genders.

Figure 1.2: Evolution of Total Work Time through MxFLS waves



Household characteristics needed for the estimation are the number of children averaging 2.3 and the mean age of children averaging 8.7 years. Other individual characteristics used are the education level and individual wages. Only 39% of men and 36% of women have an education level higher than high school. Women, on average, earn less per hour; men's hourly wage averages 22 pesos ($\approx$ 2 dollars) and women's 21 pesos.

Table 1.2: Gender Norm Questions from ENDIREH

	Survey Questions
i.	A good wife needs to obey his husband in everything he says
ii.	A woman can choose her friends even if the husband dislikes them (inverted)
iii.	The man is responsible for all the expenditures of the family
iv.	A woman has the same capacity as a man to earn money (inverted)
v.	It is the obligation of women to have sexual relations with her husband even if she does not want to
vi.	When a woman does not follow her obligations, the husband has the right to hit her

	i	ii	iii	iv	v	vi
From 100 women, those responding "no"						
2002	43	41	25	28	89	90
2005/6	66	31	35	20	94	96
2009/11	81	03	36	14	84	98

Note: Yes=0 or No=1 responses. Means for the three waves of MxFLS couples. For those noted as inverted, I took the (1-answer) before taking the first component of the principal component analysis.

1.3 Gender Norms in Mexico

As Cislaghi and Heise (2020) defines them, gender norms are "social norms defining acceptable and appropriate actions for women and men in a given group or society." This means that these informal rules or shared social expectations define the array of appropriate behavior for each gender in a given population. To analyze the role of gender norms in determining total work time disparity, I first build an index for gender norms. Given that the household dataset of the Mexican Family Life Survey (MxFLS) does not directly ask spouses about gender norm perceptions, I resort to another dataset, the National Survey for the Dynamics in Household Relationships (ENDIREH). The survey has three cross-sectional waves for 2003, 2006, and 2011, which match well with the MxFLS couples of 2002, 2005/06, and 2009/11. The ENDIREH surveyed women in 57,230 households in 2003, 127,944 in 2006, and 128,000 in 2011.

All respondents to the ENDIREH waves are asked questions designed to elicit their gender role attitudes or perceptions. Specifically, they are asked if they agree with the following statements: (i) "A good wife needs to obey his husband in everything he says"; (ii) "A woman can choose her friends even if the husband dislikes them"; (iii) "The man is responsible for all the expenditures of the family"; (iv) "A woman has the same capacity as a man to earn money"; (v) "It is the obligation of the women to have

sexual relations with her husband even if she does not want to"; (vi) "When a woman does not follow her obligations, the husband has the right to hit her". Responses are rated "0-yes" and "1-no". I define each woman's gender role index by summarizing the responses with the first component obtained from a principal component analysis after inverting the scale for questions (ii) and (iv). From this analysis, I extracted the first principal component, and from that, I used the factor loadings as weights for the questions. Hence, the weights assigned to each attitude are 0.54 for attitude i, 0.36 for attitude ii, 0.40 for attitude iii, 0.41 for attitude iv, 0.39 for attitude v, and 0.31 for attitude vi.[9] When we applied these weights to the attitudes selected, we obtained a value for each respondent in the ENDIREH samples.

Creating a gender norm index through the principal component of survey responses is a standard approach in the literature of gender norms. For example, this strategy is used in the recent papers of Campaña et al. (2018) and Cavapozzi et al. (2021). Particularly, Campaña et al. (2018) created a gender norms index for Mexico using the 2010-2014 wave of the World Values Survey (WVS). The questions in the WVS of the 2010-2014 wave are similar to those in ENDIREH waves, yet these questions are not present in the WVS 2000 and 2005 waves. For this reason, as well as the smaller sample size of the WVS, I use the ENDIREH waves. In a sensitivity analysis, I sum responses across the six responses. The estimates and model predictions are similar to those presented in the results section and therefore are not presented.

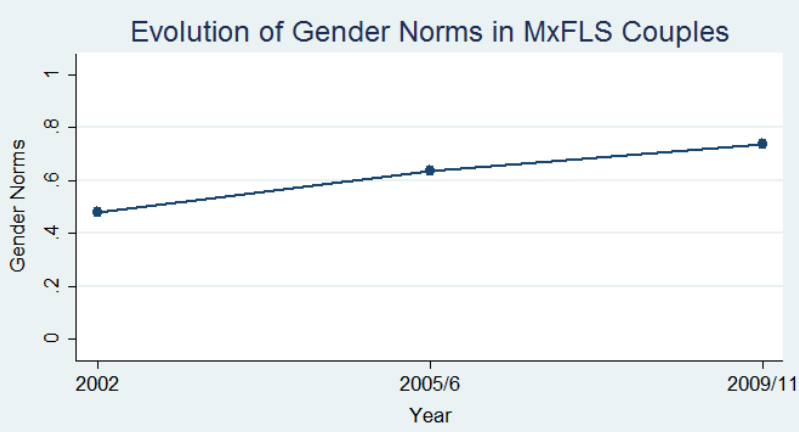
The gender norms of couples in the MxFLS are proxied using the average of individual ENDIREH gender role index across their peers, where peers are defined as women born in the same cohort, education level, number of rooms in the house, rural or urban setting, Mexican region, and year as the focal women in the MxFLS couple. Birth cohorts are divided into 5 groups: women born in 1959 or earlier, 1960-1969, 1970-1979, 1980-1989, and 1990-1999.[10] Education is stratified into low educated (lower than high school attainment) and high educated. Number of rooms is grouped for those

[9]The Kaiser-Meyer-Olkin measure of sampling adequacy for our gender role perception variables is 0.69, which is above the recommended minimum of 0.6 for justifying using principal component analysis. This measure is a statistic that indicates the proportion of variance in the variables that might be caused by underlying factors.

[10]The cohort from 1990-1999 only used for the third wave.

that have two rooms or fewer, and those with more than two rooms. For the peer groups using the six categories that happen to have fewer than 25 respondents, these peer groups were joined with the peer group that is equal in all the categories but the number of rooms.[11] Finally, the index is standardized between 0 and 1, with higher values indicating more egalitarian attitudes between the genders and lower values capturing more traditional gender norm attitudes. With zero being the most unequal peer gender norm score found in the combination of the six characteristics (1950s cohort, low educated, rural setting, southern region, fewer than three rooms, and year 2003). The most egalitarian gender norm score was normalized to one (1990s cohort, highly educated, urban setting, Central and Northwest Mexico, more than two rooms, and year 2011).

Figure 1.3: Gender Norms of MxFLS Couples by Year



Using the six characteristics of cohort, education, rural/urban, number of rooms, region, and year, there are 184 distinct peer gender norms indices that are matched to the 782 MxFLS couples with an average size of the peer group of 741 women. The average peer gender norm for the MxFLS sample is 0.62 with a standard deviation of 0.18. Figure 1.3 presents the evolution of gender norms for the 782 MxFLS couples. There were large changes through the three waves; for example, the gender norm index changed by 16 percentage points (33%) between 2002 and 2005/6 and by 10 percentage points (15%) between 2005/6 and 2009/11. The large changes in the gender norms index in such a short amount of time are mainly due to two reasons: the

[11]Number of rooms was chosen as the joining feature given that it has the lowest explainatory power for gender norms.

decade in question and the types of questions. The first is that this period of time saw larger structural changes in society and gender roles than in other decades. For example, the years between 2002 and 2006 experienced a 4.1 percentage point increase in women's labor force participation, a sign of more egalitarian gender norms, the largest of any four years between 1990 and 2020 (International Labour Organization, 2021). Second, the questions of the survey are of a conservative nature and particularly capture the kinds of changes to gender roles that occurred during this time. The change in responses for each question between each wave is shown in Table 1.2. For example, in 2003, only 43% of women did not agree with the statement "A good wife needs to obey his husband in everything he says"; by 2006, it had increased to 66%. Other variables associated with gender norms, such as wages and education levels, also saw changes but not as drastic. For example, between 2002 and 2005/6, women's median wage rose by 11%, and percentage of highly educated women rose by 5 percentage points (18%). In the paper, I assume the gender norm index to be an exogenous variable. Given that the changes in gender norms were twice as large as than those of other related variables and that the period of analysis is shorter than a decade, I find this a reasonable assumption that allows for a tractable model.

What is causing the changes in gender norms? To answer this questions, we first must consider the following: are individual gender norm perceptions fixed? If so, changes in the gender norm index come solely from new generations entering the population; thus, changes to the gender norms of peers are unlikely to directly impact already grown individuals. The literature on gender norms by social theorists and social psychologists agrees although gender norms are learned in childhood from parents and peers, they are continually reinforced or contested through social interactions (West and Zimmerman, 1987; Cislaghi and Heise, 2020). This updating of gender norm perceptions can be seen by following the cohorts in the ENDIREH sample. Take, for example, three different cohorts those born between 1950-1959, 1960-69 and 1970-1979. First, as expected, in any single survey year, the younger cohort of 1970-1979 has more egalitarian gender norm perceptions but more interesting in each successive survey cohorts as they age become more egalitarian in their perceptions. Thus, individuals update their perceptions as they interact with people, institutions,

Table 1.3: Correlations of Gender Norms with Household Characteristics

	BP_{mw}	BP_{mm}	$Wages_w$	$Wages_m$	Edu_m	Edu_w	No. kids	Kids avg age	TWTD
GN	0.003	-0.031	0.279	0.219	0.284	0.729	-0.145	-0.169	-0.076
TWTD	0.011	-0.003	-0.012	0.183	0.059	-0.004	-0.0001	-0.113	1.00

Note: BP_{mw} and BP_{mm} stand for men's bargaining power according to women's and men's responses.

and culture such as television programs. This finding means that as gender norms change, they have the possibility of impacting all individuals.

Gender norms are highly correlated with other household socioeconomic characteristics that determine time allocation. Table 1.3 presents the correlation between gender norms, men's bargaining power as reported by women and men, spouses' wages, education level, the number of children, the average age of children and the total work time disparity (wife's hours minus husband's hours). Men's bargaining power index is proxied using responses from spouses on the MxFLS on the relative decisions husbands make compared to those made by the wife. The index is built based on responses by the husband and wife, which do not always coincide. I create this index only for the purpose of this correlation exercise, while for the collective model, the bargaining power is fitted using the sample observations and the estimated parameters obtained from the model. Gender norms are negatively correlated with men's bargaining power (according to husband responses), number of children, children's average age and total work time disparity. In contrast, gender norms are positively correlated with wages and education levels of both husbands and wives. Total work time is positively correlated with men's wages and men's education level and negatively correlated with the children's average age. Given the high correlation between wages, education, and children's characteristics, a model is needed to disentangle how each factor contributes to the spouses' allocation of time.

1.3.1 Gender Norms in a Collective Model

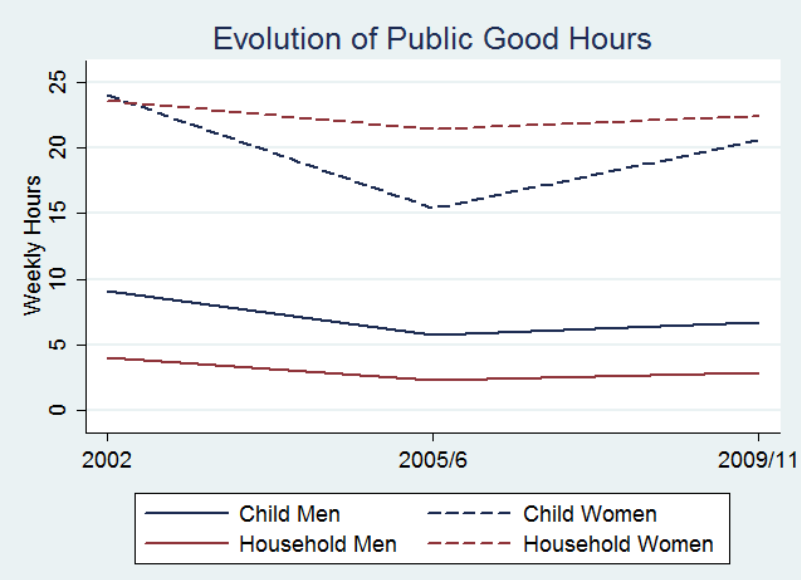
With the gender norm index built for the MxFLS couples, I now focus on how to introduce it into the collective model. A static collective labor supply model with household production can be thought of as having three main components. The model

consists, first, of preferences that depend on private consumption, leisure, and the stock of public goods; second, bargaining power; and finally, the household production functions for each public good that have as inputs spouses time and expenditures. Gender norms could enter into the collective model through any of these three channels.

One might imagine that as gender norms become more egalitarian, first, women might hold greater intra-household bargaining power. Second, we could think that as gender norms change, women and men shift their preferences. The most common traditional household gender perception is that the wife is responsible for children and household upkeep while the husband is responsible for the income. A change in this perception could be manifested as women no longer feel their full attention has to be devoted to children and the household, shifting some preferences towards leisure and private consumption (with men having the opposite change). Another change could be manifested as women and men exchange some hours between public goods and market work, relaxing the gender division of labor. Such a change would be captured by the household production function. In particular, the household production function captures both the public good productivities and the social stigma costs associated with a particular gender performing domestic work or child care. As gender norms change, these stigma costs could also shift.

For the last possibility of relaxing the division of labor through changes in social stigma costs, it is not supported by the data as men spend fewer hours on the two public goods than before. As seen in Figure 1.4, husbands dedicated 3.9 hours to housework per week in 2002; but by 2005/6, this average had decreased to 2.3 hours and to 3.0 hours in 2009/11. For child care, men shifted from an average of 8.9 weekly hours in 2002 to 6.1 and 6.6 weekly hours in the last two survey waves. Similarly, women seem to have decreased their total number of hours spent in child care and household work, particularly for the second wave in 2005/6. Given the patterns of hours spent in public goods, I allow changes in the model such that gender norms may change preferences and bargaining power. However, given that there is a decrease in the hours spent by husbands on the public goods, I consider the household production functions that absorb social stigma costs to be fixed in the medium-term window of time I am analysing.

Figure 1.4: Evolution of Public Good Hours by Gender



One possible reason for the nonincreasing time in public hours by the husbands, even though gender norms are becoming more egalitarian in Mexico, is that these changes might first affect bargaining power and preferences within the household; however, the changes may not be drastic enough to change the social stigma internalized by the couple associated with men participating in unpaid domestic work and child care. That is, larger changes in gender norms, possibly over a longer period of time, may be needed to see reductions in these stigma costs and an increase in men's participation in unpaid work. I present in the next section a static collective model and how gender norms are allowed to affect it.

1.4 A Collective Labor Supply Model with Home Production

My model begins from Cherchye, De Rock and Vermeulen's (2012) collective labor supply model, which focuses on households with two spouses $i = \{1, 2\}$ making joint decisions on time use and expenditures. The model takes real wages as exogenous and allows for the production of two public goods: children and a household good. Children do not have any bargaining power. The model abstracts from fertility and

divorce decisions.

Each spouse must allocate time between leisure l, market work m, child care hours k, and domestic household work p. For each spouse i, the time budget constraint is:

$$l_i + m_i + k_i + p_i \leq 1 \tag{1.1}$$

where the time endowment is normalized to one.

To determine the budget, each spouse i receives a wage w for each unit of market work m, and the household receives nonlabor income y. The income of the household is allocated to a Hicksian composite good with a normalized price of one. The composite good can be used for the private consumption of both spouses, c_1 and c_2, expenditures on the children c_k, and expenditures on the household public good c_p, resulting in the following household budget constraint:

$$c_1 + c_2 + c_k + c_p \leq y + w_1 m_1 + w_2 m_2 \tag{1.2}$$

I assume there are no savings.

The preferences of each spouse are represented by the following utility function:

$$u_i = u_i(c_i, l_i, f_k(c_k, k_1, k_2; \mathbf{s_k}), f_p(c_p, p_1, p_2; \mathbf{s_p}), E_i, GN) \tag{1.3}$$

c_i and l_i are private goods that only affect the utility of spouse i. The domestic good f_k stands for the children's utility that acts as a public good for the spouses. Similarly, f_p is the public good that can be interpreted as the joy of having a clean house. E represents the education level of individual i and GN the couple's gender norms; here, I have my first deviation from Cherchye, De Rock, and Vermuelen's model with the inclusion of education and gender norms within the utility function. I allow gender norms, GN, to impact the preferences of both spouses (details discussed in Section 5). I also allow the preferences to change with education, as in Mexico, the disparity of educational attainment is large, and educational level has been found in previous literature to impact the investment in children (Guryan et al., 2008). I assume that the

function $u(c_i, l_i, f_k, f_p, E_i, GN)$ is twice continuously differentiable, strictly increasing, and strongly concave in all its arguments. Furthermore, the production functions f_k and f_p are assumed to be twice continuously differentiable in all their arguments and to be strictly increasing and strongly concave in the arguments $c_k,\ k_1,\ k_2$ (respectively $c_p,\ p_1,\ p_2$). I also assume that the household production technologies have constant returns to scale.

The vectors $\mathbf{s_k}$ and $\mathbf{s_p}$ in (3) are the production shifters associated with the domestic goods f_k and f_p. A production shifter is a variable that affects individual preferences only through at least one of the household production technologies. For example, the average age of children or the age of the youngest child; it can be argued that children of a small age require more care as they are quite dependent. I assume that there is at least one production shifter for one of the two domestic goods.

I assume the spouses decide how to divide their income and time by choosing the Pareto efficient intrahousehold allocations that are the result of the following optimization program:

$$\max_{l_1,l_2,k_1,k_2,p_1,p_2,c_1,c_2,c_k,c_p} \lambda(w_1, w_2, y, GN, \mathbf{z})\ u_1(c_1, l_1, f_k(c_k, k_1, k_2; \mathbf{s_k}), f_p(c_p, p_1, p_2; \mathbf{s_p}), E_1, GN)$$
$$+ \quad (1 - \lambda(w_1, w_2, y, GN, \mathbf{z}))\ u_2(c_2, l_2, f_k(c_k, k_1, k_2; \mathbf{s_k}), f_p(c_p, p_1, p_2; \mathbf{s_p}), E_2, GN)$$

subject to

$$c_1 + c_2 + c_k + c_p \leq y + w_1 m_1 + w_2 m_2 \tag{1.4}$$

$$l_i + m_i + k_i + p_i \leq 1 \text{ for } i \in \{1, 2\}$$

The Pareto weight $\lambda(w_1, w_2, y, GN, \mathbf{z})$ represents the relative bargaining power of spouse 1, and $(1-\lambda(w_1, w_2, y, GN, \mathbf{z}))$ represents the power of spouse 2. The bargaining power depends on their exogeneous wages w_1 and w_2, the household's nonlabor income y, gender norms GN, and a vector of distribution factors $\mathbf{z}$.[12] The distribution factors

[12]I do not allow for the women's wages to increase with more egalitarian gender norms. The assumption of exogenous wages may potentially affect my gender norm estimates, since in a world where women's wages depend on gender norms women may reduce total work time directly from the impact of more egalitarian gender norms on preferences and bargaining power but also indirectly from an increase in women's wages. Both more egalitarian gender norms and higher women's wages decrease the total work time disparity. In the case that wages are endogenous, then the impact of gender norms would be underestimated.

are variables affecting the bargaining position of the spouses without affecting the preferences or budget constraint after controlling for total income.[13] For example, the difference in age between the spouses could be argued to affect the bargaining between them while not changing their preferences. Within the bargaining power, the second distinction from Cherchye, De Rock and Vermeulen's model is made by including gender norms GN in the intrahousehold bargaining power equation.

The maximization problem in equation (4) can be understood as a two-stage allocation process. In the first stage, spouses decide on the level of domestic goods f_k and f_p and an intrahousehold allocation of the residual nonlabor income $(y - c_k - c_p)$. I denote spouse i fraction of this residual nonlabor income by ρ_i $(i = 1, 2)$, also known as the conditional sharing rule. In the second stage, each spouse maximizes her or his utility by choosing his or her own leisure and private consumption conditional on the levels of both domestic goods and the budget constraint defined in the first stage $\bar{f}_k$ and $\bar{f}_p$.

I assume that the Pareto weight is continuously differentiable in all its arguments, which implies the use of continuous distributional factors. Let the vector $\mathbf{s}$ contain the different production shifters in $\mathbf{s_k}$ and $\mathbf{s_p}$. Then, the household's optimal choices are observable functions of the following variables: spouses' wages w_1 and w_2, the household's nonlabor income y, education levels E_1 and E_2, gender norms GN, the distributional factors $\mathbf{z}$, and the production shifters $\mathbf{s}$. For spouse $i = \{1, 2\}$, we have that:

$$\begin{aligned}
l_i &= l_i(w_1, w_2, y, E_1, E_2, GN, \mathbf{z}, \mathbf{s}) \\
c_i &= c_i(w_1, w_2, y, E_1, E_2, GN, \mathbf{z}, \mathbf{s}) \\
k_i &= k_i(w_1, w_2, y, E_1, E_2, GN, \mathbf{z}, \mathbf{s}) \\
p_i &= p_i(w_1, w_2, y, E_1, E_2, GN, \mathbf{z}, \mathbf{s}) \\
c_k &= c_k(w_1, w_2, y, E_1, E_2, GN, \mathbf{z}, \mathbf{s}) \\
c_p &= c_p(w_1, w_2, y, E_1, E_2, GN, \mathbf{z}, \mathbf{s})
\end{aligned} \tag{1.5}$$

The system brought to the data thus consists of ten equations where $l_1,\ l_2,\ c_1,\ c_2,\ k_1,$

[13] See Bourguignon, Browning, and Chiappori (2009) for a more detailed discussion.

k_2, c_k, p_1, p_2, and c_p are observable function of w_1, w_2, y, E_1, E_2, GN, $\mathbf{z}$, and $\mathbf{s}$, which are assumed to be completely observed. As proven in Cherchye, De Rock and Vermeulen (2012) Proposition 1 (page 3383) this system of equations allows me to recover the underlying structural model that consists of both spouses individual preferences, household production technologies, and the Pareto weight. For the proof of Cherchye, De Rock and Vermeulen's (2012) Proposition 1 refer to their online appendix page 1.

The model is identified under a specific weak nonsingularity condition. This condition pertains to the partial derivatives of the production functions f_k and f_p with respect to one distribution factor z of $\mathbf{z}$ and one production shifter s of $\mathbf{s}$, where the functions of f_k and f_p are assumed to exhibit constant returns to scale. How is identification achieved through this characterization of a two-stage allocation process and the existence of at least one distribution factor and one production shifter? First, the identification of household production technologies using the assumption of constant returns to scale and cost minimization can be identified in a standard way by means of variation in wages and expenditures on the inputs. To identify the sharing rule ρ_1 and ρ_2 and preferences over leisure and private consumption, there needs to be a production shifter s and distribution factor z. If there were no public goods, the variations in wages, nonlabor income, and gender norms would be enough to use Chiappori's (1992) identification strategy. The problem is that with household goods, the variation in wages, nonlabor income, and gender norms also impacts the stock of household goods chosen. So we need additional variables to keep household goods constant. Thus, the importance of the distribution factor z and production shifter s is that they serve to keep the domestic goods constant while allowing variation in the individual wages, nonlabor income, gender norms, and education levels. Because there are two domestic goods, we need at least one production shifter in addition to a distribution factor to keep the output of the domestic goods constant at $\bar{f}_k$ and $\bar{f}_p$.

1.5 Parametric Specification

I now turn to the parametric specifications for the preferences, production function, and bargaining power. Referring to the two-stage allocation representation mentioned above, we begin with a utility form for the second stage of determining leisure and private consumption, conditional on the quantity of domestic goods produced $\bar{f}_k$ and $\bar{f}_p$ and the sharing rule $\bar{\rho}_i$, which is the residual nonlabor income decided in the first-stage allocation.

Using an individual indirect utility function allows for the derivation of flexible equations for the observables while also allowing leisure and individual consumption to not be separable from the unobserved outputs of the household production process. The second stage's spouses' preferences concerning leisure and private consumption, conditional on $\bar{f}_k$, $\bar{f}_p$, and $\bar{\rho}_i$ can be represented with the following utility function of the PIGLOG class for spouse $i = \{1, 2\}$:

$$\upsilon_i(w_i, \bar{\rho}_i, \bar{f}_k, \bar{f}_p, E_i, GN) = \frac{ln(w_i + \bar{\rho}_i) - ln\ T_i(w_i, E_i, GN; \bar{f}_k, \bar{f}_p)}{(w_i)^\beta} \tag{1.6}$$

where

$$ln\ T_i(w_i, E_i, GN; \bar{f}_k, \bar{f}_p) = \Big(\alpha_i^1(\mathbf{d_i}) + (\alpha_i^2\ GN + \alpha_i^3 + \alpha_i^4 E_i)\ ln\ \bar{f}_k + (\alpha_i^2\ GN + \alpha_i^5)\ ln\ \bar{f}_p\Big) ln\ w_i.$$

Here α_i^1 captures the trade off between leisure and consumption. α_i^2 reflects how much gender norms affect the joint value of the public goods with a positive value, meaning a positive relationship between gender norms and the valuation of public goods in comparison to consumption and leisure. α_i^3 captures the valuation of the public good of the children and α_i^4 captures the change in valuation when the person is highly educated. Similarly, α_i^5, captures the valuation of the public good of the household. I consider age as a taste shifter and use $\alpha_i^1(\mathbf{d^i}) = \alpha_i^{10} + \alpha_i^{11} age_i$ for $i = \{1, 2\}$. As a sensitivity analysis, I compare the estimation results and counterfactual if I consider a distinct indirect utility where α_i^2 affects the preference for the public goods proportionally rather than additively. The main conclusions of the effects of

gender norms remain very similar between the two indirect utility parametric forms and can be seen in the appendix of the paper.

For the household production technologies that transform the expenditures and time spent on home production into the domestic good f_k and f_p, I assume the following technologies with constant elasticity of substitution form and constant returns to scale:

$$f_j(c_j, j_1, j_2; \mathbf{s_j}) = (\gamma_{1j}(j_1)^{g_j(\mathbf{s_j})} + \gamma_{2j}(j_2)^{g_j(\mathbf{s_j})} + \gamma_{3j}(c_j)^{g_j(\mathbf{s_j})})^{\frac{1}{g_j(\mathbf{s_j})}} \quad \text{for } j = \{k, p\} \quad (1.7)$$

where $g_j(\mathbf{s_j})$ is assumed to depend on the production shifters $\mathbf{s_j}$. I consider the number of children (*kids*) and the mean age of the children (*meanagekids*) for the production shifters for the two domestic goods by assuming $g_j(\mathbf{s_j}) = g_{0j} + g_{1j}\ kids + g_{2j}\ meanagekids$ for $j = \{k, p\}$. Within Cherchye et al. (2012), γ_{1j} and γ_{2j} are interpreted as the productivity of each hour by spouse 1 and spouse 2. I diverge from this interpretation. In this simple setup of the collective model, all work hours, market or public, have the same disutility. However, in a context of unequal gender norms, it is unlikely that men are indifferent between an hour working in the market and an hour cleaning the house. To capture the social stigma of men participating in areas that are thought to pertain to the sphere of women without complicating the model with distinct disutilities for types of work, I reinterpret the meaning of γ_{1j} and γ_{2j}. These parameters capture the combination of the productivity of each hour by the spouses plus the social stigma for that gender to partake in the production of the public good j. Meaning γ_{1j} is likely to be very low to explain the lack of participation of men in public good hours for almost all wages and levels of bargaining power.

Recall that γ_{1j} and γ_{2j} are assumed to be fixed even as changes in gender norms occur. I make such an assumption as I only take a relatively small window of time between 2002 and 2011, where men, if anything, decrease the number of average hours dedicated to the public goods. Given the timeline of analysis and data averages, I think it a fair assumption to fix the productivities and social stigma for men.[14] If a longer

[14] As a sensitivity analysis, I reestimate the model allowing for different productivities plus social stigma for couples with low and high egalitarian gender norm couples and find the estimates statistically the same. For simplicity, I continue with a model with fixed productivities and social stigma.

window of time was analyzed where even larger changes to gender norms might have shifted the social stigma of men performing child care and housework, then it would be suitable to allow gender norms to change the public goods production functions.

Finally, I assume that the Pareto weight λ takes the following logistic form:

$$\lambda(w_1, w_2, y, GN, \mathbf{z}) = \frac{exp(\Lambda^0 + \Lambda^1 \frac{w_1}{w_2} + \Lambda^2 y + \Lambda^3 GN + \Lambda^{4\prime}\mathbf{z})}{1 + exp(\Lambda^0 + \Lambda^1 \frac{w_1}{w_2} + \Lambda^2 y + \Lambda^3 GN + \Lambda^{4\prime}\mathbf{z})} \tag{1.8}$$

thus, restricting λ to be between 0 and 1. For the two distribution factors $\mathbf{z}$, I consider the age difference between the husband and the wife and the difference in educational years.

With the parametric forms set up for preferences, production functions, and bargaining power, then the system of equations shown in the system (5) can be solved using the two-stage allocation representation. To account for unobservable heterogeneity across household, I add additive errors to the system equations. The errors are uncorrelated across households. The system of ten equations used for estimation can be found in the Appendix Section A.1. For derivations of the equations, refer to Appendix Section B. Parametric Specification of Cherchye et al. (2012) remains the same in my case. The parameters are estimated using the feasible generalized nonlinear least squares estimator (see Greene 2008). All the exogenous variables needed for estimation are those previously shown in Table 1.1 in addition to the gender norms index. The production shifters are the mean age of children and the number of children, while the distributional factors are the age difference and educational years difference.

1.6 Results

I pool all 782 couple observations from the three distinct survey years and estimate the 31 parameters of the model by means of the feasible generalized nonlinear least squares, and the standard errors are computed using the numerical derivative method. There are 13 parameters for gender preferences, 12 for the public good production functions, and 6 for bargaining power. With the identification strategy presented in

Section 4, the individual preferences are generically identified up to a strict increasing transformation, and the Pareto weight or bargaining power is exactly identified. Table 1.4 shows our estimation results where spouse 1 is the husband and spouse 2 is the wife in the household.[15]

Most parameters turn out to be statistically significant. Both domestic goods have a significant impact on spouses' leisure and consumption decisions, see estimates of $\alpha_1^2, \alpha_1^3, \alpha_1^4, \alpha_1^5, \alpha_2^2, \alpha_2^3, \alpha_2^4$, and α_2^5. Particularly in the case of the domestic good of children k, a higher education level of individuals E_i increases women's and men's preference for this good, as seen in estimates of α_1^4 and α_2^4. As education rises, both women and men invest more in the children, consistent with the literature on education and child investment (Duflo, 2003; Imai et al., 2014). A positive β means that as their individual wages increase, as does the weight on leisure in comparison with consumption, meaning leisure is a luxury good for both genders.

For gender norms, I find that it is statistically significant for both its impact on gender preferences and bargaining power. A positive α_1^2 means that with more egalitarian gender norm, the husband's preference for the public goods increases with respect to consumption and leisure, while women's negative α_2^2 means that the preference for the public goods decreases. For bargaining power, with more egalitarian gender norms, the negative parameter of Λ^3 decreases the power of the husband and thus increases for women. The stories support the changes I propose to the collective model to better fit the Mexican population, gender norms, and their work time evolution.

Recall that to keep the model simple and avoid having to include the disutility costs for distinct types of work, the public good "productivities" also include the social cost or social stigma internalize by the couple when men perform these activities. Thus, for the household production process, it turns out that when internalizing the social stigma, one extra time unit spent on children by the wife benefits f_k much more than one unit of time spent by the husband ($\gamma_{1k} = 0.03$ vs $\gamma_{2k} = 0.30$) and similarly for the household domestic good ($\gamma_{1p} = 0.01$ vs $\gamma_{2p} = 0.21$). Thus, I do not claim that men are significantly less productive than women in the production of public goods

[15]The system of equations forming our structural collective model is highly nonlinear. From the several local minima found, I selected the lowest.

Table 1.4: Structural Estimation Results

	Estimate	SE
Preference parameters		
α_1^1	0.837	0.023
$\alpha_1^{11}[age/10]$	0.016	0.005
$\alpha_1^2[GN]$	0.657	0.167
$\alpha_1^3[f^k]$	-3.720	0.125
$\alpha_1^4[f^k EduH]$	0.392	0.076
$\alpha_1^5[f^p]$	-3.541	0.119
α_2^1	0.801	0.034
$\alpha_2^{11}[age/10]$	0.016	0.006
$\alpha_2^2[GN]$	-0.717	0.208
$\alpha_2^3[f^k]$	-3.630	0.233
$\alpha_2^4[f^k EduH]$	1.465	0.172
$\alpha_2^5[f^p]$	-1.926	0.133
β	0.096	0.010
Child care production		
γ_1^k	0.028	0.018
γ_2^k	0.302	0.108
γ_3^k	0.670	0.327
g_0^k	-3.588	0.810
g_1^k $[kids]$	0.183	0.107
g_2^k $[meanagekids/10]$	1.223	0.344
Housework production		
γ_1^p	0.009	0.004
γ_2^p	0.213	0.050
γ_3^p	0.778	0.118
g_0^p	-0.300	0.104
g_1^p $[kids]$	-0.062	0.021
g_2^p $[meanagekids/10]$	-0.024	0.044
Pareto weight parameters		
Λ^0	-1.450	0.069
Λ^1 $[w_1/w_2]$	1.407	0.027
Λ^2 $[y]$	-0.012	0.003
Λ^3 $[GN]$	-0.255	0.094
Λ^4 $[age_1/10 - age_2/10]$	-0.191	0.030
Λ^5 $[edu_1 - edu_2]$	0.012	0.008

Coefficients estimated using feasible generalized nonlinear least squares estimator. SE computed using the numerical derivative method

but rather that once the spouses internalize the social stigma they follow a path choice similar to one where men are immensely unproductive in comparison to women.[16]

The production shifter of the mean age of children affects the production of the domestic good of children f_k, and the number of children affects the production of the household good f_p. Recalling the identification strategy, one requires at least one significant production shifter for at least one domestic public good. The estimates of the household production parameters are easier to interpret through the elasticity of substitution, which is defined as $1/(1 - g_j(\mathbf{s}^j))$ for the given technology. For the children's good f_k, this elasticity becomes 0.32 for an average household. It increases with the number of children and with the average age of children, although the former is not statistically significant. The elasticity of substitution for f_p amounts to 0.68 for an average household. It decreases with the number of children (statistically significant) and with their average age.

In terms of the Pareto weight parameters, the husband's relative wage and husband's difference in years of education turn out to have a positive (negative) impact on the husband's (wife's) Pareto weight, ceteris paribus. In contrast, the household's non-labor income, more egalitarian gender norms, and age difference decrease (increase) the husband's (wife's) Pareto weight. As before, these results are useful given the earlier identification argument, which showed that at least one significant distribution factor in addition to a significant production shifter is sufficient for identification. I conclude that the setting under study with two statistically significant production shifters and one statistically significant distributional factor fulfills the identification conditions. Furthermore, given the statistical significance of the Pareto weight parameters and the different preferences for the husband and wife, this implies a strong rejection of the unitary model. The fitted average bargaining power for the couples is 0.537 for men and 0.463 for women. The median husband's bargaining power is 0.49, the 25th percentile is 0.34, and the 75th percentile is 0.67.

To understand some of the mechanics of the model and how responsive some choices

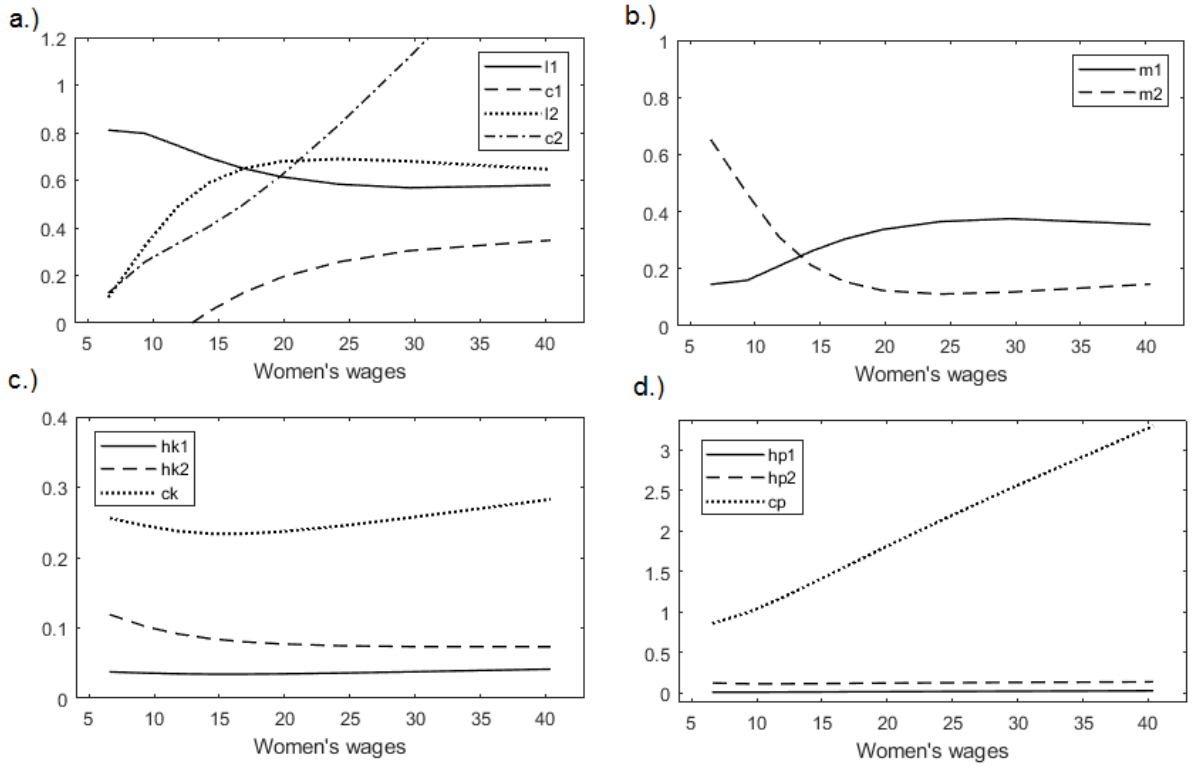
[16]This result comes into contrast with the γs found in Cherchye et al. (2012) where the productivities found are mostly equal with $\gamma_{1k} = 0.36$ vs $\gamma_{2k} = 0.42$ and$\gamma_{1p} = 0.20$ vs $\gamma_{2p} = 0.20$. Thus, in the Netherlands there is likely no large social stigma for men engaged in household public goods.

are to the exogenous variables, I present in Figure 1.5 how the different hours and expenditures respond as women's average wage moves while keeping all other variables (e.g., number and mean age of children, age of spouses, men's wages, and gender norms) at their means. The comparative statics of a change in women's wages on leisure and private consumption is shown in Figure 1.5 panel a. For women with lower wages, leisure is more responsive than her consumption, but around the median wage of 17 pesos per hour, the increase in wages barely impacts her leisure, while her private consumption steadily increases. Similarly, at the median women's wage with all the exogenous variables at the mean gives an almost equal leisure time between husband and wife, yet their consumption is not on par. The largest disparity in total work time is when women have the lowest wages. The changes in leisure from women's wages come primarily from changes in market work rather than from changes in public good hours (panel b). With the low γ_{1k} and γ_{1p} (public productivities plus social stigma) of men there are almost no changes in men's public good hours regardless of women's wages (panels c and d). This exercise emphasizes the importance of wages in determining total work time disparity. Before turning into the importance of gender norms I dedicate the next section to exploring the fitness of the model.

1.6.1 Model Fit

Having now a sense of some of the mechanics of the model, in this section, I present how well the model's predictions match with the couple's choices. Although the model chooses the estimates that reduced the nonlinear least squares for all the time and expenditure equations considered, given the large nonlinearity of the model, it is not necessarily obvious that the model estimates would satisfactorily get close to all of the data means. First, I present in Table 1.5 the sample means and model predictions for weekly hours and monthly expenditures. The average leisure per week is 91 hours for women and 108 for men, while the model predicts 95 and 114 hours. The small overprediction of men's leisure comes from an underprediction of men's market work hours, as the data gives 49 and the model predicts 45. Women's market hours are more closely predicted at 37, only two hours more than the sample's average. The

Figure 1.5: Spouses' Choices as Women's Average Wage Changes



small overprediction of wives' leisure comes from a three-hour underprediction of child care with the model predicting 16 hours. For men's housework, the hours are exactly predicted at 7 and less than an hour off for men's child care. In general, the model does a good job at fitting the means. The total work time difference of couples is only overpredicted by an hour, with the model giving 19.3 hours per week, while the data gives 17.7 hours per week. In terms of expenditures, there is an overprediction of private consumption for both spouses, as shown in Table 1.5.

With the model fitting well the data at the means, I now explore the fit of the model by using an out-of-sample case with single mothers' choices. The model is able to replicate the general trends of single mothers, which can be seen in Appendix A.2 in Table 9. Although this out-of-sample can only speak to the fitness of the female estimates of the model, it gives general confidence in the model's estimates.

As shown earlier in Section 2, there was a reduction in total work time disparity from 2002 to the more recent waves of the MxFLS; can the model predictions also mimic this trend? Figure 1.6 shows in blue the trend for total work time disparity as

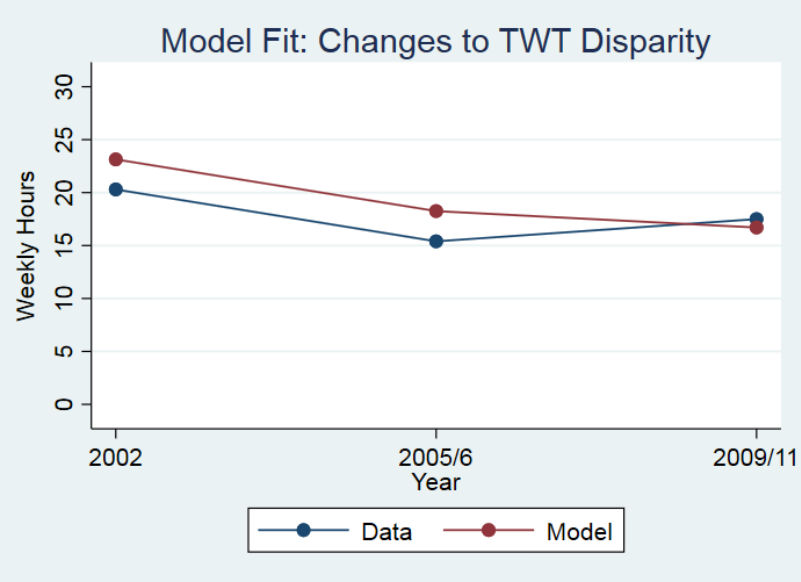
Table 1.5: Data vs. Model Means

	Wife		Husband	
	Data	Model	Data	Model
Leisure	90.7	94.6	108.4	113.9
Market Work	35.1	37.0	49.3	44.8
Child Care	19.9	16.3	7.2	7.0
Housework	22.3	20.2	3.1	2.4
Private Expenditures	0.30	0.63	0.35	0.40
	Household			
	Data	Model		
Child Care Expenditures	0.35	0.31		
House Expenditures	2.10	1.87		
Total Work Time Difference	17.7	19.3		

Note: Expenditures in the thousands of Mexican pesos

seen in the MxFLS sample and in red the model's predictions. Similarly to model's fit of the means, there is a small overprediction of the total work time disparity by the model; however, the model reproduces the drop in total work time disparity quite well particularly for the change between 2002 and 2005/6. The data shows a drop of 4.9 hours, which is perfectly predicted by the model. That none of the moments used for estimation were separated by year, yet the model mimics the drop in total work time disparity gives confidence in the model's estimates. The decision to extend the collective to include gender norms and education was not only to address the research question but also because I believe it to better adapt the model to a developing country context. To test this, I estimate the original Cherchye et al. (2012) (CDV) model without my extensions. The CDV model overpredicts the total work time disparity in all years and the shift between 2002 and 2005/6 from 4.9 hours to 6.4 hours. For more information on the parameters and mean predictions for the original CDV model, refer to Appendix A.4.

Figure 1.6: Model Fit for Total Work Time Disparity



1.6.2 Gender Norms as a Determinant of Total Work Time Disparity

The magnitudes of the estimated parameters concerning gender norms in the collective model are not easy to grasp by themselves. To better understand how these parameters affect total work time, I generate some counterfactual analyses. Given the strong match between data and model in the change in total work time disparity between 2002 and 2005/6, I conduct a counterfactual analysis using the change between these two waves of the survey.

In 2002, the gender norms in the sample had an average of 0.477, while in 2005/6, they increased to 0.638. Between these two survey waves, there was also a decrease in the total work time disparity by 4.9 hours in both the sample and the model's predictions. How much did the changes in gender norms impact the changes in total work time disparity? To answer this question, I analyze a counterfactual setting where the gender norms index would not have changed from 2002 to 2005/6. Thus, the 2002 gender norms index would have remained the same in 2005/6. Figure 1.7 normalizes the total work time disparity in 2002 to zero. It shows the model predicts exactly the drop of 4.9 hours with the 2005/6 sample. If the gender norm indices of the couple's peers would have remained the same as the 2002 indices, then the total work time

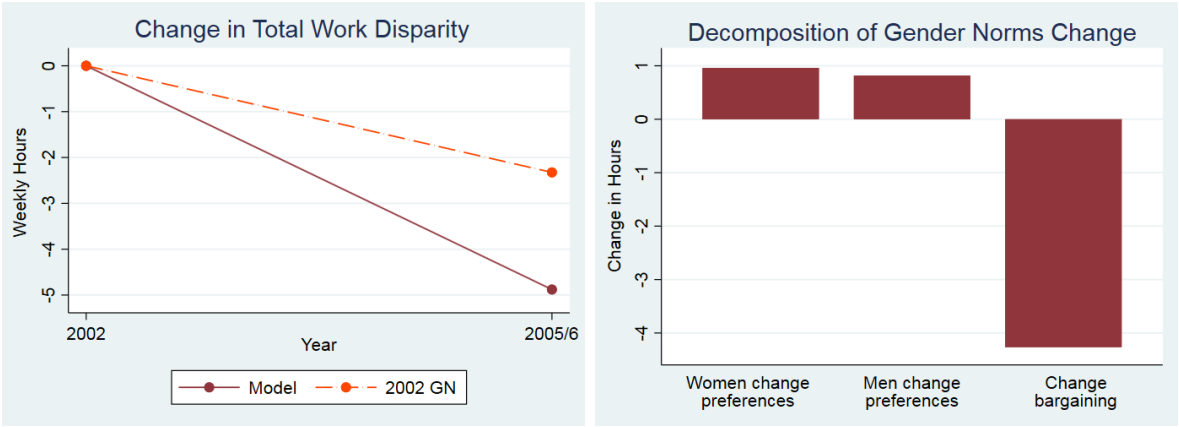


Figure 1.7: Counterfactual Changes in Gender Norms from 2002 to 2005/5

disparity would have only decreased by 2.3 hours instead of its observed 4.9 hours. The drop of 2.3 hours is attributed to other changes besides gender norms between the 2002 and 2005 samples, such as changes in wages and the number of children. This finding means the gender norm index changes from an average of 0.477 to 0.638, an increase of 16 percentage points or 33.8 percent, accounted for a fall of 2.6 hours or 14% in the total work time disparity in 2005.

The drop of 2.6 hours from changes in the gender norms index comes from three mechanisms. Recall that gender norms impact the preferences of men and women and their bargaining power. More egalitarian gender norms increase the preference for public goods with respect to private consumption and leisure for husbands but decrease the preference for wives. The third mechanism is the increase in women's bargaining power as gender norms become more egalitarian. I provide the decomposition effect of the three mechanisms in the second panel of Figure 1.7. Both the changes in preferences of men and women increase the work time disparity. When women care more about their leisure and consumption in comparison to the public goods, they reduce the hours spent on the two public goods, but so do the husbands because of the complementarity of their hours (see Table 1.6). With women's increasing emphasis on leisure and consumption, they increase some hours in market work to consume more and gain some extra leisure hours. Even if men do not change their preferences, they still reduce their market work even more than women, providing them an increase in leisure in response to women's change in public hours. The resulting increase in

Table 1.6: Sample of 2005/6 with Changes in Gender Norms Decomposition

	GN 2002	GN 2005	Only Pref Women 2005	Only Pref Men 2005	Only Bargaining Power 2005
Total Work Time Disparity	20.81	18.26	21.77	21.63	16.55
Leisure_m	112.66	111.67	114.65	112.22	110.12
Leisure_w	91.85	93.41	92.88	90.59	93.57
Market_m	46.09	47.06	44.5	46.12	48.6
Market_w	41.28	39.42	41.67	40.92	39.42
child care_m	6.65	6.74	6.39	6.99	6.65
child care_w	14.99	15.41	14.51	15.89	14.99
Housework_m	2.61	2.53	2.46	2.66	2.63
Housework_w	19.89	19.76	18.94	20.59	20.02
Expenditures House	1.95	1.91	1.85	2.00	1.97
Expenditures Child	0.30	0.31	0.29	0.32	0.30
Private_m	0.44	0.37	0.42	0.39	0.45
Private_w	0.65	0.71	0.68	0.68	0.65
Bargaining Power	0.53	0.52	0.53	0.53	0.52

men's leisure is greater than that of women. When men increase their preference for the public goods, both spouses increase their public good hours; women decrease their market work but by not enough to offset the public good hour increase, and, thus, women's total leisure decreases.

The decrease in total work time disparity by gender norms is driven completely by the third mechanism of increasing women's bargaining power. If the sample of couples in 2005/6 had kept the 2002 gender norms index, the fitted bargaining power would have averaged 52.7% power for husbands and 47.3% for wives. The change to the 2005 gender norms index lowers the fitted bargaining power of husbands to 51.9%. This means that women's bargaining power increased by 0.8 pp or 1.7%, reducing the total work time disparity by more than 4 hours. The 4-hour change is driven by men increasing their market work and women reducing theirs.

Therefore, changes in gender norms can reduce total work time disparity through changes in bargaining power. To obtain a better grasp of the changes in hourly choices when bargaining power shifts, be it through gender norms or other venues, I provide a graphical illustration in Figure 1.8. This graphic represents an exogenous shock to bargaining power while keeping all other exogenous variables fixed, such as wages and

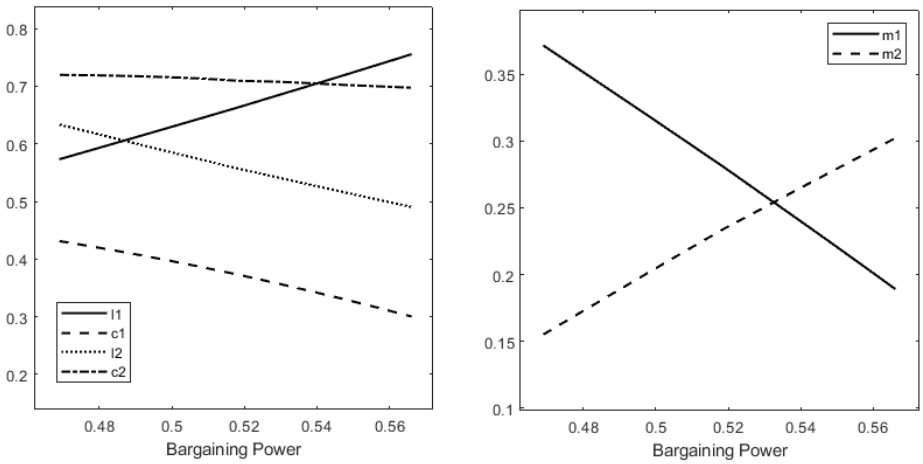
gender norms indices. In the middle of the graph, there is the baseline average predicted bargaining power in 2005 at 0.519. I allow changes in the bargaining power of men to shift from approximately 0.47 to 0.57 to see the impact on the dependent variables of leisure, private consumption, and market work. As women's power increases (and men's decreases), both her leisure l_2 and private consumption c_2 increase as well, with the rate of change higher for leisure than consumption. Men's leisure decreases as men's bargaining power decreases, but their consumption remains almost the same. The changes in leisure are mainly driven by the changes in market work seen in the second panel of Figure 1.8. In the model, most shifts to bargaining power come from the change in wages and gender norms, which also affect other parts of the model: thus, we would rarely see such linear dynamics.

Figure 1.8 also shows that women's and men's leisure would be equal for the 2005 sample at approximately 0.49 bargaining power for men and 0.51 for women. Why is parity not achieved with an average bargaining power of 0.5 for both? The necessary bargaining power for couples to reach parity in leisure or work time depends on other variables and is particularly responsive to wages. The reason wages matter, beyond that of affecting bargaining power, is that women and men hold different preferences over the trade-off between leisure and consumption, α_i^1, and how leisure is valued as wages increase, β. For example, in 2002, where women's wages were lower than in 2005, the bargaining power of men needed to achieve work time parity was 0.52. In 2009/11, where the recession lowered both men's and women's wages, but not to 2002 levels, the bargaining power of men needed for parity was 0.51. Thus, the division of bargaining power needed for equal total work time is constantly changing, particularly as the average wages of women and men shift.

1.6.3 The Impact of Gender Norms in Comparison to Relative Wages

To capture the importance of the magnitude of gender norms as a determinant of total work time disparity, I compare it to another gender disparity, the gender relative wages. A study realized by Kaplan and Piras (2019) finds that Mexico is the country

Figure 1.8: Total Work Time as Bargaining Power Shifts



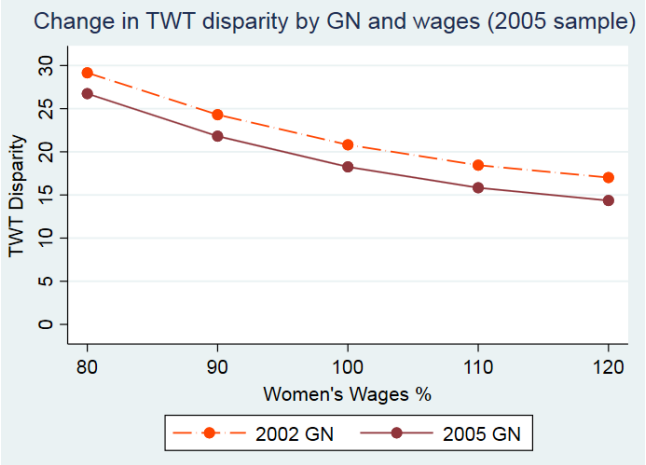
with the largest wage gap of the 14 Latin American countries studied. They find that in 2015, on average, women who work received a salary 28% lower than that of men when considering control variables such as education. Given the intrahousehold comparison, I focus on relative wages between spouses rather than the global wage gap. The average relative wage between spouses in the 2002 sample was 1.56; thus, husbands had on average 56% higher wages than their wives.[17] In 2005, the relative wage changed by 22 percentage points or 13 percent to 1.36.

I compare the effects of gender norms and relative wages by calculating what would be the needed change in gender-relative wages in 2005 to achieve the same 2.6-hour drop in work time disparity caused by gender norms. An increase in women's wages of approximately 11% would be needed, bringing the relative wage to 1.23 to decrease the total work time disparity by 2.6 hours. Similarly, the same drop of hours can be achieved by a decrease of 2005 men's wage by 9% to a relative wage of 1.25. Of course, other combinations of women's wage increases and men's wage decreases could achieve a 2.6-hour drop within the relative wage range of 1.23 and 1.25.

The changes to work time caused by relative wages and gender norms depend on the rest of the exogenous variables. Particularly, the impact of a change in wages or relative wages depends on the starting point of wages. For example, looking at Figure 1.9 at the center, we have women's wages in 2005 (100%) and to the left and right

[17] The average difference between all female and male wages is not as stark as those between couples, with men having wages 16% larger on average in 2002.

Figure 1.9: Changes in GN and Women's Wages for 2005 sample



changes to women's wages by ten percent increments. Following the red line using the gender norms of 2005, we see that increasing their wages from the 90 percent of the initial 2005 wage to 100 percent drops the total work time disparity by 3.6 hours, while a change from 100 percent to 110 percent would drop it by only 2.4 hours. Thus, reducing the wage gap has large effects when going from larger relative wage gaps; in contrast, the impact of gender norms remains quite sizable regardless of the state of women's wages and the relative wage. Ranging women's 2005 wages from 80 percent up to 120 percent, slightly changes the gender norms impact on work time disparity, with the range of impact between 2.4 to 2.7 and the larger impacts coming as women's wages are higher.

In these counterfactual exercises, gender norms and wages are exogenous in the model and, thus, changing one of them does not affect the other; however, it is important to note that if we consider a longer-term horizon, this might not hold. Most likely, in the long-run, egalitarian gender norms and women's wages have a positive feedback or virtuous loop. For example, as gender norms become more egalitarian more women may decide to enter the labor force, further changing gender norms and shrinking the gender wage gap, which then, in turn, may lead to more women joining the labor force, and so on.

1.6.4 Disparity in Public Good "Productivities" and Gender Leisure Preferences

In this subsection, I explore two possible gender differences that could affect total work time disparity: social stigma/public good 'productivities' and gender leisure preferences.

One of the key features in the model, which allows replicating the public good hours for the spouses, is the unequal public good productivities discounted by the 'social' costs or stigma, γ_1 and γ_2. Recall that this strategy helps keep the model's simplifying assumption of having equal disutility for each hour worked, be it market work, household work, or child care. How would the choice of hours be different if spouses had the same public good "productivities" for child care and domestic housework while keeping the productivity of the consumption good constant? Table 1.7 compares the 2005 baseline with unequal γs to that of equal γs (column (1) vs. (2)). It shows that having equal public good productivities with social stigma would result in men and women splitting the hours on the public goods almost equally; however, the total work hour disparity would barely change. Thus, γ_1 and γ_2 help the model replicate the pattern where the wife dedicates more hours to the public goods than the husband. Thus, γ_1 and γ_2 do not determine the amount of leisure or total work time, but rather only how working hours are split conditional on the work hours chosen. If such is the case, the impact on the gender norm index should remain similar at distinct levels of γ_1 and γ_2. In column (3) and column (4) of Table 1.7, I compare the impact of an increase in the gender norms index of 10 percent in the case of spouses having the unequal γs in column (1) and equal γs in column (2). In both cases of unequal or equal γs, the total work time disparity between spouses would be reduced, with only a slightly larger absolute and percentile change when the γs are unequal.

Another potential gender difference that could lead to gender disparities in work time are differences in preferences between the genders. Gender preferences are shown in the preference parameters in Table 1.4, but their differences are better interpreted through a counterfactual analysis. Particularly, if gender preferences are too distinct, this could lead to differences in leisure and work choices regardless of spouses' bar-

Table 1.7: Hour and Expenditure Comparison with Changes in GN and Equal Public Good Productivities

	Baseline 2005	Baseline 2005 and equal prod	GN *1.1	GN*1.10 and equal prod
TWT disparity	18.26	17.90	17.27	17.02
Leisure_m	111.67	112.99	111.2	112.58
Leisure_w	93.41	95.09	93.93	95.56
Market_m	47.06	30.44	47.49	30.90
Market_w	39.42	50.36	38.66	49.75
child care_m	6.74	11.12	6.81	11.21
child care_w	15.41	11.1	15.63	11.24
Housework_m	2.53	13.45	2.50	13.31
Housework_w	19.76	11.45	19.78	11.45
Expenditures house	1.91	1.83	1.90	1.82
Expenditures child	0.31	0.28	0.31	0.28
Consumption_m	0.37	0.22	0.34	0.18
Consumption_w	0.71	0.58	0.73	0.61

gaining power or wages. To explore this possibility, I explore this possibility using a counterfactual where both spouses have equal wages, ages, education, and bargaining power. I allow for spouses to have the same wage and age based on the mean between the husband and wife's characteristics, as well as have the wife's education being the same as the husband's education. Table 1.8 shows the choice in hours and expenditures under equal wages, education, age, and bargaining power. Column (2) also includes equal public good productivities and column (3) leaves the estimated unequal public good productivities. Similar to Table 1.7, having equal public good productivities lowers the total work time disparity by less than an hour. According to column (2), the differences in gender preferences, on average, account for a slightly less than 2 hours in the total work time disparity. Women slightly prefer private consumption to leisure in comparison to men. This finding means that gender differences in preferences on leisure, consumption and the two public goods have a limited impact on total work time disparities.

Table 1.8: Countefactual: Differences in preferences under equal characteristics

	Baseline 2005	=BP, =Charact & = public good productivities	BP=0.5 & = Characteristics
	(1)	(2)	(3)
TWT disparity	18.26	1.98	2.82
Leisure$_m$	111.67	109.9	108.58
Leisure$_w$	93.41	107.92	105.76
Market$_m$	47.06	37.35	51.23
Market$_w$	39.42	39.32	30.17
child care$_m$	6.74	10.02	6.13
child care$_w$	15.41	10.02	13.61
Housework$_m$	2.53	10.74	2.06
Housework$_w$	19.76	10.74	18.45
Expenditures house	1.91	1.79	1.88
Expenditures child	0.31	0.26	0.29
Consumption$_m$	0.37	0.26	0.38
Consumption$_w$	0.71	0.49	0.69
Pareto	0.52	0.50	0.50

1.7 Conclusion

In this paper, I proposed a collective labor supply model with household production that includes exogenously determined gender norms. Individuals care about their own leisure and private consumption of market goods but also about two public goods: children and the home. I estimated the model using Mexican couples participating in the labor force with at least one child. The empirical results of the collective model for labor force participating Mexican couples with children reveal some interesting patterns. First, the findings show that an evolution towards gender egalitarian norms closes the gender disparity in total work time. The mechanism through which more egalitarian gender norms have an impact is through increasing men's preference and decreasing women's preferences for the two public goods relative to leisure and private consumption. Simultaneously, egalitarian gender norms increase women's intrahousehold bargaining power. While the changes in preferences increase the total work time disparity, the increases in women's bargaining power drive the final impact of reducing total work time. The significance and magnitude of the parameters for gender norms support the idea that collective models for countries where gender norms are perhaps

unequal might be better adapted to the population if they are included.

From 2002 to 2005, the increase in gender norms of 16 percentage points accounts for a reduction of approximately three hours in the total work time disparity. The impact of gender norms is sizable in comparison to other gender disparities. For example, the equivalent reduction of three hours in total work time in 2005 would require women's wages to increase by 11 percent. Furthermore, the impacts of increases in women's wages decrease as the wage gap diminishes, while the effects of gender norms on work time disparity remain sizable and continue to do so regardless of relative wages.

Within the collective model presented, preferences and bargaining power were allowed to change with gender norms, but the social stigma experienced by husbands performing unpaid work was assumed to be fixed. An extension for future research would be to consider how social sigma costs may change within the public good production functions as gender norms change. However, since social stigma costs may be slower to change, either a longer time horizon or a policy program trying to change gender norms may be needed to disentangle these changes. Another extension of the model that I leave for future work is implementing distinct disutility for different types of works. One way to allow for distinct disutility for market work in comparison to public good hours could be done by allowing the bargaining power to depend on the labor income achieved, not just on the hourly wage. This change would make the bargaining power endogenous to the model and a step harder to solve but would allow highlighting that some spouses might also increase market work not only for income but also for gains in bargaining power within the household.

Bibliography

Afridi, F., Bishnu, M., and Mahajan, K. (2019). What determines women's labor supply? the role of home productivity and social norms.

Burda, M., Hamermesh, D., and Weil, P. (2013). Total work and gender: facts and possible explanations. *Journal of Population Economics*, 26(1):239–261.

Campaña, J. C., Giménez-Nadal, J. I., and Molina, J. A. (2018). Gender norms and the gendered distribution of total work in latin american households. *Feminist Economics*, 24(1):35–62.

Cavapozzi, D., Francesconi, M., and Nicoletti, C. (2021). The impact of gender role norms on mothers' labor supply. *Journal of Economic Behavior Organization*, 186:113–134.

Cherchye, L., Rock, B. D., and Vermeulen, F. (2012). Married with Children: A Collective Labor Supply Model with Detailed Time Use and Intrahousehold Expenditure Information. *American Economic Review*, 102(7):3377–3405.

Chiappori, P.-A. (1992). Collective Labor Supply and Welfare. *Journal of Political Economy*, 100(3):437–467.

Cislaghi, B. and Heise, L. (2020). Gender norms and social norms: differences, similarities and why they matter in prevention science. *Sociology of health & illness*, 42(2):407–422.

Donni, O. and Matteazzi, E. (2012). On the importance of household production in collective models: evidence from us data. *Annals of Economics and Statistics/ANNALES D'ÉCONOMIE ET DE STATISTIQUE*, pages 99–125.

Donni, O. and Molina, J. A. (2018). Household collective models: Three decades of theoretical contributions and empirical evidence.

Duflo, E. (2003). Grandmothers and granddaughters: old-age pensions and intrahousehold allocation in south africa. *The World Bank Economic Review*, 17(1):1–25.

Fernández, R. and Fogli, A. (2009). Culture: An empirical investigation of beliefs, work, and fertility. *American economic journal: Macroeconomics*, 1(1):146–77.

Field, E., Pande, R., Rigol, N., Schaner, S., and Moore, C. (2019). On her own account: How strengthening women's financial control impacts labor supply and gender norms. *Yale University Economic Growth Center Discussion Paper*, (1068).

Fortin, N. M. (2005). Gender role attitudes and the labour-market outcomes of women across oecd countries. *oxford review of Economic Policy*, 21(3):416–438.

Gurría, A. (2019). La Igualdad de Género y el Empoderamiento de las Mujeres para el Crecimiento Incluyente en México. https://www.oecd.org. Accessed: 2021–09-17.

Guryan, J., Hurst, E., and Kearney, M. (2008). Parental education and parental time with children. *Journal of Economic perspectives*, 22(3):23–46.

Hwang, J., Lee, C., and Lee, E. (2019). Gender norms and housework time allocation among dual-earner couples. *Labour Economics*, 57:102–116.

Imai, K. S., Annim, S. K., Kulkarni, V. S., and Gaiha, R. (2014). Women's empowerment and prevalence of stunted and underweight children in rural india. *World Development*, 62:88–105.

International Labour Organization (2021). Labor force participation rate, female (% of female population ages 15+) (modeled ilo estimate) - mexico. Technical report.

Kaplan, D. and Piras, C. (2019). Brechas de género en el mercado laboral mexicano: comparaciones internacionales y recomendaciones de política pública. Revista de Economía Mexicana, Anuario unam 4, Banco Interamericano de Desarrollo.

Lise, J. and Yamada, K. (2019). Household sharing and commitment: Evidence from panel data on individual expenditures and time use. *The Review of Economic Studies*, 86(5):2184–2219.

Oreffice, S. (2014). Culture and household decision making. balance of power and labor supply choices of us-born and foreign-born couples. *Journal of Labor Research*, 35(2):162–184.

Rapoport, B., Sofer, C., and Solaz, A. (2011). Household production in a collective model: some new results. *Journal of Population Economics*, 24(1):23–45.

Van Klaveren, C., Van Praag, B., and van den Brink, H. M. (2008). A public good version of the collective household model: an empirical approach with an application to british household data. *Review of Economics of the Household*, 6(2):169–191.

West, C. and Zimmerman, D. H. (1987). Doing gender. *Gender & society*, 1(2):125–151.

1A Appendix

1A.1 Model's System of Equations

With the estimation of the system of equations, it allows me to recover the underlying structural model. This comes from the fact that the optimization program shown in equation (4) can be represent as coming from a two-stage allocation problem. In the first stage, the couple agrees on the level of domestic goods f_k and f_p and the conditional sharing rule or the residual non labor income ρ_1 and ρ_2. In the second stage, the individuals maximize their own utility by choosing leisure and private consumption condition on the fixed level of domestic goods $\bar{f_k}$ and $\bar{f_p}$ and the budget constraint defined in the first stage.

The ten distinct equations as functions of observable $w_1, w_2, y, E_1, E_2, GN, \mathbf{z}$, and $\mathbf{s}$ are:

$$l_i = \left[A_i + \beta^i \left(\ln \left(\frac{w_1 + w_2 + y}{X(w_1, w_2, E_1, E_2, GN, \lambda)} \frac{\lambda_i}{(w_i)^{\beta^i}} \right) - A_i \ln w_i \right) \right] \times \frac{\left(\frac{w_1 + w_2 + y}{X(w_1, w_2, E_1, E_2, GN, \lambda)} \frac{\lambda_i}{(w_i)^{\beta}} \right)}{w_i}$$

$$c_i = \left[(1 - A_i) - \beta^i \left(\ln \left(\frac{w_1 + w_2 + y}{X(w_1, w_2, E_1, E_2, GN, \lambda)} \frac{\lambda_i}{(w_i)^{\beta^i}} \right) - A_i \ln w_i \right) \right] \times \frac{\left(\frac{w_1 + w_2 + y}{X(w_1, w_2, E_1, E_2, GN, \lambda)} \frac{\lambda_i}{(w_i)^{\beta}} \right)}{w_i}$$

and

$$
\begin{aligned}
k_1 =&\Big(\frac{w_1}{\gamma_k^1}\Big)^{\frac{1}{g^k(\mathbf{s_k})-1}} \frac{w_1+w_2+y}{X(w_1,w_2,E_1,E_2,GN,\lambda)}(g^k(w_1,w_2))^{\frac{-g^k(\mathbf{s_k})}{g^k(\mathbf{s_k})-1}} \\
&\times\Bigg[-\frac{\lambda}{(w_1)^\beta}\hat{\alpha}_1^k\ \ln w_1-\frac{(1-\lambda)}{(w_2)^\beta}\hat{\alpha}_2^k\ \ln w_2\Bigg] \\
k_2 =&\Big(\frac{w_2}{\gamma_k^2}\Big)^{\frac{1}{g^k(\mathbf{s_k})-1}} \frac{w_1+w_2+y}{X(w_1,w_2,E_1,E_2,GN,\lambda)}(g^k(w_1,w_2))^{\frac{-g^k(\mathbf{s_k})}{g^k(\mathbf{s_k})-1}} \\
&\times\Bigg[-\frac{\lambda}{(w_1)^\beta}\hat{\alpha}_1^k\ \ln w_1-\frac{(1-\lambda)}{(w_2)^\beta}\hat{\alpha}_2^k\ \ln w_2\Bigg] \\
c_k =&(\gamma_k^3)^{\frac{-1}{g^k(\mathbf{s_k})-1}} \frac{w_1+w_2+y}{X(w_1,w_2,E_1,E_2,GN,\lambda)}(g^k(w_1,w_2))^{\frac{-g^k(\mathbf{s_k})}{g^k(\mathbf{s_k})-1}} \\
&\times\Bigg[-\frac{\lambda}{(w_1)^\beta}\hat{\alpha}_1^k\ \ln w_1-\frac{(1-\lambda)}{(w_2)^\beta}\hat{\alpha}_2^k\ \ln w_2\Bigg] \\
p_1 =&\Big(\frac{w_1}{\gamma_p^1}\Big)^{\frac{1}{g^p(\mathbf{s_p})-1}} \frac{w_1+w_2+y}{X(w_1,w_2,E_1,E_2,GN,\lambda)}(g^p(w_1,w_2))^{\frac{-g^p(\mathbf{s_p})}{g^p(\mathbf{s_p})-1}} \\
&\times\Bigg[-\frac{\lambda}{(w_1)^\beta}\hat{\alpha}_1^p\ \ln w_1-\frac{(1-\lambda)}{(w_2)^\beta}\hat{\alpha}_2^p\ \ln w_2\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
p_2 =&\Big(\frac{w_2}{\gamma_p^2}\Big)^{\frac{1}{g^p(\mathbf{s_p})-1}} \frac{w_1+w_2+y}{X(w_1,w_2,E_1,E_2,GN,\lambda)}(g^p(w_1,w_2))^{\frac{-g^p(\mathbf{s_p})}{g^p(\mathbf{s_p})-1}} \\
&\times\Bigg[-\frac{\lambda}{(w_1)^\beta}\hat{\alpha}_1^p\ \ln w_1-\frac{(1-\lambda)}{(w_2)^\beta}\hat{\alpha}_2^p\ \ln w_2\Bigg] \\
c_p =&(\gamma_p^3)^{\frac{-1}{g^p(\mathbf{s_p})-1}} \frac{w_1+w_2+y}{X(w_1,w_2,E_1,E_2,GN,\lambda)}(g^p(w_1,w_2))^{\frac{-g^p(\mathbf{s_p})}{g^p(\mathbf{s_p})-1}} \\
&\times\Bigg[-\frac{\lambda}{(w_1)^\beta}\hat{\alpha}_1^p\ \ln w_1-\frac{(1-\lambda)}{(w_2)^\beta}\hat{\alpha}_2^p\ \ln w_2\Bigg]
\end{aligned}
$$

where

$$A_i \equiv \alpha_i^1(\mathbf{d_i}) + \hat{\alpha}_i^k \ln\left(\frac{w_1+w_2+y}{X(w_1,w_2,E_1,E_2,GN,\lambda)}\frac{1}{g^k(w_1,w_2)}\left[-\frac{\lambda}{(w_1)^\beta}\hat{\alpha}_1^k \ln w_1 - \frac{(1-\lambda)}{(w_2)^\beta}\hat{\alpha}_2^k \ln w_2\right]\right)$$
$$+ \hat{\alpha}_i^p \ln\left(\frac{w_1+w_2+y}{X(w_1,w_2,E_1,E_2,GN,\lambda)}\frac{1}{g^k(w_1,w_2)}\left[-\frac{\lambda}{(w_1)^\beta}\hat{\alpha}_1^p \ln w_1 - \frac{(1-\lambda)}{(w_2)^\beta}\hat{\alpha}_2^p \ln w_2\right]\right)$$

$$\hat{\alpha}_i^k \equiv (\alpha_i^2 GN + \alpha_i^3 + \alpha_i^4 E_i)$$
$$\hat{\alpha}_i^p \equiv (\alpha_i^2 GN + \alpha_i^5)$$

$$X(w_1,w_2,E_1,E_2,GN,\lambda) = \left[\frac{\lambda}{(w_1)^\beta}[1 - \left(\hat{\alpha}_1^k + \hat{\alpha}_1^p\right) \ln w_1] + \frac{(1-\lambda)}{(w_2)^\beta}[1 - \left(\hat{\alpha}_2^k + \hat{\alpha}_2^p\right) \ln w_2]\right]$$
$$\lambda = \lambda(w_1,w_2,y,GN,\mathbf{z}) = \frac{exp(\Lambda^0 + \Lambda^1\frac{w_1}{w_2} + \Lambda^2 y + \Lambda^3 GN + \Lambda^{4\prime}\mathbf{z})}{1 + exp(\Lambda^0 + \Lambda^1\frac{w_1}{w_2} + \Lambda^2 y + \Lambda^3 GN + \Lambda^{4\prime}\mathbf{z})}$$

1A.2 Model Fit using Single Mothers in the Labor Force

Another way to test the model estimates is to analyse out of sample predictions. In this case, the model can be easily used to predict the hours and expenditures of single mothers.[18] In this case, I adapt the collective model by assuming that a single mother holds all the bargaining power in the household and all inputs by the husband, which in this case is not-existing, are fixed to zero. Table 9 presents the data and model means. Leisure of single mothers is higher than married mothers by about 15 hours with a mean of 95 hours and the model predicts 110 hours. The number of hours in market work is 40 in the data and 33 for the prediction. Child care is under-predicted by 2 hour at 11.3 hours per week. Expenditures for the wife and children are similarly under-prediction given the extra hours of leisure. But the relation between private expenditure and child care remains the same. House expenditures are exactly predicted. Overall for a sample not used to estimate the model and with particularly

[18]The same exercise for single fathers is not feasible given the low number of observations

Table 9: LFP Single Mothers: Data vs. Model Means

	Single Mother	
	Data	Model
Leisure	94.5	109.7
Market Work	39.9	33.0
Child Care	14.2	11.3
Housework	19.5	14.0
Private Expenditures	0.23	0.15
Child Care Expenditures	0.23	0.18
House Expenditures	0.99	0.99

different challenges, the estimates for women from the collective model are able to predict the expenditures means quite well, and are comparable for the sample time hours.

1A.3 Sensitivity Analysis: Distinct Parametric Specification

In this appendix section, I compare the estimation results for the 'main model' discussed in the main text to the ones with a indirect utility specification where the estimates for gender norms impacts the public goods proportionally. This comparison should provide insight into the extent to which using an additively impact of gender norms on the preference of public goods can affect the estimation results and corresponding counterfactual analyses for the couples.

The only change to the model is the way α_i^2 enters in the indirect utility function, I consider this alternative form:

$$v_i(w_i, \bar{\rho}_i, \bar{f}_k, \bar{f}_p, E_i, GN) = \frac{ln(w_i + \bar{\rho}_i) - ln\ T_i(w_i, E_i, GN; \bar{f}_k, \bar{f}_p)}{(w_i)^\beta}$$

where

$$ln\ T_i(w_i, E_i, GN; \bar{f}_k, \bar{f}_p) = \Big(\alpha_i^1(\mathbf{d_i}) + (1 + GN)^{\alpha_i^2}\big((\alpha_i^3 + \alpha_i^4 E_i)\ ln\ \bar{f}_k + \alpha_i^5\ ln\ \bar{f}_p\big)\Big) ln\ w_i.$$

Table 10: Estimation Results with a Different Indirect Utility Specification

	Estimate	SE		Estimate	SE
Preference parameters			*Housework production*		
α_1^1	0.840	0.023	γ_1^p	0.017	0.005
$\alpha_1^{11}[age/10]$	0.014	0.005	γ_2^p	0.297	0.050
$\alpha_1^2[GN]$	1.281	0.266	γ_3^p	0.686	0.087
$\alpha_1^3[f^k]$	-3.934	0.146	g_0^p	-0.193	0.086
$\alpha_1^4[f^k EduH]$	0.376	0.075	$g_1^p\ [kids]$	-0.022	0.016
$\alpha_1^5[f^p]$	-3.807	0.142	$g_2^p\ [meanagekids/10]$	-0.056	0.040
α_2^1	0.776	0.038			
$\alpha_2^{11}[age/10]$	0.018	0.006	*Pareto weight parameters*		
$\alpha_2^2[GN]$	-1.072	0.311	Λ^0	-1.451	0.068
$\alpha_2^3[f^k]$	-3.543	0.235	$\Lambda^1\ [w_1/w_2]$	1.413	0.027
$\alpha_2^4[f^k EduH]$	1.471	0.165	$\Lambda^2\ [y]$	-0.010	0.003
$\alpha_2^5[f^p]$	-1.854	0.151	$\Lambda^3\ [GN]$	-0.245	0.094
β	0.097	0.010	$\Lambda^4\ [age_1/10 - age_2/10]$	-0.191	0.030
Child care production			$\Lambda^5\ [edu_1 - edu_2]$	0.015	0.008
γ_1^k	0.026	0.017			
γ_2^k	0.302	0.102			
γ_3^k	0.672	0.309			
g_0^k	-3.706	0.788			
$g_1^k\ [kids]$	0.213	0.091			
$g_2^k\ [meanagekids/10]$	1.272	0.334			

Coefficients estimated using feasible generalized nonlinear least squares estimator. SE computed using the numerical derivative method

Following the same identification strategy stemming from the two-stage allocation process, the system of equations presented in Appendix A.1 has only one small change, with only the following changes to $\hat{\alpha}_i^k$ and $\hat{\alpha}_i^p$:

$$\hat{\alpha}_i^k \equiv (1 + GN)^{\alpha_i^2}\ (\alpha_i^3 + \alpha_i^4 E_i)$$
$$\hat{\alpha}_i^p \equiv (1 + GN)^{\alpha_i^2}\ \alpha_i^5$$

I get the new estimates found in Table 10. The estimates are quite similar, with the exception of the new preferences parameters based on the change of how gender norms enters into the indirect utility.

The model mean are presented in Table 11 with this different model parametric form having a slight large total work time disparity predicted, coming mainly from a further under-prediction of market work of men in comparison to the 'main' model. Once again the model is able to predict well the drop in total work time disparity between 2002 to 2005/6 with the model predicting 4.1 hours, while the 'main' model

Table 11: Data vs. Model Means (Distinct Indirect Utility)

	Wife		Husband	
	Data	Model	Data	Model
Leisure	90.7	94.4	108.4	113.9
Market Work	35.1	37.0	49.3	44.8
Child Care	19.9	16.4	7.2	7.0
Housework	22.3	20.1	3.1	2.3
Private Expenditures	0.30	0.62	0.35	0.40
	Household			
	Data	Model		
Child Care Expenditures	0.35	0.31		
House Expenditures	2.10	1.85		
Total Work Time Difference	17.7	19.4		

Note: Expenditures in the thousands of Mexican pesos

predicted 4.9. Here, the gender norms change from 2002 to 2005/6 caused a 2.3 hour reduction in the gender work disparity similarly to the 2.6 found in the 'main' model.

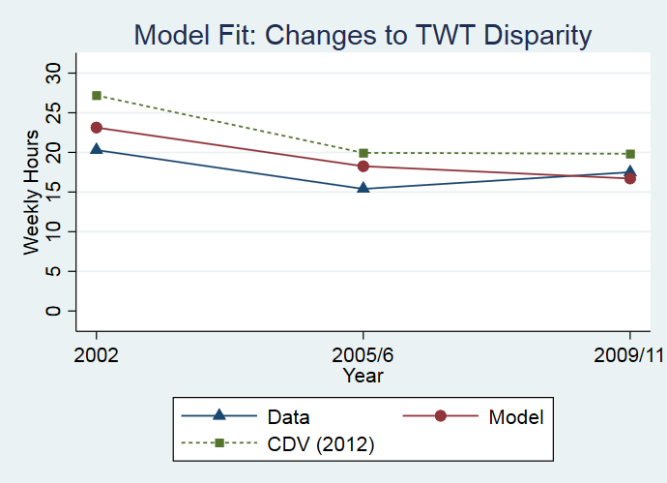
1A.4 Estimating Cherchye, De Rock and Vermeulen (2012) Model for Mexico

From my model estimates, I found that the estimates for the gender norm index and education level are statistically significantly different from zero. To better understand how they help better adjust the model to Mexico, it is useful to compare the results to those if I had used Cherchye et al. (2012) model without any changes.

The predicted means would be men's leisure at 114.4 hours and women's leisure at 94.4 hours (instead of my model's prediction of 113.9 and 94.6). The model would not have been as good at predicting the changes in total work time through time as shown in Figure 10. we see in blue the data means, in red my model prediction and, now, in the green dotted line the model fit for CDV's model. The CDV model predictions give a disparity of 26.5 hours in 2002, 20.1 hours in 2005/6, and 22.6 hours in 2009/11. With larger over-predictions every year. Particularly, the difference between 2002 and

2005/6 is of 6.4 hours, which is over predicting from the 4.9 hours in the data and my model's prediction.

Figure 10: Model Fit in Comparison to CDV (2012) Model



Chapter 2

The Effect of Violence on Intra-household Decision-Making During the Mexican Drug War

2.1 Introduction

Millions of households live in violent environments. In 2017, over 220 million people lived in close proximity to conflict areas with battle-related deaths (Corral et al., 2020), yet for many more millions of people in non war-torn areas the threat of homicide, rape, or kidnapping represent "everyday" crimes. Households are permeable to the outside environment, which may interfere with families' dynamics and well-being. For example, exposure to violent crime increases the probability of parents hitting their children (Cuartas, 2018) and domestic violence (La Mattina, 2017). Yet, the effects of violence on spouses' decision-making dynamics remains an open question.

Household decision-making are taken through a bargaining process among household members where genders may differ in their preferences, and who holds power may lead to distinct household outcomes. For example, studies have shown that when decision-making power shifts towards women, expenditures related to children, public goods, and women increase (Brown, 2009; and Behrman and Hoddinott, 2005). Most

studies that investigate the impact of violence on spouses dynamics have looked at the effects of forced migration, civil wars, or genocide and have focused on their impact on domestic violence. Only a few studies have investigated the impact of violence on decision-making. For example, La Mattina (2017) finds that women in Rwanda who married after the 1994 genocide had lower decision-making power as well as higher rates of domestic violence. Calderón et al. (2011) finds women in households displaced by conflict in Colombia have higher earnings but experience higher rates of domestic violence and have no change to their bargaining power. It is not obvious that these studies results can be extrapolated to the effects of a more immediate impact of a rise in violent crime in communities and the impact it has on decision-making of already formed couples.

How could violent surroundings impact spouses' decision-making dynamics? Violent surroundings have been shown to increase individual's fearfulness (Gutierrez-Romero, 2016), anxiety (Brown et al., 2019), risk aversion (Brown et al., 2019; Moya, 2018; and Nasir et al., 2020), and conservative preferences (Beall et al., 2016 and Karwowski et al., 2020). In contrast, after long exposure to war conflict, people's preferences have been shown to become more prosocial at the local level (Blattman, 2009; Gilligan et al., 2014; and Voors et al., 2012). It is not straightforward to imagine how these individual changes in mental states and preferences shift intra-household decision-making. Do spouses make more joint decisions in order to join forces in the face of a common outside threat? Do they decide to split decisions in order to lower the mental bandwidth and strain caused by higher anxiety levels? Or do spouses revert to historically more gender spheres of decision making as insecurity and risk-aversion rise and the "default" setting is seemed as "safer"?

In this paper, I study whether an increase in violence, proxied by the homicide rate, in Mexico changes decision-making between spouses. Particularly, I analyze if the increase in homicides from the War on Drugs during its escalation years from 2008 to 2012 impacted the type of decisions each spouse made in the household and their relative decision power. The Mexican context allows to study the impact of a violent surrounding on intrahousehold decision-making as the country saw a large shock in the rise of violent crime, particularly homicides, across municipalities as a consequence of

the Mexican War on Drugs. Furthermore, Mexico has a rich longitudinal dataset, the Mexican Family Life Survey, which surveyed households before and after the escalation of violence and directly asked spouses about which decisions they participate in, such as decisions on food, children's clothing, large expenditures and more. Moreover, I also test possible mechanism through which violent crimes impact spouses' decision-making. To do so, I explore whether personal fear or mental health mediates the impact of violence on decision-making. Finally, I shows that the Great Recession (Dec 2007- June 2009) does not drive the results.

This context and panel dataset provide an ideal setting to study this question for three reasons. First, the increase in violence during the War on Drugs was pervasive. As the army persecuted drug traffickers, it resulted in power struggles between different organized crime groups, which created up to eighty new drug trafficking gangs and spurred an increase in violence that spread not only along previously established drug routes but also across new areas (Dell, 2015). It is estimated that more than 60,000 people died (Miroff and Booth, 2012) and more than 25,000 others disappeared in drug related violence (Booth, 2012) during Calderón's presidency from 2006 to 2012. Second, researchers have argued the surge in the homicide rate to be exogenous (Brown, 2018, and Velásquez, 2020); yet, to be on the cautious side, I also perform the regression analysis using an instrumental variable for the change in homicides. Third, the longitudinal Mexican Family Life Survey provides relevant outcomes for the same households both before and after the surge of violence in Mexico and allows to control for fixed effects of spouses, their households, and municipalities.

Using the differential rise of homicides across Mexican municipalities in the mid and late 2000s, I find three main results. First, I find that for couples in municipalities with a higher homicide rate both husband and wife reduce the number of decisions they participate in. Yet, the number of total decisions does not change because of increases in violence; thus, violence increases the number of decisions that each spouse makes solely and reduces joint decisions. The average increase of 9.3 homicides per 100,000 people over 12-month period between 2005 and 2010 decreased the number of joint decisions by 6.4%.

Second, violence did not uniformly affect all types of decisions in the household. As violence rises, women are less likely to make decisions on male private consumption goods and large expenditures, while men are less likely to decide on children's education and clothing. While the first half of the decade saw an increasing in joint decisions and a slow breakdown of historical gender spheres of decisions, the effects of greater violence vanished some of these changes decreasing joint decisions back to more historical gender spheres. Given that a rise in violence decreases the number of decisions women and men take by similar magnitudes, if considering all decisions with equal weight then an increase in violence keeps the bargaining power of each spouse unchanged. Yet, considering less women decide on large expenditures and the reversal into more historical sphere of decisions, it is likely violence decreases women's bargaining power.

Finally, I find that increases in the homicide rate increased spouses' fear of being robbed. Yet, I find no evidence that feeling fearful or deterioration of mental health of spouses are mechanisms through which the homicide rate decreases the number of joint decisions.

This paper contributes to the understanding of the hidden costs of exposure to violence on household behavior and the role it plays as a determinant in intra-household decision-making. Most studies have focused on the impact of extreme violence, such as in civil wars, or the impact of individual crime on victims. This paper is one of the few studies that explores increases in commonplace violence and the immediate impact it has on spouses' division of decision-making. In closely related papers to mine, Tsaneva et al. (2019) find a rise of homicides in Mexico decrease the number of decisions the wife takes; while, Hernandez-de Benito (2022) finds the increase in homicides decreases the household's share dedicated to female private goods and moving towards male goods. My paper is able to build on their results by simultaneously analyzing the changes to men's decisions in order to capture the overall impact to intra-household bargaining and choices. Specifically, I contribute to this literature by analyzing the gender-differentiated effects of exposure to community violence on each spouse's number and type of decisions.

The remainder of this paper is organized as follows. Section 2 provides background information about bargaining power determinants. Section 3 provides a description of the War on Drugs in Mexico. The description of the household and municipal data can be found in Section 4. Section 5 provides the methodology. Section 6 presents and discusses the main results. Finally, Section 7 considers the relevance of the findings from a policy perspective and concludes.

2.2 Intra-household Decision-making

It is essential to understand the determinants of women's bargaining power, not only because women's empowerment within households is an important development goal in itself, but also because the balance of power between women and men determines household outcomes. Various studies have shown that preferences for men and women differ and, thus, who makes decisions determines household outcomes. For example, women have higher preferences for public good and child investment (Thomas, 1990; Thomas, 1993; Duflo, 2012), as higher bargaining power held by women has been shown to impact positively prenatal care (Beegle et al., 2001), survival rates of girls (Qian, 2008), children's education level (Qian, 2008; Chakraborty and De, 2017), children's nutrition (Majlesi, 2016), family's health (Thomas, 1990), and lower child labor (Reggio, 2011). Some of the previously studied positive determinants of women's bargaining power within the household include: sex-ratios in the marriage market (Angrist, 2002; Chiappori et al., 2002), women's assets (Makino, 2019), labor and non-labor income (Anderson and Eswaran, 2009), employment status (Antman, 2014), female labor demand (Majlesi, 2016), and the outside option value (Chiappori et al., 2002). Our understanding of the impact of exposure to violence in the community on intra-household decisions remains limited.

When trying to measure changes in spouses' bargaining power or estimate the effect of changes in women's bargaining power on household decisions, researchers usually face two sorts of challenges. The first challenge is that researchers do not observe spouses' bargaining power directly. Because of that, the literature usually examines

the changes in household outcomes over which spouses might have different preferences. Examples of these outcomes are spending on men's, women's, and children's clothing (Bobonis, 2009; Lundberg et al., 1997), on alcohol and tobacco (Bobonis, 2009; Wang, 2014), and children's health and education (Duflo and Udry, 2004; Thomas, 1990). The other challenge is to find an exogenous determinant of bargaining power. Using variables that could be correlated with unobserved household characteristics, that directly affect household outcomes over which spouses have different preferences, would lead to biased estimates. I am able to tackle both challenges by using a direct consequence of bargaining power, the number of decisions made within households. Thus, using the number of decisions made within households, this paper tries to shed light on the effects of a large rise of violent crime on decision-making and women's bargaining power.

2.2.1 Intra-household Decision-making and Violence

Most of the literature on violence has focus on either high violent events like civil wars or on individual victimization experiences. In the case of extreme conflicts and violence, these events can create skewed sex ratios and affect marriage markets (Brainerd, 2017; La Mattina, 2017), usually leading to lower bargaining power for women. For example, women that married after the genocide in Rwanda experienced lower decision-making power and higher rates of domestic violence (La Mattina, 2017). It is unlikely that the violence in Mexico affected the sex ratios and males in the marriage market, although it may have affected marriage decisions. Yet, this paper focuses on changes to couples already formed before the onset of higher violence.

In the setting of individual victimization experiences, such as sexual harassment in urban neighbourhoods or domestic violence, the literature has found that female victimization has a negative effect on women's labour force participation (Mishra et al., 2021 and Siddique, 2022). Given that women's labour force participation is an important determinant of bargaining power, this reduction in participation is likely to translate into lower bargaining power of women.

In comparison to war violence and individual victimization, less is known about

medium-size continual non-gendered violent crime incidents and their effects on intra-household decision-making for already formed couples. One of the few papers in this literature and a closely related paper to mine is Tsaneva et al. (2019), they analyze the rise of homicides in Mexico and find that it reduced the number of decisions taken by women. Furthermore, they find evidence that women's leisure decreased as their labor activity increased and suggests violence may disproportionately affect women resulting in decreased mental health and leisure. Yet, without analyzing the changes experienced by men, one cannot capture the overall change to decision-making. Hernandez-de Benito (2022) finds the increase in homicides decreases the household's share dedicated to female private goods and shifts it towards male goods, which is likely a sign of women's lower bargaining power. The effect of non-gendered violent crime on decision-making dynamics within the household remains an open empirical question.

As households are permeable to the outside environment, how could violent surroundings impact spousal decision-making? If the main mechanism of violence is through economic impacts, such as impacting women's labor force participation then controlling for the socio-economic status, changes in income and work opportunities should absorb most of the impact of violence on decision-making. Yet the literature suggests that violence changes preferences and behaviors beyond those related to economic changes. Violent surroundings have been shown to increase individual's fearfulness (Gutierrez-Romero, 2016), anxiety (Brown et al., 2019), risk aversion (Brown et al., 2019; Moya, 2018; and Nasir et al., 2020), and conservative preferences (Beall et al., 2016 and Karwowski et al., 2020). Furthermore, after long exposure to war conflict, people's preferences have been shown to become more prosocial and cooperative at the local level (Bauer et al., 2016; Blattman, 2009; Gilligan et al., 2014; and Voors et al., 2012). Investigating if shifts in preferences are mechanisms through which violence impacts spouses decision-making is difficult. Most household surveys are not designed to capture preferences and those that do tend to capture them with subjective questions where responses can vary in a three or four point scale. Thus, small changes in preferences are likely to be mudded by noise measurement for questions about preferences. Nonetheless, this paper investigates as possible mechanisms

changes in preferences and mental states using the available data.

2.3 Rise of Violence During the War on Drugs in Mexico

The Drug War in Mexico began in December of 2006 when the newly elected president, Felipe Calderón, from the National Action Party (PAN) increased the army presence on the streets and confronted drug trafficking cartels. The drug enforcement policy was spearheaded by the PAN federal government's war on drug trafficking, as well efforts by PAN mayors (Dell, 2015). The sudden increase in the persecution of drug traffickers resulted in power struggles between different organized crime groups, which created up to eighty new drug trafficking gangs and spurred an increase in violence that spread not only along previously established drug routes but also across new areas (Dell, 2015).

The annual number of drug-related homicides increased more than sixfold from 2005 to 2011. The Mexican government estimates that from January 2007 to late 2010 around forty-five thousand homicides took place (roughly twelve per hundred thousand people), more than thirty-two thousand or seventy percent were drug related (Shirk, 2011). In 2010 alone, a Mexican newspaper alone documented more than eleven thousand killings (Shirk, 2011). It is estimated that more than 60,000 people died (Miroff and Booth, 2012) and more than 25,000 other disappeared (Booth, 2012) in drug related violence during president Calderón's presidency (2006-2012). The violence was public and brutal, in some extreme cases with bodies hung from busy overpasses and severed heads placed in public spaces. Besides the public displays of brutality, activities such as kidnapping and extortion affected the general public. This rise in violence impacted more than just the victims of the crimes, it changed the social fabric of communities with a greater exposure to violence. A 2011 public opinion survey found that security was more likely than the economy to be chosen as the largest problem faced by the country (Heimlich, 2011).

None of the households in our dataset suffer any deaths from homicides. Yet even

if people were not victims themselves, Gutierrez-Romero (2016) finds that during this time period overall fear and perceptions of insecurity increased and people living in areas with drug-related homicides were more likely to take extra security precautions. This suggest that fear of bodily harm, kidnapping, extortion and/or sexual assault is one possible mechanism that may lead households to divide decision-making differently.

Recent studies of the impact of Mexico's War on Drugs have also documented the significant negative effects of violent crime on economic outcomes, such as lower growth among businesses, employment, and labor earnings (Dell, 2015; Enamorado et al., 2014; Robles et al., 2013; and Velásquez, 2020). Given that the rise of violence had also an economic impact which in turn itself impacts decision-making, I make sure to control for the economic situation of couples. Thus, I analyze the impact of a rise of violence above that of the impact it may have to economic status, wages, and labor force participation. Furthermore, as a robustness check, I explore in detail the economic shocks experienced by households and the rise of violence.

2.4 Household Survey and Municipal Data

This paper combines household-level data from two waves of the Mexican Family Life Survey and municipality-level data from the Mexican National Institute for Statistics and Geography.

2.4.1 Mexican Family Life Survey and Sample Construction

This paper focuses on the household-level data from two waves of the Mexican Family Life Survey (MxFLS). The MxFLS is a nationally representative longitudinal multi-thematic survey of about an initial 8400 households in 150 municipalities throughout country. The first wave was conducted in 2002 (MxFLS-1), the second (MxFLS-2) was conducted largely during 2005-2006 before the surge in drug-related violence, and the third (MxFLS-3) during 2009-2012 after the surge in violence. I use the first wave of the MxFLS to present the summary statistics of the dependent variables and view the patterns of household and municipality characteristics before the second and third

wave. Yet, I do not use the first wave for the regression analysis as the homicide rates between the first and second wave did not change significantly and the methodology to distinguish the causal effect of homicides on decision making is only applicable between the second and third wave.

Importantly, the survey also include questions on decision-making within the household. The decision-making module is responded by both husband and wife. In this paper, I examine the effect of violence on the total number of decisions the spouses participate in and on the probability of each individual decision. I restrict the sample to married couples interviewed in the second and third wave of the MxFLS. I drop the few households that moved municipalities between the two waves to avoid the problem of selective migration due to the changes in homicides. The remaining sample use for the analysis consists of 3,053 households in 143 distinct municipalities.

Table 2.1: Summary Statistics of Household and Municipality Characteristics

	MxFLS-2		MxFLS-3		Difference
	mean	sd	mean	sd	mean
Characteristics of Individual and Household					
Wife's age	38.997	9.209	43.239	9.264	4.242***
Husband's age	41.770	9.707	46.071	9.716	4.301***
Wife's labor force participation	0.258	0.438	0.300	0.458	0.042***
Husband's labor force participation	0.956	0.204	0.918	0.275	-0.039***
More than two rooms in home	2.186	1.041	2.298	0.987	0.113***
Number of household members	2.703	1.117	3.243	1.443	0.540***
Number of children in household	2.213	1.533	2.078	1.553	-0.135***
Number of children younger than six	0.596	0.789	0.511	0.774	-0.085***
Household owns their home	0.818	0.386	0.843	0.364	0.025**
Household owns a car	0.461	0.499	0.466	0.499	0.006
Household owns a washing machine	0.894	0.308	0.922	0.268	0.028***
Characteristics of Muncipality					
Percentage of female informal workers	0.290	0.078	0.302	0.072	0.012***
Percentage of male informal workers	0.263	0.054	0.269	0.045	0.006***
Percentage of women economically active	0.210	0.019	0.213	0.017	0.003***
Percentage of men economically active	0.364	0.015	0.355	0.014	-0.010***
Percentage of men unemployed	0.012	0.004	0.019	0.006	0.007***
Percentage of women unemployed	0.008	0.003	0.012	0.005	0.003***
Log median income	7.184	0.436	7.411	0.411	0.227***
Gini coefficient (base 2000)	0.276	0.033	0.263	0.031	-0.013***
Food poverty percentage	0.198	0.149	0.201	0.152	0.003
Log total population	11.575	1.617	11.649	1.609	0.075
% of Population without secondary schooling	0.265	0.063	0.039	0.007	-0.226***
Urban Population Ratio	0.547	0.389	0.557	0.384	0.010
Observations	3053		3053		6106

Table 2.1 presents the descriptive statistics for the independent variables for the analysis sample before the surge in violence in MxFLS-2 (2005/6) and after the surge in violence in MxFLS-3 (2009/12). For the individual characteristics, people got older between the waves and wives increased and husbands decreased their labor force participation. The increases in women's labor force participation are likely a combination of more egalitarian gender norms (Carrillo, 2021), and, to an extent, the financial crisis that began in 2008 pushed wives to enter labor market for the specific purpose of protecting against deteriorating income whether from husband's job less or anticipated increases in unemployment (Young, 2014). In terms of household characteristics, the number of households with more than two rooms, and owners of their home and a washing machine increased. At the household level, between the two waves of surveys there are less children under 15 and younger than six, but higher number of members living in the household.

I further complement the information of household characteristics with municipal characteristics from the Mexican National Institute for Statistics and Geography (INEGI) which include information on the rate of the economically active, informal workers, and unemployment by sex, log median income, poverty indices, percentage of population without secondary schooling, and urban population ratio.

Table 2.2 presents the descriptive statistics for the decision-making dependent variables for the analysis sample during the three waves of the MxFLS. The MxFLS module on decision-making asks each spouse who gets to decide on twelve different topics, either themselves alone, jointly, their spouse, or someone else in the household. These topics include decisions on food, spouses' clothing, children's clothing, education and health, large expenditures, money to parents of spouses, spouses' labor force participation, and contraceptive usage.

The average total number of decisions taken by spouses out of the twelve decisions topics is 10.4 in the first wave, 10.2 in the second wave, and 9.6 in the last wave. Some of the decline in the third wave's total number of decisions reflects the natural changes in decisions given the aging of couples. For example, as couples age they are less likely to have a parent alive, have children in the house or need contraception;

Table 2.2: Summary Statistics of Decisions by Husband and Wife Across MxFLS Waves

	Mean MxFLS 1 (1)	Mean MxFLS 2 (2)	Mean MxFLS 3 (3)	Difference MxFLS 2 - 1 (4)	Difference MxFLS 3 - 2 (5)
Number of decisions	10.373	10.160	9.583	-0.21***	-0.58***
Number of decisions of 6 most common	5.320	5.381	5.281	0.06**	-0.10***
Number of joint decisions	5.364	5.953	4.853	0.59***	-1.10***
No. joint decisions of 6 most common	2.309	2.649	2.299	0.34***	-0.35***
Wife's number of decisions	8.183	8.353	7.514	0.17*	-0.84***
Husband's number of decisions	7.554	7.760	6.922	0.21**	-0.84***
Husband's bargaining power	0.469	0.479	0.477	0.01*	-0.00
12 Decisions					
Wife- Food	0.965	0.982	0.972	0.02***	-0.01*
Husband -Food	0.407	0.354	0.333	-0.05***	-0.02
Wife- Female clothing	0.918	0.937	0.928	0.02**	-0.01
Husband - Female clothing	0.601	0.655	0.632	0.05***	-0.02
Wife- Male clothing	0.647	0.614	0.559	-0.03*	-0.05***
Husband - Male clothing	0.570	0.624	0.574	0.05***	-0.05**
Wife- Child clothing	0.920	0.918	0.887	-0.00	-0.03***
Husband - Child clothing	0.530	0.626	0.552	0.10***	-0.07***
Wife- Child education	0.918	0.943	0.907	0.02***	-0.04***
Husband - Child education	0.814	0.871	0.796	0.06***	-0.08***
Wife- Child health	0.913	0.934	0.910	0.02**	-0.02**
Husband - Child health	0.787	0.867	0.784	0.08***	-0.08***
Wife- Large expenditures	0.673	0.723	0.684	0.05***	-0.04***
Husband -Large expenditures	0.909	0.949	0.926	0.04***	-0.02***
Wife- Money to wife's parents	0.779	0.825	0.782	0.05***	-0.04***
Husband - Money to wife's parents	0.792	0.795	0.743	0.00	-0.05***
Wife- Money to husband's parents	0.663	0.680	0.596	0.02	-0.08***
Husband - Money to husband's parents	0.817	0.821	0.810	0.00	-0.01
Wife-Wife works	0.716	0.780	0.772	0.06***	-0.01
Husband - Wife works	0.831	0.854	0.822	0.02*	-0.03***
Wife- Husband works	0.487	0.513	0.409	0.03	-0.10***
Husband - Husband works	0.843	0.851	0.829	0.01	-0.02*
Wife- Contraceptives	0.883	0.897	0.857	0.01	-0.04***
Husband - Contraceptives	0.826	0.891	0.851	0.07***	-0.04***
Observations	2684	3053	3053	5737	6106

Note: The table includes the available decisions of the 2002 wave for those couples that were kept for 2005/6 and 2009/12. Means on the individual decisions are conditional on that being a decision taken by at least one of the spouses.
* $p < 0.05$, ** $p<0.01$, *** $p<0.001$

thus, the decisions on these topics are less likely to be applicable to the couple. There are six out of the twelve questions that most households make a decision on: food, female clothing, male clothing, large expenditures, and labor force participation of each spouse. As can be seen in Table 2.2, the number of decisions of the six most common are more stable.

The number of joint decision rose between the first and second wave but decrease between the second and third. The decline in joint decisions in the third wave also applies to the six most common decisions. Similarly, both husband and wife's number of decisions increased between the first and second wave, but decreased between the second and third wave. The bargaining power slightly decreased for men between the first and second wave, but had no change in the third wave.[1] The second part of Table 2.2 shows the disaggregated twelve individual decisions that spouses can have decision power over. A noticeable contrast between the three waves is that between the first and second wave there is a pattern of increasing the number of decisions that spouses partake in across most topics, while between the second and third wave there is a general pattern of taking less decisions. For example, of the households that make a decision on children's education, in the first wave about 81.4% of husbands and 91.3% of women took this decisions, it increases in the second wave to 87.1% and 93.4%, and finally decreases in the third wave to 79.6% and 90.7%.

2.4.2 Municipal Data and the Rise in the Homicide Rate

The homicide data at the municipal-level data used for this paper is taken from the Mexico's National Institute for Statistics and Geography (INEGI). Table 2.3 presents the changes in the homicide rates in the 143 distinct municipalities of our sample of 3,053 households. Given that households in a same municipality were surveyed in different months or years, there is homicide rate variation across space and time in municipalities. The number of homicides per 100,000 inhabitants increased by 5.6 homicides on average from 4.7 to 10.3 in a span of six months between the second

[1]The husband's bargaining power is calculated as the ratio of number of decisions taken by the husband over the sum of decisions taken by the husband and the wife.

and third wave of MxFLS, while the increase was of 10.04 homicides in a 12 month period. Since the sample in the MxFLS is a representative sample of the population, it is a reassuring fact that the changes in the homicide rate are close to the municipality population weighted averages of the entire country. The unweighted means of municipalities are much smaller, meaning that most of the increase in homicides happen in more populous municipalities.

Table 2.3: Homicide Rate Statistics Comparing the Study Sample to the Average Municipality

	2005		2010		Difference
	Mean	SD	Mean	SD	Mean
Sample unweighted means					
Homicide rate past 3 months	2.37	3.16	5.54	8.64	3.17***
Homicide rate past 6 months	4.67	5.27	10.26	14.22	5.59***
Homicide rate past 12 months	8.86	10.33	18.90	24.02	10.04***
Household observations	3053		3053		6106
All municipalities unweighted means					
Homicide rate in past 3 months	2.22	10.51	3.24	11.37	1.02**
Homicide rate in past 6 months	4.05	12.45	6.84	25.92	2.79***
Homicide rate in past 12 months	8.25	19.45	12.79	44.22	4.54***
All municipalities population weighted means					
Homicide rate in past 3 months	2.10	3.42	5.02	9.54	2.92***
Homicide rate in past 6 months	4.17	5.68	9.84	18.87	5.67***
Homicide rate in past 12 months	8.43	10.67	17.73	30.87	9.30***
Number of municipalities	2172		2172		4344

Note: The homicide rate in all municipalities are computed starting on January 2010 and going backwards for the specific number of months and comparing them to January 2005 going the same months backwards.

2.5 Methodology

One of the challenges of identifying the causal effect of crime on the number of decisions taken by spouses is that there may be other individual, household or municipal unobserved factors that affect both these outcomes and cause spurious correlations. For example, crime may be correlated with poverty and work opportunities which are also correlated to women's bargaining power. In order to deal with this problem, I take advantage of the longitudinal nature of the dataset and use an individual and municipality fixed effects model to estimate the effect of violence on decision-making. Where the model for total number of decisions, D, of household respondent i of sex s

living in municipality j in month m and year t is as follows:

$$D_{i,s,j,m,t} = \beta_{1,s} + \beta_{2,s}\, Violence_{j,m,t} + \mathbf{X_{i,j,m,t}}\, \beta_{3,s} + \mathbf{H_{j,m,t}}\, \beta_{4,s} + \mathbf{R_{j,t}}\, \beta_{5,s} + \eta_{i,s} + \mu_t + \epsilon_{i,s,j,m,t} \tag{2.1}$$

where $Violence$ is the homicide rate in the prior months to the survey month m, η is the individual fixed effects, and μ is the survey year fixed effects. Individual characteristics, $\mathbf{X}$, include age and labor force participation. Household characteristics, $\mathbf{H}$, measured at month m and year t that include the number of years living in their house, the number of children under 15, the number of children under 5, household size, ownership of house, car and washing machine. Municipality characteristics, $\mathbf{R}$, control for log median income, log security expenditures, log total population, asset poverty index, food poverty index, gini coefficient, the rate of economically active, informal employment, and unemployment for both genders.

Even with the advantage of the fixed effects model, which accounts for any time invariant individual and locality unobservable characteristics, there still remains concerns of the potential endogeneity of $Violence_{j,t}$. The sudden increase in persecution of drug traffickers resulted in power struggles between different organized crime groups that spurred an increase in violence not only in previously established drug routes but also across many new areas. Brown (2018) formally addresses the exogeneity of the increase in municipal homicide rates. He finds no evidence that pre-escalation trends in 2005 predicted 2009 municipal homicide rates or the change in rates between these two years.

Yet, the possibility remains that unobservable factors that affected which municipalities had more power struggles between drug traffickers and thus higher homicide rates could be correlated with unobservable factors that affect the decision-making of a couple. To combat the possible sample selection problem, I use an instrument that conditional on appropriate controls determines the treatment status or the homicide rate, but uncorrelated with the error term, meaning the outcome of couple's decision-making process. Dell (2015) finds that there are increases in drug-related violence and homicides during 2007-2010 when municipalities have mayors from the PAN political

party as their party spearhead the efforts against drug trafficking and rival traffickers' attempted to take over territories after government crackdowns weakened incumbent criminals. Taking the finding of Dell (2015), I use as an instrumental variable a dummy variable equal to one if the mayor of the municipality belong to the political party of PAN during the year that the household was surveyed during the 2009-2012 wave. Given that the instrumental variable only applies to the third wave or the change between the second and third wave, I use the first difference of equation (1) between the 2005/06 and 2009/12 wave of the survey when using the instrumental variable.

2.6 Results

In this section, first, I present the effects of the homicide rate on the total number of decisions taken in a household, total number of decisions by each spouse, number of joint decisions, and relative bargaining power of the husband. The results are explored using a fixed effect regression (FE), comparable to the previous literature, and a first-difference instrumental variable regression (FD-IV). Next, I estimate the effects of changes in the homicide rate on individual topic decisions. Then, I test possible mechanisms through which changes in homicides affects decision-making in the household. Finally, I studied an extension about victimization and joint decisions, as well as verifying that the Great Recession and the worsening economic situation does not drive the main effects of the paper.

2.6.1 The Impact of Violent Crime on the Number of Decisions and Bargaining Power

I begin by examining whether an increase in the homicide rate changes the total number of decisions taken in the household. Table 2.4 presents the point estimates from Equation 2.1 and its first difference with an instrumental variable for the homicide rate. The homicide rate is explored using three, six, and twelve months before the interview month. Columns 1 and 2 present the impact of the homicide rate on the total number of decisions. Controlling for individual, household, and municipality

characteristics (see Table 2.1), the homicide rate has a small negative impact on the total number of decisions taken but this effect loses its statistical significance when using the instrumental variable.

Table 2.4: Impacts on Spouses Decision-Making from Homicide Rates

	Total no. of decisions		Wife's no. of decisions		Husband's no. of decisions	
	(1) FE	(2) FD-IV	(3) FE	(4) FD-IV	(5) FE	(6) FD-IV
Homicide rate past 3 months	-0.008	-0.036	-0.015***	-0.056**	-0.004	-0.058*
	(0.005)	(0.027)	(0.005)	(0.027)	(0.006)	(0.030)
Homicide rate past 6 months	-0.008**	-0.021	-0.007**	-0.033**	-0.002	-0.034*
	(0.003)	(0.016)	(0.003)	(0.016)	(0.004)	(0.017)
Homicide rate past 12 months	-0.004**	-0.012	-0.005**	-0.019**	-0.002	-0.019*
	(0.002)	(0.009)	(0.002)	(0.009)	(0.003)	(0.010)
DV Mean 2005/6	10.16	10.16	8.35	8.35	7.76	7.76
	No. of joint decisions		Husband's bargaining power			
	(7) FE	(8) FD-IV	(9) FE	(10) FD-IV		
Homicide rate past 3 months	-0.019**	-0.114***	0.000	-0.000		
	(0.008)	(0.040)	(0.000)	(0.002)		
Homicide rate past 6 months	-0.008	-0.067***	0.000	-0.000		
	(0.005)	(0.022)	(0.000)	(0.001)		
Homicide rate past 12 months	-0.007**	-0.038***	0.000	-0.000		
	(0.003)	(0.012)	(0.000)	(0.001)		
DV Mean 2005/6	5.95	5.95	0.48	0.48		
N	6106	3053	6106	3053	6106	3053

All regressions have as covariates the characteristics in Table 2.1 and, furthermore, columns 3 to 10 have as a control the total number of decisions taken in the household. For the entire regression refer to the Appendix A. Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Columns 3 and 4 explore the impact of the homicide rate on the number of decisions taken by the wife, and columns 5 and 6 repeat the process for the husband's decisions. Controlling for individual, household, and municipality characteristics as well as the total number of decisions in a household, I find that an additional homicide per 100,000 people in the past three months decreases the number of decisions the wife participates in by 0.015, by 0.007 for six months, and 0.005 for twelve months.[2] Using the sample's average homicide rate in municipalities during the third wave of

[2]Table 11 in the Appendix presents the full regression estimation showing the effects of the different characteristics on decision-making.

the survey (see Table 2.3), the average homicide rate would decreases women participation by 0.07 decisions for 3 months, 0.08 decisions for 6 months, and 0.09 decisions 12 months, the latter representing a 1.1% decrease in the wife's average decisions, compared if the couple lived in a municipality without homicides.

In column 4, I repeat the exercise using an instrumental variable to deal with the potential endogeneity of the homicide rate. The instrument consists of a dummy variable equal to one if during the year of the third wave survey the mayor of the household's municipality belonged to the political party of PAN. The effort by the government to eliminate drug trafficking organizations was spearheaded by the Mexican president belonging to the PAN political party, who worked more closely with mayors of his own political party. The dummy variable of whether the municipality has a PAN mayor is significant at the 0.1% level and increases the homicides by 10.5 homicides per 100,000 people; furthermore, the F-statistic demonstrates that the instrument has sufficient power (see Table 10 in the appendix). Using the instrumental variable,0 the magnitudes for the homicide rate more than triples. With the instrumental variable, the average impact of the twelve-month homicide rate in 2009/12 decreases the wife's decisions by 4.3% compared to living in a municipality with a zero homicide rate. These estimates are the first divergence with the results from Tsaneva et al. (2019). Although the magnitudes of the fixed effect regression are similar to those of Tsaneva et al. (2019), the results from the instrumental variable regression suggest that the estimates of the fixed effect regression are biased downwards.[3]

The second difference with Tsaneva et al. (2019) comes from the exploration of the husband's number of decisions in Table 2.4 columns 5 and 6. With an additional homicide per 100,000, the decrease in the number of decisions by the husband shown in column 5 is smaller than those of the wife and not statistically significant. The impact of the homicide rate on the husband's decisions is larger when considering the first difference instrumental variable regression, like in the wife's case. With the reduction

[3]Given the dependent variables in columns 1, 3, 5 and 7 are count variables, a Poisson regression would be better suited for the analyses. However, using a Poisson regression would not allow for a simple comparison with the first difference instrumental variable results. Furthermore, keeping the fixed effect regression allows the results to be comparable to those of Tsaneva et al. (2019). Results from a Poisson estimation are shown in Table 12 of the appendix and are qualitatively similar to those in Table 2.4.

of decisions for each additional homicide per 100,000 people being 0.058 for 3 months, 0.034 for 6 months, and 0.019 for 12 months. These negative impacts are similar to those experienced by wives.

Given the previous results that both husband and wife reduce their decisions when controlling for the total number of decisions, it is consistent to see that as homicide rates increase the number of joint decisions between spouses decreases while controlling for the total number of decisions (see Table 2.4 columns 7 and 8). I find one extra homicide per 100,000 people over a 12-month period decreases the number of joint decisions by 0.038. Thus, the average increase of 10.04 homicides in the twelve-month homicide rate during the War on Drugs caused couples to decrease the number of joint decisions by 0.38 or 6.4% from the 2005/6 baseline. Moreover, for the average municipal 12-month homicide rate during 2009/12, a couple decreases joint decisions by 0.72 or 12.1% than if they would live in a zero homicide municipality.

Did the changes in decisions caused by the increases in the homicide rate change the balance of power between spouses? I define the husband's bargaining power as the ratio of the number of decisions the husband takes divided by the sum of decisions taken by the husband and wife, where the wife's bargaining power is one minus the husband's bargaining power. This particular bargaining power assumes equal weight to the 12 distinct decisions asked in the survey hold. Given that homicides decrease wife's and husband's number of decisions evenly, it is no surprise I find a magnitude close to zero and not statistically significant for the effect of homicides on husband's bargaining power (see columns 9 and 10 in Table 2.4). Thus, Table 2.4 showed that as homicides increase the husbands and wives reduce the number of decisions they participate in, which reduces joint decisions but bargaining power seems to be left unchanged. Given the strong assumption that all types of decisions hold the same weight, I analyse each decision separately in the next section to get a better sense of the changes in decision-making and possible changes in bargaining power.

Table 2.5: Violence Impact on Individual Decisions

	Food		Women's clothes		Men's clothes		Women's work		Men's work		Large expenditures	
	Wife	Husband	Wife	Husband	Wife	Husband	Wife	Husband	Wife	Husband	Wife	Husband
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Homicide rate	-0.000	-0.001*	-0.000	0.000	-0.002**	0.001	-0.000	-0.001	0.001	-0.000	-0.002***	-0.000
past 12 months	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
DV Mean	0.98	0.34	0.93	0.64	0.59	0.60	0.78	0.84	0.46	0.84	0.70	0.94
N	6007	6007	6079	6079	3722	3722	6040	6040	4705	4705	6019	6019
	Child Clothes		Child education		Child health		Money woman's parents		Money man's parents		Contraceptives	
	Wife	Husband	Wife	Husband	Wife	Husband	Wife	Husband	Wife	Husband	Wife	Husband
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
Homicide rate	0.000	-0.002**	0.000	-0.001**	0.000	-0.001	0.000	0.000	-0.001	0.001	-0.000	-0.001**
past 12 months	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
DV Mean	0.90	0.59	0.93	0.84	0.92	0.83	0.80	0.77	0.64	0.82	0.88	0.87
N	3888	3888	5074	5074	5204	5204	4735	4735	3916	3916	4927	4927

Note: All fixed effects regressions have as covariates the characteristics in Table 2.1. Robust standard errors presented in parenthesis.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

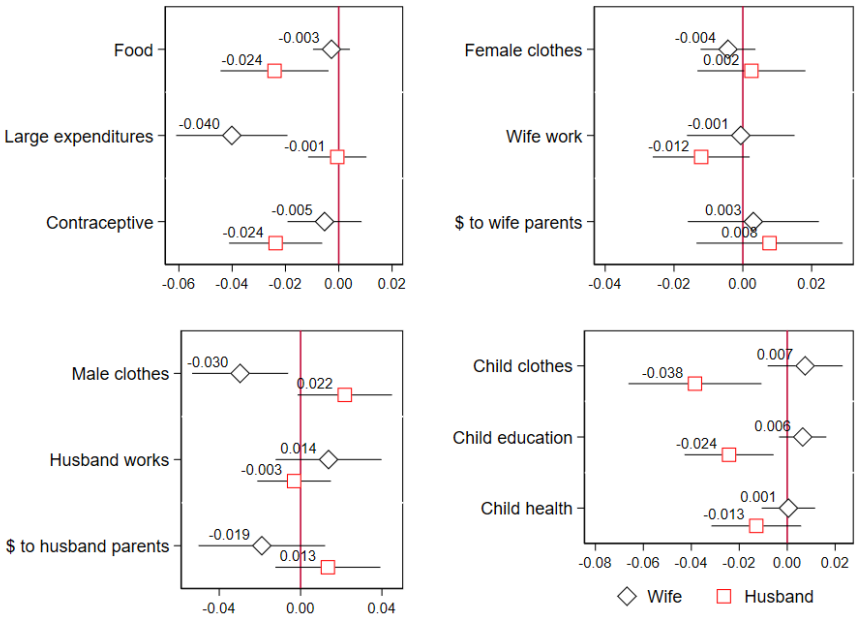
2.6.2 Which decisions are more likely to be taken separately?

In this section, I present the point estimates on the impact for each of the twelve decisions from a fixed effects equation interpreted as a linear probability model.[4] Table 2.5 provides the estimates for each decision, and Figure 2.1 provides a visual representation of these estimates using the sample's twelve-month mean homicide rate. I find that various decisions drive the decrease in the number of decisions by both spouses. An increase in the homicide rate decreases the probability that husbands decide on food, contraceptives, children's clothes and children's education. In contrast, an increase in the homicide rate decreases the probability of the wife deciding on male clothes and large expenditures. For example, Figure 2.1 shows that wives living in a municipality with the average homicide rate during the third wave are four percentage points less likely to decide on large expenditures than if they lived in a municipality with a zero homicide rate. In contrast, men would be 3.8 and 2.4 percentage points less likely to decide on children's clothes and education. The results of increases in homicides reducing women's decisions over large expenditures and male clothing are also found by Tsaneva et al. (2019).

Furthermore, decreases in joint decisions seem to revert decisions back into historical gender spheres. Women in the data across the waves are more likely to decide on children and food-related decisions, while men are more likely to make decisions on large expenditures. The only exception to this pattern is contraceptives, where both usually always share the decision. It is interesting to note that the first decade of the 2000s saw gender norm perceptions move towards a more egalitarian society (Carrillo, 2021) with women increasing their bargaining power within the household. Particularly, there was a rise in joint decision-making between the first and second wave of the MxFLS, particularly concerning public good decisions (see Table 2.4); yet, from the second wave to the third wave, the opposite pattern is observed with a decrease in joint decisions. Thus, the effects of greater insecurity in the country vanished some of the changes achieved in the first part of the decade concerning an increase in share

[4] A probit model with fixed effects would better account for the binary nature of the dependent variable of interest, being a decision maker-1 or not-0 for a single topic. However, to make the results comparable to Tsaneva et al. (2019) I keep a fixed effect model.

Figure 2.1: The Effects of the Average Homicide Rates in 2009/12 on Individual Decisions



Note: Estimates from the figure come from Table 2.5 multiplied by the sample average municipal homicide rate from the MxFLS third wave of 18.9 homicides per 100,000 people over a 12 month period. Confidence intervals at 90% level.

decisions and thus diminishing historical gender spheres. Interestingly, the reversal back to more historical gender spheres of decision-making was not accompanied by a reduction in women's bargaining power when considering all decisions have equal weight. Yet, given the importance of deciding on large expenditures and women's lowering decision power on this topic, it could reflect a lower bargaining power that would go in line with Hernandez-de Benito's (2022) results.

2.6.3 Possible Mechanisms

This subsection examines potential mechanisms that could explain the relationship between the surge in violent crime and changes in joint decisions in the household. As possible mechanisms, I look at changes in fear of being robbed, fear towards some things in the past four weeks, self-reported probability of being robbed in the next 12 months, and mental health. For the analysis of possible mechanisms in Table 2.6, I pool husbands' and wives' responses for the analysis.

Table 2.6: Impacts on Fear and Mental Health State from the Homicide Rate

	Fear of being robbed		Fear of things		Likelihood of being robbed		Anxiety or depression	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	FE	FD-IV	FE	FD-IV	FE	FD-IV	FE	FD-IV
Homicide rate	0.003***	0.014***	0.001	-0.002	0.002*	-0.005	0.000	-0.002
	(0.001)	(0.004)	(0.001)	(0.002)	(0.001)	(0.004)	(0.000)	(0.001)
DV Mean - Women	1.80	0.01	1.19	0.05	1.86	0.06	0.16	0.01
DV Mean - Men	1.57	0.09	1.19	0.05	1.88	0.05	0.09	0.01
N	11956	5974	11912	5947	11950	5968	11912	5930

Dependent variable (DV) mean is computed for the baseline year for the fixed effect regression and is the average change for the first difference instrumental variable regression. Fear of being robbed and likelihood of victim can be integers from 1-feel no fear/improbable to 4-very fearful/very probable. Standard errors clustered at the municipal level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The first column in Table 2.6 explores as the dependent variable the extent that individuals experience fear of being robbed during the night. Respondents were asked if they felt (1) no fear, (2) little fear, (3) somewhat fearful, or (4) very fearful of being robbed during the night.[5] Although the responses in this 4-point scale are categorical, given the values can be seen as equidistant in terms of agreement with the statement, I treat the responses as integers. I find that the homicide rate increased the feeling of fear for spouses. When using the instrumental variable, the magnitude of the impact is more than quadrupled in line with the downward bias seen before during the fixed effect regressions.

The next mechanism explored is the extent that spouses have felt fear of some things in the past four weeks. The question is kept broad, so answers could be related to insecurity related to violence or other fears. The responses range from a 4-point scale (1) none, (2) sometimes, (3) many times, and (4) all the time. Although there was a rise in the number of people feeling fear of things in the third wave, there is no evidence that the homicide rate impacts this variable.

Similarly to self-reported fear, there is a question about their perception of the probability of being robbed in the next twelve months. Respondents can answer they feel it is (1) improbable, (2) a little probable, (3) probable, and (4) very probable.[6]

[5] At baseline, about 66% of spouses feel no fear, 19% little fear, 12% somewhat fearful, and 3% very fearful.

[6] At baseline, about 46% of spouses feel it is improbable, 30% a little probable, 19% probable, and 5% very probable.

The probability of thinking you will be an assault victim in the next twelve months increases as the homicide rate increases. Yet, the estimate becomes noisier and losses statistically significance when implementing the instrumental variable.

The MxFLS contains an emotional well-being module. Within the module, there are 20 questions where individuals self-report the level of experiencing depressive symptoms, sadness, or anxiety over the past four weeks. Using all responses and following the normal mental range score created by the researchers that designed the module, I created a dummy equal to zero for a normal level of anxiety and one for an anxiety or depressive state.[7] The magnitudes of the effects for anxiety or depression in both the fixed effect and first difference instrumental variable regression are quite low and not statistically significant.

The results from Table 2.6 suggest that an increase in the municipality's homicide rate increases fear of being robbed, but I find no evidence that it increases anxiety or depression. Is the individual's change in fear of being robbed and therefore of their feelings of insecurity in the municipality a mechanism through which a rise in homicides impacts joint decisions? To find out, I include the four individual perceptions from Table 2.6 as regressors in the main regression of the impact of homicides on the number of joint decisions. Table 2.7 presents the results of these regressions. One would expect that including one of the main mechanisms through which the homicide rate impacts joint decisions as a regressor in the model would then reduce the magnitude of the effect of the homicide rate. Furthermore, one would expect a main mechanism added as a regressor to have a large enough (relative to the homicide rate magnitude) and statistically significant magnitude. I explore the individual mechanisms separately for the changes experienced by the husband and wife.

The number of observations in Table 2.7 is smaller than that of main Table 2.4 given some missing information in the new regressors. The sample is kept constant

[7]The indicator for depressive symptoms is the Calderón depression score designed specifically for the Mexican context (Calderon-Narvaez, 1997). The depression score is calculated using 20 questions with a 4-point scale from the emotional well-being module of the MxFLS. The final scale adds these values leaving a range from 20 to 80. Following Calderon-Narvaez (1997), the score can be interpreted as follows: 20-35 = Normal levels of depression and anxiety; 36-45 = A low level of anxiety; 46-65 = Moderate depression; 66-80 = Severe depression. Only scores above 45 are qualified as clinical depression.

Table 2.7: Possible Mechanism for Homicides to Impact Joint Decisions

	No. of joint decisions				
			FD-IV		
	(1)	(2)	(3)	(4)	(5)
Homicide rate 12 months	-0.040***	-0.042***	-0.041***	-0.040***	-0.040***
	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)
Fearful of being robbed × wife		0.082*			
		(0.046)			
Fearful of being robbed × husband		-0.006			
		(0.053)			
Fearful of things × husband			-0.113		
			(0.093)		
Fearful of things × wife			-0.159**		
			(0.081)		
Probability being robbed × wife				-0.009	
				(0.047)	
Probability being robbed × husband				-0.025	
				(0.047)	
Anxiety or depression × wife					-0.047
					(0.128)
Anxiety or depression × husband					0.051
					(0.160)
N	2895	2895	2895	2895	2895

Note: All first difference instrumental variable regressions have as covariates the total number of decisions taken by spouses and the characteristics in Table 2.1. Robust standard errors presented in parenthesis.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

throughout the specifications for comparison purposes. As shown in Table 2.6, feeling fearful of being robbed during the night increases as the homicide rate increases in the municipality. The inclusion of being fearful of being robbed by the spouses in column 1 only slightly increase the magnitude of the homicide rate rather than taking away some of its predictive power, and it is itself not statistically significant. The inclusion of the regressors of being fearful of things by the wife in column 2 is statistically significant and negatively impacts the number of joint decisions. The husband's negative magnitude is of similar size but not statistically significant. With the inclusion of these regressors, there is only a slight increase in the magnitude of the homicide rate over twelve months. Thus, even though being fearful of things seems to negatively impact joint decisions, there is no evidence that this a mechanism through which the homicide rate impacts the number of joint decisions. For columns 3 and 4, the regressors of the probability of being robbed and anxiety/depression have a negative tendency but are otherwise not statistically significant.

Thus, I find no evidence that individual fear of robbery, a fearful state of mind,

perceptions about robberies, or depression/anxiety are mechanisms through which the homicide rate impacts intra-household decisions making. Yet, the lack of evidence may be due to the the small four point-scale in which these variables are measured in the survey, which might not be rich enough to capture changes in these attitudes or mental states. Furthermore, the variable of fear only considers individual fear and not the fear for the safety other household members. Warr and Ellison (2000) found that altruistic fear or fear for others has a distinctive structure in family households and is more common and often more intense than personal fear. This paper leaves for future research the further exploration of viable mediums through which a more violent or unsafe surrounding may reduce joint decisions between spouses.

2.6.4 Extension: Victimization and Joint Decisions

Can the finding that a rise in the homicide rate decreases joint decisions be extended for households that become victims of a violent crime? The MxFLS survey has a module on victimization and records all the crimes experienced by the household. I built a dummy equal to one if at least one spouse experienced a crime that they declared as either very serious or serious crime in the past 36 months from the interview date. I focus on the changes in victimization for households between the first (2002) and second wave (2005/6) before the rise in homicides from the War on Drugs. I focus on households that in 2002 had not experienced a crime in the past twelve months. 115 households or 4% of the sample were victims of very serious or serious crimes in the 36 months before the second-wave surveyed month.

Table 2.8: Impact on Joint Decisions from Victimization

	Poisson Fixed Effect
Dependent variable: no. of joint decisions	(1)
Crime victim in the past 36 months	-0.109**
	(0.050)
N	6294

Note: All regressions have as covariates the total number of decisions by the couple and those shown in Table 2.1. Being a crime victim refers to at least one of the spouses was a victim of a crime of a serious nature in the past 36 months. Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.8 presents the results from a Poisson fixed effect household regression using

the survey waves of 2002 and 2005/6. I find that households that experienced serious victimization take less joint decisions. Having at least one of the spouses be a victim of a serious crime in the past 36 months lowers the number of joint decisions by 0.11 or 2%. Thus, it seems that both exposure to higher violence and being a victim oneself decrease the number of joint decisions taken by spouses. Interestingly, the magnitude of the effects of three more homicides per 100,000 inhabitants in your municipality surpasses the effect of at least one spouse being a victim of a serious or very serious crime on the number of joint decisions. This points to the importance of an individual's expectation of experiencing violence and its impact being similar or even greater than the experience of victimization and violence itself on intra-household decision-making.

2.6.5 Robustness Check: Recession and Worsening Economic Situation

In this subsection, I explore the potential confounding variable of the worsening economic situation during and in the aftermath of the global financial crisis. In 2008, Mexico's real GDP expanded by only 1.2%, much lower than the 3.3% attained the year before, and in 2009, it decreased by 6.1%, the sharpest drop in real GDP in Latin America (Moreno-Brid and Padilla-Pérez, 2012). Unemployment also increased significantly to 5.7% in the second quarter of 2008. Despite an immediate but slight decrease in the following quarters, the unemployment rate took almost ten years to return to pre-crisis levels (Lopez-Acevedo et al., 2020).

If the differential impact of the financial crisis is correlated with the geographic heterogeneity in crime and its correlation is not well controlled for in the regression, the estimates for the homicide rate would be biased. Although a number of papers have found that the differences in the economic impact of the global financial crisis are uncorrelated with the rise in the homicide rate across municipalities (Ajzenman et al., 2015 and Velásquez, 2020), I still provide evidence that the results are not driven by the differential impact of the recession during the third wave of the survey.

Studies have shown that the financial crisis negatively impacted total employment with small differences in how the recession affected women and men. Men were more

likely to stay in their jobs, while women were proportionately more affected during the fall but also more favored during the recovery (Freije et al., 2011). Table 2.9 explores the negative impacts to the household's finances or economic stability using three different variables. The first variable takes the logarithm of the household's self-reported assets, the second and third variables asked the spouses if the household's earnings decreased or experienced a loss of job as a consequence of the financial crisis. The number of observations is smaller than that of main Table 2.4 given some missing information in the new regressors explored. To have the twelve-month homicide rate coefficient be comparable across different specifications the sample is kept constant throughout.

Table 2.9: Robustness Check: Impact on Joint Decisions from Economic Downturns in 2009-2011

	Dependent variable: number of joint decisions			
	FD-IV			
	(1)	(2)	(3)	(4)
Homicide rate 12 months	-0.036***	-0.036***	-0.037***	-0.036***
	(0.013)	(0.013)	(0.013)	(0.013)
Logarithm assets		0.022		
		(0.032)		
Crisis affected household by lowering earnings			0.106	
			(0.119)	
Crisis affected household by loss of job				-0.219*
				(0.131)
N	2762	2762	2762	2762

Note: All regressions have as covariates the total number of decisions by the couple and those shown in Table 2.1.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.9 shows that while controlling for the labor force participation of spouses like in all other result tables, the impact of the logarithm of the household's assets and a dummy if the crisis lowered earnings on the number of joint decisions cannot be distinguished from zero. In contrast, if a couple experiences a job loss because of the financial crisis it reduces the number of joint decisions. Yet, the loss of a job did not reduce the magnitude of the impact of the homicide rate. Thus, although as the War on Drugs was happening there was also a worsening of the economic stability of couples, yet this is not driving the main impacts found of the homicide rate on the number of joint decisions.

2.7 Conclusion

Households are permeable to the outside environment, which may interfere with families' dynamics and well-being. The findings of the paper show that a violent environment in a municipality, even when it does not create victims of the household members, impacts their decision-making dynamics. Using a panel dataset of Mexican households and the differential rise of homicides across Mexican municipalities from Mexico's War on Drugs, this paper identifies the effect of homicide rates, a proxy for a violent environment, on the number of decisions taken by each spouse. I find that increases in the homicide rate decrease the number of decisions taken by the husband and the wife, leading to a lower number of decisions taken jointly. This is important as it may lead to different household outcomes.

Investigating individual decisions reveals that, as the number of homicides increases, women reduce their decisions on male private goods and large expenditures. In contrast, men reduce their decisions on expenditures related to children and food. The impact of violence pushes against the historical trends of corroding gender spheres in decision-making. Although violence decreases joint decisions and strengthens traditional gender spheres of decisions, I do not find a reduction in women's bargaining power when considering all decisions equally. Yet, the reduction of women's decision-making for large expenditures and the reversal to more gender spheres of decision-making, pair with the results of Hernandez-de Benito's (2022) that find lower share of expenditures on women might suggest increases in violence decrease women's empowerment in the household.

The findings in this paper also open several important avenues for future research. To begin with, what are the mechanisms through which violence affects intrahousehold decision-making? The results showed that higher homicide rates stimulate negative emotional changes where women and men are more likely to feel afraid. Although I do not find evidence that fear nor mental health are the main mechanisms, further exploration should take place. The small four point-scale in which fear and mental variables are measured in the survey might not be rich enough to capture changes in these attitudes or mental states. Second, what are the downstream effects on

household outcomes from having fewer joint decisions? Particularly, women's loss of decision power in large expenditures in the household may lead to very different long-term impacts. Finally, I have shown the impact of homicide rates on decision-making; however, citizens are usually not aware of official statistics, and the perception of the amount of violence might be more closely related to the news. Understanding the difference between official homicide statistics and exposure to the amount of homicide-related news could help understand how perceptions about safety are built.

Bibliography

Ajzenman, N., Galiani, S., and Seira, E. (2015). On the distributive costs of drug-related homicides. *The Journal of Law and Economics*, 58(4):779–803.

Anderson, S. and Eswaran, M. (2009). What determines female autonomy? evidence from bangladesh. *Journal of Development Economics*, 90(2):179–191.

Angrist, J. (2002). How do sex ratios affect marriage and labor markets? evidence from america's second generation. *The Quarterly Journal of Economics*, 117(3):997–1038.

Antman, F. M. (2014). Spousal employment and intra-household bargaining power. *Applied economics letters*, 21(8):560–563.

Bauer, M., Blattman, C., Chytilová, J., Henrich, J., Miguel, E., and Mitts, T. (2016). Can war foster cooperation? *Journal of Economic Perspectives*, 30(3):249–74.

Beall, A. T., Hofer, M. K., and Schaller, M. (2016). Infections and elections: Did an ebola outbreak influence the 2014 us federal elections (and if so, how)? *Psychological Science*, 27(5):595–605.

Beegle, K., Frankenberg, E., and Thomas, D. (2001). Bargaining power within couples and use of prenatal and delivery care in indonesia. *Studies in family planning*, 32(2):130–146.

Behrman, J. R. and Hoddinott, J. (2005). Programme evaluation with unobserved heterogeneity and selective implementation: The mexican progresa impact on child nutrition. *Oxford bulletin of economics and statistics*, 67(4):547–569.

Blattman, C. (2009). From violence to voting: War and political participation in uganda. *American political Science review*, 103(2):231–247.

Bobonis, G. J. (2009). Is the allocation of resources within the household efficient? new evidence from a randomized experiment. *Journal of political Economy*, 117(3):453–503.

Booth, W. (2012). Mexico's crime wave has left about 25,000 missing, government documents show. *The Washington Post*.

Brainerd, E. (2017). The lasting effect of sex ratio imbalance on marriage and family: Evidence from world war ii in russia. *Review of Economics and Statistics*, 99(2):229–242.

Brown, P. H. (2009). Dowry and intrahousehold bargaining evidence from china. *Journal of Human Resources*, 44(1):25–46.

Brown, R. (2018). The mexican drug war and early-life health: The impact of violent crime on birth outcomes. *Demography*, 55(1):319–340.

Brown, R., Montalva, V., Thomas, D., and Velásquez, A. (2019). Impact of violent crime on risk aversion: Evidence from the mexican drug war. *Review of Economics and Statistics*, 101(5):892–904.

Calderón, V., Gáfaro, M., and Ibáñez, A. M. (2011). Forced migration, female labor force participation, and intra-household bargaining: does conflict empower women? *Documento CEDE*, (2011-28).

Calderon-Narvaez, G. (1997). Un cuestionario para simplificar diagnostico del sindrome depresivo. *Revista de Neuro-Psiquiatria*, 60(2):127–135.

Carrillo, P. (2021). The effects of gender norms on work time.

Chakraborty, T. and De, P. K. (2017). Mother's autonomy and child's secondary schooling enrollment in mexico. *Review of Economics of the Household*, 15(3):1037–1053.

Chiappori, P.-A., Fortin, B., and Lacroix, G. (2002). Marriage market, divorce legislation, and household labor supply. *Journal of political Economy*, 110(1):37–72.

Corral, P., Irwin, A., Krishnan, N., and Mahler, D. G. (2020). *Fragility and conflict: On the front lines of the fight against poverty*. World Bank Publications.

Cuartas, J. (2018). Neighborhood crime undermines parenting: Violence in the vicinity of households as a predictor of aggressive discipline. *Child abuse & neglect*, 76:388–399.

Dell, M. (2015). Trafficking networks and the mexican drug war. *American Economic Review*, 105(6):1738–79.

Duflo, E. (2012). Women empowerment and economic development. *Journal of Economic literature*, 50(4):1051–79.

Duflo, E. and Udry, C. R. (2004). Intrahousehold resource allocation in cote d'ivoire: Social norms, separate accounts and consumption choices.

Enamorado, T., López-Calva, L. F., and Rodríguez-Castelán, C. (2014). Crime and growth convergence: Evidence from mexico. *Economics Letters*, 125(1):9–13.

Freije, S., Freije, S., López-Acevedo, G., and Rodríguez-Oreggia, E. (2011). *Effects of the 2008-09 economic crisis on labor markets in Mexico*. World Bank Washington, DC.

Gilligan, M. J., Pasquale, B. J., and Samii, C. (2014). Civil war and social cohesion: Lab-in-the-field evidence from nepal. *American Journal of Political Science*, 58(3):604–619.

Gutierrez-Romero, R. (2016). Estimating the impact of mexican drug cartels and drug-related homicides on crime and perceptions of safety. *Journal of economic geography*, 16(4):941–973.

Heimlich, R. (2011). Crime and drug cartels top concerns in mexico.

Hernandez-de Benito, M. (2022). This is a man's world: Crime and intra-household resource allocation. Technical report, Households in Conflict Network.

Karwowski, M., Kowal, M., Groyecka-Bernard, A., Białek, M., Lebuda, I., Sorokowska, A., and Sorokowski, P. (2020). When in danger, turn right: Does covid-19 threat promote social conservatism and right-wing presidential candidates?

La Mattina, G. (2017). Civil conflict, domestic violence and intra-household bargaining in post-genocide rwanda. *Journal of Development Economics*, 124:168–198.

Lopez-Acevedo, G., Freije, S., Vergara Bahena, M. A., and Cardozo, D. (2020). Changes in female employment in mexico: Demographics, economics, and policies.

Lundberg, S. J., Pollak, R. A., and Wales, T. J. (1997). Do husbands and wives pool their resources? evidence from the united kingdom child benefit. *Journal of Human resources*, pages 463–480.

Majlesi, K. (2016). Labor market opportunities and women's decision making power within households. *Journal of Development Economics*, 119:34–47.

Makino, M. (2019). Marriage, dowry, and women's status in rural punjab, pakistan. *Journal of population economics*, 32(3):769–797.

Miroff, N. and Booth, W. (2012). Mexico's drug war is at a stalemate as calderon's presidency ends. *The Washington Post*.

Mishra, A., Mishra, V., and Parasnis, J. (2021). The asymmetric role of crime in women's and men's labour force participation: Evidence from india. *Journal of Economic Behavior & Organization*, 188:933–961.

Moreno-Brid, J. C. and Padilla-Pérez, R. (2012). The mexican economy and the international financial crisis.

Moya, A. (2018). Violence, psychological trauma, and risk attitudes: Evidence from victims of violence in colombia. *Journal of Development Economics*, 131:15–27.

Nasir, M., Rockmore, M., and Tan, C. M. (2020). Do the lessons from micro-conflict literature transfer to high crime areas?: Examining mexico's war on drugs. *The Journal of Development Studies*, 56(1):26–44.

Qian, N. (2008). Missing women and the price of tea in china: The effect of sex-specific earnings on sex imbalance. *The Quarterly Journal of Economics*, 123(3):1251–1285.

Reggio, I. (2011). The influence of the mother's power on her child's labor in mexico. *Journal of development economics*, 96(1):95–105.

Robles, G., Calderón, G., and Magaloni, B. (2013). The economic consequences of drug trafficking violence in mexico. *Poverty and Governance Series Working Paper, Stanford University*.

Shirk, D. A. (2011). *The drug war in Mexico: confronting a shared threat*. Number 60. Council on Foreign Relations.

Siddique, Z. (2022). Media-reported violence and female labor supply. *Economic Development and Cultural Change*, 70(4):000–000.

Thomas, D. (1990). Intra-household resource allocation: An inferential approach. *Journal of human resources*, pages 635–664.

Thomas, D. (1993). The distribution of income and expenditure within the household. *Annales d'Economie et de Statistique*, pages 109–135.

Tsaneva, M., Rockmore, M., and Albohmood, Z. (2019). The effect of violent crime on female decision-making within the household: evidence from the mexican war on drugs. *Review of Economics of the Household*, 17(2):615–646.

Velásquez, A. (2020). The economic burden of crime evidence from mexico. *Journal of Human Resources*, 55(4):1287–1318.

Voors, M. J., Nillesen, E. E., Verwimp, P., Bulte, E. H., Lensink, R., and Van Soest, D. P. (2012). Violent conflict and behavior: a field experiment in burundi. *American Economic Review*, 102(2):941–64.

Wang, S.-Y. (2014). Property rights and intra-household bargaining. *Journal of Development Economics*, 107:192–201.

Warr, M. and Ellison, C. G. (2000). Rethinking social reactions to crime: Personal and altruistic fear in family households. *American journal of Sociology*, 106(3):551–578.

Young, G. (2014). *Gendering Globalization on the Ground: The Limits of Feminized Work for Mexican Women's Empowerment*. Routledge.

2A Appendix

Table 10: First Stage Regression for Column 4 in Table 2.4

	Homicide rate past 12 months (1)
PAN political party mayor	10.143***
	(0.753)
Household's total number of decisions	-0.281*
	(0.145)
More than two rooms in home	0.684***
	(0.262)
Wife's age square	-0.003
	(0.002)
Husband's age square	0.001
	(0.003)
Household owns their home	-0.872
	(0.826)
Household owns a car	-1.446**
	(0.690)
Household owns a washing machine	-2.043***
	(0.778)
Wife's labor force participation	1.281**
	(0.639)
Husband's labor force participation	-1.802*
	(1.039)
Percentage of female informal workers	209.950***
	(23.323)
Percentage of male informal workers	-190.768***
	(22.362)
Percentage of women economically active	-965.491***
	(57.196)
Percentage of men economically active	-424.259***
	(44.888)
Percentage of men unemployed	-334.114*
	(181.013)
Percentage of women unemployed	762.933***
	(279.480)
Log median income	12.743***
	(4.557)
Gini coefficient (base 2000)	-91.885
	(91.076)
Food poverty percentage	-8.416
	(8.205)
% of Population without secondary schooling	0.000
	(0.000)
Number of children in household	0.283
	(0.338)
Number of children younger than six	0.248
	(0.435)
Number of household members	0.071
	(0.315)
Urban Population Ratio	-51.413***
	(3.211)
Sanderson-Windmeijer multivariate F test of excluded instruments:	
F(1, 2835) = 175.86	
Prob > F = 0.0000	

Note: Robust standard errors shown in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Second Stage Regression for Column 4 in Table 2.4

	Wife's no. of decisions (1)
Homicide rate 12 months	-0.019**
	(0.009)
Household's total number of decisions	0.793***
	(0.020)
More than two rooms in home	-0.076**
	(0.035)
Wife's age square	-0.001**
	(0.000)
Husband's age square	0.000
	(0.000)
Household owns their home	-0.024
	(0.103)
Household owns a car	-0.066
	(0.078)
Household owns a washing machine	0.000
	(0.132)
Wife's labor force participation	0.059
	(0.077)
Husband's labor force participation	-0.155
	(0.133)
Percentage of female informal workers	3.194
	(3.247)
Percentage of male informal workers	-3.435
	(2.945)
Percentage of women economically active	19.319**
	(9.495)
Percentage of men economically active	-9.008
	(7.930)
Percentage of men unemployed	9.727
	(21.958)
Percentage of women unemployed	-5.112
	(31.856)
Log median income	0.044
	(0.572)
Gini coefficient (base 2000)	-41.738***
	(12.961)
Food poverty percentage	-2.466**
	(1.174)
% of Population without secondary schooling	-0.000**
	(0.000)
Number of children in household	0.053
	(0.048)
Number of children younger than six	-0.039
	(0.054)
Number of household members	0.054
	(0.049)
Urban Population Ratio	-0.571
	(0.679)

Note: Robust standard errors shown in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Robustness of the Impacts on Spouses Decision-Making from Homicide Rates

	Total number of decisions	Wife's no. of decisions	Husband's no. of decisions	No of joint decisions	Husband's bargaining power
	(1)	(2)	(3)	(4)	(5)
	Poisson fixed effect regression				
Homicide rate	-0.001	-0.002***	-0.001	-0.004***	0.001
past 3 months	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Homicide rate	-0.001**	-0.001*	-0.000	-0.002**	0.000
past 6 months	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
Homicide rate	-0.000**	-0.000*	-0.000	-0.001**	0.000
past 12 months	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
DV 2005/6 Mean	10.16	8.35	7.76	05	0.48
N	6106	6106	6106	6106	6106

All regressions have as covariates the total number of decisions out of 12 topics and the remaining characteristics in Table 2.1. Robust standard errors shown in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

Information Provision and Court Performance: Experimental Evidence from Chile

Joint with Daniel Chen, Manuel Ramos-Maqueda, and Bernardo Silveira[1]

3.1 Introduction

Many public institutions spend a great deal of time generating, collecting, and storing administrative data. This increasing data availability raises opportunities but also analytical challenges. Recent surveys confirm that a lack of data-management and analytic skills is an important barrier to the use of readily available data for evidence-based decisions in the public sector (OECD, 2014). Therefore, identifying and relieving the constraints that prevent institutions from taking advantage of their data has become a central focus for researchers and policymakers (Savoldelli et al.,

[1]We thank the Department of Institutional Development in the Chilean Judiciary's Administrative Corporartion (CAPJ) for the partnership in this project. We would like to thank in particular Esteban Paiva, Mario Lara, Pablo Cabezas and Sebastián Montero for their coordination. This project received IRB Approval from Toulouse School of Economics on January 18, 2020. The pre-analysis plan is registered in the AEA RCT Registry on https://www.socialscienceregistry.org/trials/5512 and a copy of it can be found in Appendix 3A.1. All errors are our own.

2014).

One particularly promising application of administrative data in the public sector is the use of performance measures to assess the efficiency, effectiveness, procedural satisfaction, and productivity of daily operations, as well as to guide management and policy decisions (Ostrom and Hanson, 2010; Hanson and Ostrom, 2014; Durham and Becker, 2016). The number of public institutions using performance measures is still limited (Hanson and Ostrom, 2014), which can be attributed to competing demands on management's limited attention (Gabaix, 2014) or limited analytical capabilities (Kleiman, 2009). Reducing information frictions could help public officials take better advantage of their measurements, while also increasing the officials' intrinsic motivation by making their own performance more salient.

This paper analyzes information frictions in the context of Chilean courts. It establishes, first, that there are in fact substantial and consequential information frictions among court managers regarding their courts. Specifically, we show that, despite data availability, many managers have inaccurate beliefs about their own court's performance. Second, we use a randomized experiment to show that a simple informational intervention can significantly improve court performance: providing information to court managers in a clear and digestible way causes them to adjust their decisions and improve court efficiency. Finally, we show that more experienced and older court managers are less likely to access their court performance indicators and, thus, hold less accurate baseline beliefs. Luckily, they respond more to the information intervention, helping bridge the gap in the productivity of court managers of different ages and years of experience.

To achieve these goals, we partner with the Department of Institutional Development of the Chilean Judiciary to conduct a randomized control trial in Chilean courts. Through a factorial design, we analyze two main interventions: an email promotion of the online court management platform, and a redesign of that platform. We evaluate whether each of these interventions—or the combination of both—increases the usage and digestion of information on the online platform and enhances court productivity.

The Chilean Judiciary is an ideal setting for our analysis. In the early 2000s,

the Chilean Judiciary created a new position in family courts, the court manager, which is in charge of the administrative functioning of the court. Court managers are responsible for planning, organizing, and monitoring the performance of the court, and supervise between 40 and 60 court staff each. Based on a survey that we conducted in 2020, 70% of the court managers in the country agree or strongly agree that tracking court indicators is one of the most important tasks of their job (see subsection 3A.2 for methodology and results of the survey). Since 2018, court managers have had access to court measurements through an online platform, named Quantum. While 71% of court managers agree or strongly agree that the information in the platform is useful for their work, the usage of the online platform is limited: From July 2018 to August 2019, 20% of court managers never logged in, and there was an average of 20 logins per court manager that year. Court managers are not familiar with their own court indicators; when asked to recall their court's case clearance rate in the preceding two months, 40% of the managers were off by more than 25% from the true value.

In our first intervention, we randomly promote Quantum by sending court managers promotional emails; we refer to this as the *email promotion* intervention. In the second intervention, which we call the *feedback intervention*, we redesign Quantum and randomly assign court managers to different versions of the platform. In one arm of the feedback intervention, we provide court managers with a new dashboard on the homepage of Quantum that simplifies and summarizes the main performance statistics (*simplified feedback*). In another arm, we supplement the simplified dashboard with a pop-up window that compares the managers' best and worst statistics relative to another family court of similar size (*social comparison feedback*). The social comparison captures mechanisms that draw on social norms, comparisons, and pro-social behavior effects. Finally, the court managers in the control group of the feedback intervention continue seeing the original version of Quantum, which presents statistics in a disaggregated table.

How managers respond to the three treatments we consider—i.e., email promotion, simplified feedback, and the comparison feedback—is a non-trivial empirical question. Although the interventions facilitate access to performance measures and reduce information frictions, it is possible that managers perceive them as an accountability

mechanism that enables their superiors to control their efforts and compare them to other court managers. This, in turn, could deteriorate the sense of autonomy of the court managers and reduce effort (Humphrey et al., 2007; Rasul and Rogger, 2018). Similarly, the dashboards in the feedback intervention—notably those in the social comparison treatment—can be interpreted as providing unsolicited feedback to the court managers. According to West et al. (2018), this type of feedback may increase physiological and psychological expressions of anxiety, which, in our setting, could lead to worse court performance.

The interventions began in February 2021 and lasted until June 2021. We find a substantial effect of the treatments on court performance. The combination of email promotion and feedback increases the timely motion resolution rate (+0.2, or +0.5 sd) and hearing programming (+0.7, or +1.3 sd) but it also decreases the realized hearings (-1.3, or -1.0 sd) among the treated. The simplified feedback increases the case clearance by one standard deviation for the treated. The magnitude of the effect of the simple feedback is statistically larger than social comparison feedback for hearing programming and case clearance and less negative for the realized hearings. Our results are in accordance with the findings by West et al. (2018) that unsolicited feedback to employees can lead to worse performance.

Given that the technology to easily access court performance measures is relatively new for the Chilean court managers, court managers with different years of experience and age are not equally proficient in the use of the platform. We find that managers who are more experienced and older log in fewer times into the Quantum platform. Additionally, courts managed by more experienced and older court managers tend to underperform those managed by less experienced and younger managers. Interestingly, the treatment effects are not homogeneous across court managers. Those who are more experienced in their position respond more to the treatments. As the treatments improve the productivity of most indicators, the interventions help bridge the gap in productivity between more and less experienced court managers.

This paper makes several contributions and builds on at least two strands of literature. First, our study contributes to the growing literature on how the inner workings

of government administration can influence the quality of public service delivery. This paper studies low-cost and easily implementable changes using existing technology to the work of court administrators and its impact on public provision. This paper complements recent work testing more specialized approaches for using technology to improve governance, such as monitoring worker attendance with time clocks (Banerjee et al., 2008; Dhaliwal and Hanna, 2017), phone call monitoring (Muralidharan et al., 2021), or custom mobile phone applications (Callen et al., 2020; Dodge et al., 2021). Relative to these specialized approaches which introduce new technology, re-imagining existing technology to improve management has the advantages of (i) low fixed costs and almost no recurring costs, (ii) the flexibility to scale across places, programs and outcomes, and (iii) scope to adapt quickly as challenges and circumstances change. Particularly, improving existing technology avoids the additional learning costs that come from adopting new technology that tends to unevenly impact more senior workers.

Second, this paper contributes to the literature on the importance of bureaucratic information acquisition costs and how to reduce information acquisition frictions in the management of public institutions. After the investment of many governments in recording high-frequency data, public institutions are now concentrating on technological and operational reforms to process this high-frequency data into evidence-based decisions that management can take (Savoldelli et al., 2014). For example, in a study of payroll managers, Dodge et al. (2021) find that lowering information-access costs through mobile phone applications reduces payment processing time. Instead of mobile phone usage, our intervention creates clearer and more digestible information through dashboards. Increasingly, public officials employ dashboards to improve their decision-making in areas ranging from health (Callen et al., 2020; Whidden et al., 2018), civil service (Mattsson, 2021), education (Dizon-Ross, 2019) to tax collection (Pomeranz, 2015). Ours is the first study to use informational dashboards to evaluate the improved monitoring of public employees in the judicial context. Courts are hard to reform and have generally shielded away from evaluations designed to show the impact of programs on the system and their users (Engel, 2013; Greiner and Matthews, 2016). Our paper analyzes the effects on court productivity of a dashboard designed

to reduce information frictions using existing technology.

3.2 Court Managers and Quantum Platform

The court managers' primary role is to facilitate the administrative functioning of the court. The position arose in the early 2000s from the need to separate the jurisdictional responsibilities of judges from the administrative management of the court. Thus, court managers are responsible for planning, organizing, directing and monitoring the work of the court, supervising the performance of each of the units of the court, and optimizing the performance of court staff (Chilean Judiciary, 2020). Each court manager supervises over 40 to 60 employees. They are also responsible for providing constant feedback on the areas for improvement and achievements of court staff. For example, court managers may decide to alter the court's agenda, hearing schedule, number of staff or redistribute tasks in response to inflows in the number of incoming cases (see Table 9).

In July 2018, the Department of Institutional Development of the Chilean Judiciary created the online Quantum platform. Quantum provides court managers and other court staff with access to comprehensive information on performance metrics at the court level, such as the number of cases filed, the case clearance rate, the average duration of cases, and the percentage of realized hearings. Quantum presents the information for all courts, thus allowing users to compare the performance of different courts. The indicators are aggregated at monthly and yearly levels.

Since Quantum's launch in 2018, take-up has been limited: 20% of court managers never logged in and there is an average of 20 logins per court manager in one year and two months.[2] The platform is technologically well-developed and rich in information, yet it is unclear whether it has any impact on the management of the court.

One potential explanation for Quantum's low usage is that part of its target audience—especially those without a quantitative background—may view the information provided by the platform as overwhelming and not sufficiently clear. Upon

[2]Statistics from July of 2018 to August of 2019

login, users are taken to an initial homepage that provides a packed table. It might not be immediately clear to a user how to access disaggregated data or to look for information of a specific kind. In other words, it is easy for users to get lost in the platform. In the next section, we describe how our intervention attempts to make the information on Quantum easier to digest.

3.3 Intervention and Data

Our intervention focuses on 49 family courts in Chile—each of them supervised by one court manager. As of 2019, these courts employed 1,697 court staff, including 254 technical advisers, 1,106 administrative officers, 85 heads of unit, and 252 judges.[3] There were 698,971 family cases in 2021. During the period of our intervention, we estimate that the 49 courts resolved approximately 235,000 cases.[4] The most common case types in 2021 were breach of rights (21.1%), alimony (16.6%), intra-family violence (16.4%), and child custody (8.9%).

We randomize the 49 family courts into one of six treatment arms: (C) control; (T1) email; (T2) simplified feedback; (T3) email and simplified feedback; (T4) social comparison feedback; and (T5) email and social comparison feedback. We stratify on the number of incoming cases in 2019—a proxy for the size of the court. The randomization selected eight courts in control, eight in Treatment 1, nine in Treatment 2, eight in Treatment 3, nine in Treatment 4, and eight in Treatment 5. The factorial design implies that the same courts are randomized into the email/no email treatment and the feedback/no feedback treatment. Thus, regarding the email intervention, we have 24 courts in the email promotion treatment arm and 25 in the treatment arm with no email promotion. Similarly, for the feedback intervention, we have 16 courts in the treatment arm with no feedback and 33 in the treatment arm with either simplified

[3]In total, there are 60 family courts in Chile. However, in 2020, 10 family courts were chosen by the Department of Institutional Development of the Chilean Judiciary to design the simplified homepage of the Quantum platform, and were therefore excluded from the experiment. We also exclude another court from the analysis because we received no performance metrics on it.

[4]The yearly data is publicly available at: https://numeros.pjud.cl/Competencias/Familia. To estimate the total number of judicial cases affected by our intervention, we take the proportion of total cases that took place in 5 months and in 49 courts out of the total of 60 courts.

feedback or social comparison feedback.

The intervention was launched on January 26, 2021, and lasted 5 months, until the end of June. With our sample size and multiple month observations before and after treatment the study is able to detect "medium" size effects between 0.4 and 0.6 standard deviations (see Appendix 3A.4). All 49 courts, including those in the control group, received an initial email on January 26 of 2021. This initial email informed the courts about Quantum, how to access it, and how to retrieve their password, if forgotten. Courts in Treatments 2, 3, 4, and 5 (feedback treatment arm) were also informed about changes in their homepages. Courts in the control group (C) had no change to their Quantum homepage nor were provided with any Quantum promotion after the initial January 26 email.

The courts randomized into Treatments 1, 3, and 5 (email promotion treatment arm) received three different emails in three different months promoting access to Quantum.[5] The emails began by announcing and congratulating the top two or three managers that logged into Quantum the greatest number of times in the previous month. Furthermore, the emails were personalized using the information on recent login activity by the recipient manager. If the manager did not login at all in the previous month, the email would contain the following text: "We noticed you did not access Quantum during [month] of 2021. You are losing the opportunity to follow your indicators and evaluate the productivity of your court and compare it to other courts". For those that had at least one login in the previous month, the email read: "We trust you will all continue using Quantum to follow your indicators, evaluate the productivity of your courts and compare it with other courts." Additionally, all emails included a small paragraph stating that most users judge Quantum positively for its easy usage, clear information, trusted indicators, and relevance to one's work. The paragraph also mentioned that those using Quantum have a larger clearance rate for cases older than two years. Finally, all the emails included a Quantum link and a large linked button to Quantum to increase the accessibility and salience of Quantum. The distinct email versions sent are presented in Appendix 3A.3.

[5]The three rounds of emails were on February 22, March 22, and May 25.

The courts randomized into Treatments 2, 3, 4, and 5 (feedback treatment arm) had their homepage in Quantum updated to a simplified version. This homepage presents the key statistics at the court level. Examples of the old and new homepage versions are shown in Figure 1 and Figure 2 in Appendix 3A.2. The old version includes a condensed main table showcasing the number of incoming cases, terminated cases, hearings, protection measures in favor of children, legal writings, resolutions, proceedings, notifications, and people attended by customer service. These statistics are shown at the monthly level and in cumulative terms for the current year. The changes made to the new homepage, which we refer to as simplified feedback, include the streamlining of information and the addition of statistics in percentage or rate terms.

For courts in Treatments 4 and 5, a pop-up window appears on top of the simplified homepage upon accessing Quantum (see Figure 3 in Appendix 3A.2). This window highlights three performance indicators: one of them is the indicator in which the court performed best and the two other are the indicators in which the court performed worst, relative to a peer family court.[6] The pop-up window thus stresses the court's relative strengths and weaknesses, which lean into social comparison motivation. Note that the pop-up window always appears in addition to the simplified feedback homepage. In our analysis, we refer to the social comparison feedback treatment as a combination of the pop-up window and the simplified feedback.

Table 10 presents a summary of the different category labels discussed in this section, which we continue to use in the remainder of the paper.

Table 3.1: Treatment Category Labels

	C	T1	T2	T3	T4	T5
Email		✓		✓		✓
Simplified feedback			✓	✓		
Social comparison feedback					✓	✓
Feedback			✓	✓	✓	✓
Email x Feedback				✓		✓

(C) control; (T1) email; (T2) simplified feedback; (T3) email and simplified feedback; (T4) social comparison feedback; and (T5) email and social comparison feedback.

[6]Peer courts are established by the Chilean Judiciary as being similar courts given their competence (in this case family) and their "size" which is proxied by the number of incoming cases in a year.

3.3.1 Data

We observe five monthly court indicators, starting four months prior to the start of the intervention and up until the fifth and final month of the intervention. The first indicator is the *case clearance rate*, defined as the number of cases resolved over the number of incoming cases in a month. Second, the *timely motion resolution rate* is computed as the number of motion signed within three working days as a ratio of total signed motion resolutions in a month.[7] Third, the *rate of realized hearings* is the number of hearings held after the first scheduling divided by the total scheduled hearings. Fourth, the *timely case resolution rate* is computed as the inventory of cases with entry date shorter than two years as a ratio of the total cases in inventory. Finally, we define the *rate of timely hearing programming* as the number of hearings scheduled within 70 days as a ratio of total number of scheduled hearings. These five indicators allow measuring changes across different case components and across short and long term horizons (e.g., timely case resolution vs. case clearance).

Other pre-treatment measures about the courts' characteristics and their efficiency include indicators from 2019: number of incoming cases, number of hearings, case clearance, inventory older than one year, and inventory older than two years. We have information on the managers' age, sex, and tenure. About two thirds of court managers are men. They are on average 52 years old and have been in their position for on average 12 years.

Furthermore, we have information on logins into Quantum one month before the intervention and during the five months in which the intervention took place. The login information gives the exact date and time of login for each manager. We have neither information on the duration of the access nor on which pages in the platform were visited. In the pre-treatment month of January 2020, 33 percent of court managers logged into Quantum at least once.

Tables 3.2 and 3.3 summarize our key variables at baseline for the treatment arms of feedback and email promotion, respectively. The first and second columns in each

[7] A motion is a written request or proposal to the court to obtain an asked-for order, ruling, or direction. There are a variety of motions, and it has become standard practice to file certain kinds of motions with the court based on the type of case.

Table 3.2: Court Characteristics for No-Feedback and Feedback at Baseline

	No Feedback (1)	Feedback (2)	Difference (1-2) (3)	SE (4)
Panel A. Manager's Characteristics				
Age	52.06	52.00	0.06	(1.90)
Men	0.69	0.67	0.02	(0.15)
Years in position†	12.31	12.30	0.01	(1.61)
January Quantum logger	0.31	0.33	-0.02	(0.15)
Number of logins in January	0.94	0.85	0.09	(0.41)
Panel B. Court Indicators (2019)				
2019 incoming cases	8799.19	8938.61	-139.42	(1668.42)
2019 hearings	5320.13	5275.33	44.79	(960.60)
2019 case clearance	75.54	74.30	1.24	(2.55)
2019 inventory older than 1 year	0.32	0.70	-0.38	(0.28)
2019 inventory older than 2 years	0.06	0.17	-0.11	(0.11)
Observations	16	33	49	49
Panel C. Court Indicators (0ct-Jan)				
Case rate	70.39	70.48	-0.09	(1.75)
Timely motion resolution rate	96.13	97.53	-1.41	(0.41)
Realized hearings rate	69.16	60.01	9.15	(2.38)
Timely case resolution rate	92.16	91.62	0.54	(1.90)
Timely hearing programming rate	56.05	63.14	-7.09	(4.37)
Observations	64	132	196	196

The no-feedback group includes those in the control group (C) and those receiving the email promotion (T1).
†The years in the position of court manager has an upper bound of 16 years when the position was first created. 26 out of the 49 court managers have been in the position since its inception.

Table 3.3: Court Characteristics for No-Email and Email at Baseline

	No Email (1)	Email (2)	Difference (1-2) (3)	SE (4)
Panel A. Manager's Characteristics				
Age	51.35	52.78	-1.44	(1.77)
Men	0.62	0.74	-0.12	(0.14)
Years in position†	11.85	12.83	-0.98	(1.51)
January Quantum logger	0.35	0.30	0.04	(0.14)
Number of logins in January	0.85	0.91	-0.07	(0.39)
Panel B. Annual Court Indicators (2019)				
2019 incoming cases	9074.15	8688.39	385.76	(1566.84)
2019 hearings	5347.85	5224.52	123.32	(902.47)
2019 case clearance	73.39	76.18	-2.79	(2.37)
2019 inventory older than 1 year	0.67	0.46	0.21	(0.26)
2019 inventory older than 2 years	0.18	0.08	0.10	(0.10)
Observations	26	23	49	49
Panel C. Monthly Court Indicators (0ct-Jan)				
Case rate	70.27	70.66	-0.39	(1.65)
Timely motion resolution rate	97.28	96.84	0.44	(0.39)
Realized hearings rate	62.63	63.40	-0.77	(2.32)
Timely case resolution	93.00	90.43	2.57	(1.76)
Timely hearing programming rate	58.72	63.20	-4.47	(4.12)
Observations	104	92	196	196

The no-email group includes those in the control group (C) and those receiving the feedback promotion (T2 and T4). †The years in the position of court manager has an upper bound of 16 years when the position was first created. 26 out of the 49 court managers have been in the position since its inception.

table present the mean of the control group and the treatment arm. The third and fourth columns give the difference between the treatment and control groups and the standard errors. We consider five manager characteristics, nine annual court indicators for 2019, and five monthly pre-treatment indicators from October 2020 to January 2021.

Before the treatment, the groups were similar on most variables considered. For the feedback treatment arm, out of the 19 variable differences reported in the third column in Table 3.2, there are two instances where the estimated differences are statistically significant. For monthly indicators, the rate of resolutions is statistically higher in the treatment group and the rate of realized hearings is statistically lower in the treatment group. There is no statistically significant difference between those that received the email promotion and those who did not, as seen in Table 3.3. Overall, we conclude that the random assignment of courts to treatment and control groups was largely successful; nevertheless, in our analyses below, we include as covariates the court manager's characteristics and pre-treatment monthly indicators to control for any potential pre-treatment differences that, although not statistically significant, are not negligible in terms of magnitude and likely to be imprecisely estimated given the small sample size.

3.3.2 Survey Data: Manager's Indicator Knowledge

We developed a survey to measure court manager's perceptions on indicators, managerial practices, and the Quantum platform. Particularly, the survey asked about knowledge of their own court indicators in absolute terms and relative to others, their opinion on managerial practices, their actions as court managers, and, for certain randomly selected managers, their perception and usage of Quantum.

The survey was sent to the work email of all 346 Chilean court managers[8] on January of 2020, including the family courts that our intervention focuses on. Note that this survey was sent one year before the intervention started. We received 121

[8]There are more courts than court managers as the civil courts have not yet been reformed, leaving some courts without managers.

surveys back from distinct court managers, of which 21 are within our intervention.[9]

Among survey respondents, 70% agree or strongly agree that tracking court indicators is one of the most important tasks in the job of court managers[10]. Furthermore, 71% of court managers agree or strongly agree that the information in the Quantum platform is useful for their work. Yet, it is not evident that court managers know their court's indicators well. When asked to recall the two-months previous case clearance, 40% of court managers were off by more than 25% from the true value. The average age and gender composition of respondents and non respondents were similar; yet, respondents had higher clearance rate than those that did not respond. One could expect that those that perform better are also more likely to better guess their indicators. Thus, we see the figures on the knowledge gap provided by the survey as a lower bound for the the true gap.

3.4 Empirical Model

This section explores the methodology to measure the impact of the email promotions and feedback treatments on court efficiency. Denote by n one of the five productivity indicators discussed in Section 3.3.1, and consider the following specification:

$$Y_{nit} = \alpha_{n0} + \alpha_{n1} Treatment_i + \alpha_{n2} \mathbf{X}_i + \alpha_{n3} \mathbf{Y}_{nit_0} + \alpha_{n4} \gamma_t + \xi_{nit}, \tag{3.1}$$

where Y_{nit} denotes the value of indicator n for court i in month $t \in$\{February, March, April, May, June\}; $Treatment$ may consist of different combinations of the six treatment arms; $\mathbf{X_i}$ is a vector of characteristics of court i, which includes the court's manager's sex, age, and tenure, and the stratified variable of a dummy for large courts computed using the 2019 number of incoming cases; $\mathbf{Y}_{nit_0}$ is a vector of the values of the indicator n in the pre-intervention months of October, November, December, and

[9]This constitutes a 35% response rate for all courts and 43 % rate within the intervention courts. The relatively small response rate limits the usage of the information from the survey to investigate heterogeneous effects of the intervention. Still, it gives us the opportunity to assess manager's knowledge on their own indicators prior to the intervention.

[10]Among the remaining percentages, 1% strongly disagrees, 4.5% disagree and 24% are neutral about the statement

January; and γ_t is a month fixed effect.

For the feedback treatments, α_{n1} captures the average treatment effects. Note that not every court manager who was assigned to the feedback treatments received them—as being exposed to the feedback treatments requires the manager to log into Quantum. We are able to estimate the local average treatment effect (LATE) of the feedback treatments—which, in our setting, is equivalent to the treatment on the treated or the compliers, as no one in the no-feedback control group was exposed to feedback. The local average treatment effect consists of the impact of the treatment among managers who: (i) were in the feedback treatment group; and (ii) logged into Quantum at least once in the month. Thus, the LATE of the feedback treatment uses as *effective treatment* logins into the platform and as instrument the treatment assignment. This estimator relies on two key assumptions: monotonicity—that is, the assignment to treatment does not make one less likely to be treated; and the exclusion restriction—that is, individuals respond to the treatment itself rather than to treatment assignment.

In all our regressions, we standardize the dependent variable Y_{nit} using October to June's observations of the control treatment (C). As a result, all indicators have a standard deviation of one.

3.5 Results

Tables 3.4 and 3.5 present the local average treatment effects (LATE) of the distinct treatments for the five court indicators. Both tables show three regressions, using distinct treatment groupings: email promotion (column 1); simplified feedback and social comparison feedback (column 2); and email, feedback, and email and feedback interaction (column 3). The respective reference control groups for each regression are shown at the bottom of each section. To compare the average treatment effects (ATE) and local average treatment effects (LATE), please refer to Table 11 in Appendix 3A.6.

Table 3.4 shows that the impact of the email promotion on case clearance cannot be distinguished from zero. In contrast, the simplified feedback has a large and statis-

Table 3.4: The Local Average Treatment Effects for Email and Feedback

Case Clearance	(1)	(2)	(3)
Email promotion	0.222		0.0270
	(0.284)		(0.607)
Simplified feedback		0.957**	
		(0.475)	
Social comparison feedback		0.500	
		(0.656)	
Feedback			0.807
			(0.591)
Email * feedback			-0.0675
			(0.805)
Timely Motion Resolution	(1)	(2)	(3)
Email promotion	0.208***		0.0909
	(0.0711)		(0.186)
Simplified feedback		0.360***	
		(0.112)	
Social comparison feedback		0.587***	
		(0.149)	
Feedback			0.492***
			(0.167)
Email * feedback			-0.0743
			(0.255)
Realized Hearings	(1)	(2)	(3)
Email promotion	-0.794***		-1.361***
	(0.217)		(0.401)
Simplified feedback		-0.397	
		(0.260)	
Social comparison feedback		-0.627*	
		(0.374)	
Feedback			-1.011***
			(0.308)
Email * feedback			1.148**
			(0.558)
Control Group	No email promotion	No feedback	No email and no feedback
N	245	245	245

All regression have as covariates the strata variable, three manager characteristics, four pre-treatment lagged values of the dependent variable, pre-treatment login dummy, and month fixed effects. Standard errors are clustered at the court level and bootstraped. Average treatment effects and the comparison to the local average treatments are found in Table 11

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

tically significant positive impact on the case clearance rate—increasing the average case clearance rate by one standard deviation for those that were treated, relative to courts that did not receive feedback. The magnitude of the impact of social comparison feedback is lower and not statistically significant, but the impacts of the simplified feedback and social comparison feedback are not statistically different from each other. The feedback intervention (both simplified and social comparison) increases the case clearance by 0.8 standard deviations compared to the control group, although the magnitude is not significant.

For the timely motion resolution, almost all treatment groupings had a statistically significant positive impact. Particularly, the impact of the social comparison feedback seems to be the highest, followed by the simplified feedback and finally by the email promotion. The social comparison feedback's average treatment on the treated increases motion resolution by 0.6 standard deviations. Considering both types of feedback together, the impact was an increase in motion resolution by 0.5 standard deviations. In the third regression, the effect of email promotion is positive but loses its statistical significance. Furthermore, the impact of the interaction of email promotion and feedback is of similar magnitude but has an opposite sign to that of email promotion. This suggests that the email and feedback treatments are not additive in their impact.

In contrast to the previous findings, the last indicator in Table 3.4 shows negative impacts of the treatments on realized hearings. The simplified feedback and social comparison feedback both have negative coefficients of large magnitude, although only the latter is significant at conventional levels; there are no statistically significant differences between the two effects. When considering social comparison and simplified feedback together, the feedback caused the rate of realized hearings to drop by one standard deviation for those treated, and the email promotion caused a drop of 1.4 standard deviations. Once again, the interaction between email and feedback shows an opposite sign to their individual impact and a similar magnitude to one of them. Thus, the rate of realized hearings or the number of hearings held after the first scheduling over total scheduled hearings, lowers with the feedback and email promotion treatments.

Table 3.5: The Local Average Treatment Effects for Email and Feedback

Timely Case Resolution	(1)	(2)	(3)
Email promotion	0.118*		0.0266
	(0.0695)		(0.188)
Simplified feedback		0.0880	
		(0.0927)	
Social comparison feedback		-0.0891	
		(0.157)	
Feedback			-0.0607
			(0.109)
Email * feedback			0.153
			(0.245)
Timely Hearing Programming	(1)	(2)	(3)
Email promotion	-0.127		0.737**
	(0.133)		(0.310)
Simplified feedback		0.563**	
		(0.236)	
Social comparison feedback		-0.0212	
		(0.320)	
Feedback			1.330***
			(0.353)
Email * feedback			-1.656***
			(0.441)
Control Group	No email promotion	No feedback	No email and no feedback
N	245	245	245

All regression have as covariates the strata variable, three manager characteristics, four pre-treatment lagged values of the dependent variable, pre-treatment login dummy, and month fixed effects. Standard errors are clustered at the court level and bootstraped. Average treatment effects and the comparison to the local average treatments are found in Table 11.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We can discard the possibility of the negative impact been driven by a lack of exposure in the new dashboard as the new simplified feedback homepage displays the number of realized hearings and its comparison to the previous month in a prominent area and, furthermore, the impact is also seen for those experiencing only the email promotion treatment. Another possibility for a negative impact on realized hearing could be that managers could divert staff's effort away from hearings to other activities such as motion resolution or other cases that may not require hearings. Yet, this story is unlikely to be at play give that timely hearing programming increases with the treatments (see below), meaning more hearings are scheduled within 70 days. Furthermore, the negative impact of the treatments on realized hearings does not seem to be explained by the increase in timely hearing programming number or more earlier scheduling of hearings.[11] Other possibilities for the negative impact of the treatments on realized hearings could reflect compositional effects about the type of cases scheduled for hearings or it could reflect a larger flexibility in scheduling hearings. With greater flexibility for scheduling hearings, litigants may request changes in scheduled hearings to better met their schedules. This would reduce the number of hearings held after the first schedule but may represent higher a quality of the judicial system and end up speeding up the process of cases.

As for the the rate of inventory with an entry date earlier than two years, there is little impact of the treatments (see Table 3.5). Given the longer-term horizon of this indicator and the relatively short-term nature of the intervention we consider, this finding may come as no surprise.

Finally, for timely scheduling of hearing programming, the treatments have a positive impact. The email promotion increased hearing programming by 0.7 standard deviations, while the feedback increased it by 1.3 standard deviations. The impact of the simplified feedback is statistically higher than that of social comparison feedback. Once again, we see the interaction term of email and feedback having an opposite sign but of similar magnitude to the email or feedback individual treatments. Given that this result appears in many of our specifications, we conclude that the impacts of email

[11] Hearing programming of different months as a regressor for realized hearing gives us a small, non statistically significant coefficient.

and feedback are not additive.

Thus, we find that when controlling for manager characteristic and pre-treatment behavior both treatment arms of email promotions and feedback increase the performance for timely motion resolution and hearing programming, while they decrease it for the rate of realized hearings. The largest impact on the motion resolution comes from the social comparison feedback, while for hearing programming most of the impact comes from the simplified feedback. The negative impact on first scheduled realized hearings does not seem to suggest that managers divert effort away from hearings as timely hearing programming improved. Also the negative impact on this particular indicator does not necessarily point to a worsening of the productivity of the judicial system since case clearance, which is largely consider the main tribunal indicator that encompasses all other indicators, improved with the simplified feedback and also had positive magnitudes for the other treatments.

Furthermore, the evidence suggests that the impacts of the email and feedback on the indicators, which always go in the same direction, are not additive; thus, there is no extra gain for courts that received both email and feedback treatments, relative to courts that receive just one of these treatments.

As a complement to the findings reported in this section, in Appendix 3A.6 we assess the impact of the email promotion and feedback interventions on the login behavior of court managers. Our results from that analysis indicate that the email causes a substantive increase in the usage of Quantum. Specifically, three rounds of email promotions increase the number of logins per month by 1.5 logins—a 127 percent increase compared to the control group. As for the feedback treatments, the evidence is suggestive that they affect the number of logins positively. We refer the interested reader to Appendix 3A.6 for further details.

3.5.1 Heterogeneity

In this section, we explore the heterogeneity of the treatment effects by the number of years of experience and age of the court managers. Although envisioned in the Pre-Analysis Plan (see Appendix 3A.1), we cannot explore the changes in the knowledge

of Quantum indicators as a post-treatment survey was not conducted; yet, given the negative correlation found between age and years of experience and the Quantum platform usage, we instead explore using these variables for heterogeneity treatment effects.

Given that Quantum was only launched in 2018, it is likely that more experienced court managers developed management strategies and habits that do not depend largely on accessing court statistics. As one may expect, younger and less experienced court managers log more into to Quantum than their older and more experienced counterparts, as seen in Table 14 in Appendix 3A.6.

It is not ex ante clear whether age and experience would increase or decrease the impact of the information treatments. On one hand, managers with fewer years of experience or younger are more likely to be already aware of their court statistics from their previous Quantum usage and thus may learn less from the new information of Quantum intervention; on the other hand, these managers may be more willing to adapt their management, given the new information provided by the treatments. The opposite can be true for older and more experienced court managers: they may have larger information gains from the information in Quantum, but may be less willing to explore new alternatives to their management strategies, as their management habits may have solidified. The dominant effect for each group will depend on the trade-off between new information and the responsiveness or ability to adjust management practices based on the new information.

Table 3.6 shows the heterogeneous effects of the treatments by experience for the five distinct indicators. Case clearance has noisy estimates (column 1) but seems to follow the pattern of timely motion resolution (column 2). At a one percent significance level, the impacts of simplified feedback and social comparison feedback on motion resolution are higher for more experienced court managers. With each extra year of experience, receiving the simplified and social comparison feedback increases the timely motion resolution rate by 0.09 and 0.1 standard deviations, respectively. For example, among court managers with 16 years of experience—the median value in our sample—the simplified and social comparison feedback treatments lead to in-

Table 3.6: Heterogeneous Treatment Effects for Feedback by Experience

	Case Clearance	Timely Motion Resolution	Realized Hearings
	(1)	(2)	(3)
Simplified feedback	-0.824	-0.634**	-2.227**
	(1.182)	(0.316)	(0.987)
Simplified feedback * Experience	0.152	0.086***	0.142*
	(0.104)	(0.028)	(0.079)
Social comparison feedback	-0.435	-0.246	-1.137***
	(0.674)	(0.162)	(0.423)
Social comparison feedback * Experience	0.104	0.101***	0.033
	(0.093)	(0.038)	(0.054)
Experience	-0.001	-0.022***	-0.020
	(0.024)	(0.006)	(0.014)

	Timely Case Resolution	Hearing Programming
	(4)	(5)
Simplified feedback	-0.744**	1.494***
	(0.302)	(0.399)
Simplified feedback * Experience	0.072**	-0.086**
	(0.029)	(0.039)
Social comparison feedback	-0.355*	0.429
	(0.196)	(0.273)
Social comparison feedback * Experience	0.034	-0.050
	(0.036)	(0.041)
Experience	-0.009	0.028***
	(0.006)	(0.006)
N	245	245

All regression have as covariates the strata variable, three manager characteristics, four pre-treatment lagged values of the dependent variable, pre-treatment login dummy, and month fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

creases in the motion resolution rate by 0.7 sd and 1.4 sd, respectively. For managers with 12 years of the experience—the average value—the effects are 0.4 sd and 1.0 sd, respectively. Furthermore, those that are more experienced tend to underperform in the motion resolution rate; thus, the feedback treatments are able to shrink the gap that exists between newer and more experienced court managers. Similarly, although only at a ten percent level, higher years of experience increase the positive impact of the simplified feedback on realized hearings, counterbalancing the generally negative impact of simplified feedback on the indicator.

We see once again that years of experience positively impact the effects of simplified feedback on timely case resolution. Table 3.6 suggests that higher experience is likely to increase the positive impact of the feedback on the court's productivity.

Table 3.7: Heterogeneous Treatment Effects by Experience and Age

	Case Clearance	Timely Motion Resolution	Realized Hearings	Timely Case Resolution	Timely Hearing Programming
	(1)	(2)	(3)	(4)	(5)
Feedback	-0.467	-0.351***	-2.194***	-0.324**	0.720***
	(0.645)	(0.128)	(0.427)	(0.131)	(0.213)
Feedback*Experienced	1.311	0.740***	1.933***	0.490**	-0.843***
	(0.910)	(0.179)	(0.646)	(0.208)	(0.311)
Feedback*Experienced*Older	1.569	1.207	0.144	0.078	1.211*
	(2.206)	(0.785)	(1.423)	(0.403)	(0.699)
Experienced	0.044	-0.323***	-0.484**	-0.134**	0.425***
	(0.353)	(0.059)	(0.220)	(0.066)	(0.090)
Older	-0.292	-0.062	-0.103	0.064	-0.258***
	(0.273)	(0.101)	(0.276)	(0.072)	(0.100)
N	245	245	245	245	245

All regression have as covariates the strata variable, three manager characteristics, four pre-treatment lagged values of the dependent variable, pre-treatment login dummy, and month fixed effects. Experienced is equal to one for court managers with more than six years of experience in their position. Older is a dummy if the court manager is 53 or older.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Given that experience and age are positively correlated, we further explore if the interaction of years of experience and age has a differential impact; yet, given the number of observations and the introduction of a third interaction, we interpret the results as suggestive evidence on these heterogeneous effects. Table 3.7 shows the impact of the feedback treatment arm plus two interaction terms: (1) feedback and being experienced and (2) feedback, experienced, and older, where experienced is a

dummy for having more than six years of experience in the current position. Less than or equal to six years of experience is the lowest quartile. Given that there are no court managers with experiences from 7 to 9 years, six years gives a good cutting point for considering less experienced court managers. Older is a dummy for being older than the mean age of 52.

Once more, we find suggestive evidence that receiving the feedback treatment while being more experienced positively impacts all indicators except hearing programming. Furthermore, the results suggest that among those who are more experienced, being older has an extra positive impact when receiving the feedback treatment; although the estimates for the triple interaction have large magnitudes, only that associated with timely hearing programming is statistically significant.

3.6 Conclusion

Many court systems adopted measures to keep track of their performance indicators, but this information is seldom used to assist court management. In this paper, we partnered with the Department of Institutional Development of the Chilean Judiciary to investigate the impact of providing performance information to court managers.

We first found through survey responses that information frictions exist. Many court managers are unable to recall their court's indicators. We investigate the possible reduction of information frictions through a factorial randomized control trial through an open-source online platform that shows a repertoire of court statistics. First, we incentivize the use of the platform through email promotions. Second, within this platform, we simplify the main homepage feedback containing the information on court statistics. Third, we explore providing performance feedback in comparison to a peer court.

We find that both email promotions of the court statistics platform and a simplified feedback about court statistics to court managers improved court performance. As a result of these interventions, we find an increase in timely motion resolution and hearing programming rates, although we also find a reduction in the rate of realized

hearings. The largest impact on the motion resolution rate comes from the social comparison feedback, while the effect of simplified feedback is larger for hearing programming. Simplified feedback significantly increased case clearance as well, which is a key metric of court efficiency.

Finally, we find the effects of the treatments are larger for older and more experienced court managers. Future research can explore the benefits of using AI trained on court managers to present best practices and reduce information frictions within and between court managers.

Bibliography

Banerjee, A. V., Duflo, E., and Glennerster, R. (2008). Putting a band-aid on a corpse: incentives for nurses in the indian public health care system. *Journal of the European Economic Association*, 6(2-3):487–500.

Callen, M., Gulzar, S., Hasanain, S. A., Khan, M. Y., and Rezaee, A. B. (2020). Data and policy decisions: Experimental evidence from pakistan. *NBER Working Paper*, 27678.

Chilean Judiciary (2020). Design and description of positions in the country's tribunals: Family courts.

Cohen, J. (1988). Statistical power analysis for behavioral science. hillsdale, new jersey: Lawrance erlbaum associates. *Inc., Publishers*.

Dhaliwal, I. and Hanna, R. (2017). The devil is in the details: The successes and limitations of bureaucratic reform in india. *Journal of Development Economics*, 124:1–21.

Dizon-Ross, R. (2019). Parents' beliefs about their children's academic ability: Implications for educational investments. *American Economic Review*, 109(8):2728–65.

Dodge, E., Neggers, Y., Pande, R., and Moore, C. (2021). Updating the state: Information acquisition costs and public benefit delivery. Technical report, Working paper, July.

Durham, C. and Becker, D. (2016). A case for court governance principles.

Engel, C. (2013). Legal experiments-mission impossible? *Available at SSRN 2276566*.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4):1661–1710.

Greiner, D. J. and Matthews, A. (2016). Randomized control trials in the united states legal profession. *Annual Review of Law and Social Science*, 12:295–312.

Hanson, R. A. and Ostrom, B. J. (2014). Achieving better court management through better data. *J. App. Prac. & Process*, 15:19.

Humphrey, S. E., Nahrgang, J. D., and Morgeson, F. P. (2007). Integrating motivational, social, and contextual work design features: a meta-analytic summary and theoretical extension of the work design literature. *Journal of applied psychology*, 92(5):1332.

Kleiman, M. (2009). Building a better understanding of trial courts. *Justice System Journal*, 30(3):321–326.

Mattsson, M. (2021). Information systems, service delivery, and corruption: Evidence from the bangladesh civil service. *Service Delivery, and Corruption: Evidence from the Bangladesh Civil Service (December 16, 2021)*.

Muralidharan, K., Niehaus, P., Sukhtankar, S., and Weaver, J. (2021). Improving last-mile service delivery using phone-based monitoring. *American Economic Journal: Applied Economics*, 13(2):52–82.

OECD (2014). Data-driven innovation for growth and well-being.

Ostrom, B. J. and Hanson, R. A. (2010). *Achieving high performance: A framework for courts*. National Center for State Courts.

Pomeranz, D. (2015). No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review*, 105:2539–69.

Rasul, I. and Rogger, D. (2018). Management of bureaucrats and public service delivery: Evidence from the nigerian civil service. *The Economic Journal*, 128(608):413–446.

Savoldelli, A., Codagnone, C., and Misuraca, G. (2014). Understanding the e-government paradox: Learning from literature and practice on barriers to adoption. *Government Information Quarterly*, 31:S63–S71. ICEGOV 2012 Supplement.

West, T., Thorson, K., Grant, H., and Rock, D. (2018). Asked for vs. unasked for feedback: An experiemental study. *NeuroLeadership Journal*.

Whidden, C., Kayentao, K., Liu, J. X., Lee, S., Keita, Y., Diakité, D., Keita, A., Diarra, S., Edwards, J., Yembrick, A., et al. (2018). Improving community health worker performance by using a personalised feedback dashboard for supervision: a randomised controlled trial. *Journal of global health*, 8(2).

3A Appendix

3A.1 Pre-Analysis Plan

AEA RCT Registry

The American Economic Association's registry for randomized controlled trials

Welcome back, Paloma!

Sign out Profile My Trials

About Registration Guidelines Data FAQ Advanced Search SEARCH

INFORMATION PROVISION AND COURT PERFORMANCE: EXPERIMENTAL EVIDENCE FROM CHILE

LAST REGISTERED ON JUNE 17, 2021

VIEW TRIAL HISTORY

Restricted Access Edit Trial

Pre-Trial

Trial Information

GENERAL INFORMATION

Title

Information Provision and Court Performance: Experimental Evidence from Chile

RCT ID

AEARCTR-0005512

Initial registration date

December 06, 2020

Initial registration date is when the trial was registered.

It corresponds to when the registration was submitted to the Registry to be reviewed for publication.

First published

December 07, 2020, 10:54 AM EST

First published corresponds to when the trial was first made public on the Registry after being reviewed.

Last updated

June 17, 2021, 4:42 AM EDT

Last updated is the most recent time when changes to the trial's registration were published.

LOCATIONS

Country

Chile

Region

National

PRIMARY INVESTIGATOR

Name
Paloma Carrillo

Affiliation
Toulouse School of Economics

Email
paloma.carrill@gmail.com

OTHER PRIMARY INVESTIGATOR(S)

PI Name
Daniel Chen

PI Affiliation
IAST/ Toulouse School of Economics

PI Email

PI Name
Bernardo Silveira

PI Affiliation
University of California, Los Angeles

PI Email

PI Name
Manuel Ramos

PI Affiliation
World Bank

PI Email
mramosmaqueda@worldbank.org

ADDITIONAL TRIAL INFORMATION

Status
In development

Start date
2020-01-01

End date
2021-09-30

Keywords
Behavior, Governance

Additional Keywords
Judicial, tribunal, judicial system, behavioral, information provision, Chile

JEL code(s)

D04

Secondary IDs

H43, D63

Abstract

Previous studies have shown that behavioral nudges can be a cost-effective tool to influence changes in people's actions. In this study, we aim to test whether nudging court managers through informing them on how their court performs in absolute and relative terms can improve court productivity. Moreover, we test if there is any difference if the information about the court performance is given in contrast and relation to self past performance or if the information is relative to other courts' performance.

External Link(s)

REGISTRATION CITATION

Citation

Carrillo, Paloma et al. 2021. "Information Provision and Court Performance: Experimental Evidence from Chile." AEA RCT Registry. June 17. https://doi.org/10.1257/rct.5512-1.1

Sponsors & Partners

PARTNER

Name

Department of Institutional Development (DDI) at the Administrative Corporation of the Chilean Judiciary (CAPJ)

Type

government

URL

Experimental Details

INTERVENTIONS

Intervention(s)

In this study, we aim to test whether nudging court managers through informing them on how their court performs in absolute and relative terms can improve court productivity.

Intervention (Hidden)

The Department of Institutional Development (DDI) of the Chilean judicial system developed an electronic platform called Quantum in 2018. Quantum displays comprehensive information on court indicators, such as the number of cases filed, the case clearance rate, the average duration of cases finished in a month, and the percentage of realized hearings. It also allows users to compare their courts to other courts in the

same jurisdiction. Quantum was launched in 2018, but take-up has been limited: 20% of court managers have never logged in, and overall there is an average of 20 logins per court manager in 1 year and 2 months. The platform is technologically well developed and rich in information, yet it is unclear whether it has any impact on the management of the court. Our project consists of evaluating an intervention through a randomized controlled trial (RCT) with two main branches. First, we will (randomly) promote the Quantum platform in multiple ways, such as sending court managers a survey that implicitly markets the platform, making phone calls, and sending them emails. Second, we will (also randomly) provide court managers a new dashboard that summarizes the main statistics displayed in Quantum and compares them to themselves to a reference group of courts. There would be a total of six distinct intervention or treatments:

Treatment 0: Control. No change to their Quantum dashboards nor provided with any Quantum promotion.

Treatment 1: Quantum Promotion
The tribunals randomized into Treatment 1 will have their court staff receive both emails with a Quantum link to increase accessibility and salience of Quantum, and a small baseline and post-intervention survey that includes Quantum promotion. The baseline survey given to the court managers at the beginning of the RCT will ask them about their beliefs about some productivity metrics, how much these metrics affect their decision at work, inform them that these metrics can be seen in Quantum, and describe the effect of Quantum usage on people's productivity through the results of an event study using historical data. A sample survey is provided in the appendix.

Treatment 2: No Quantum Promotion + New Dashboard
The tribunals randomized into treatment 2 will not receive any promotion but will have their home page in Quantum, what we call the dashboard, present various statistics at the tribunal level.

Treatment 3: Quantum Promotion + New Dashboard
The tribunals randomized into treatment 3 will receive the same promotion as that in treatment 1 and the new dashboard as in Treatment 2.

Treatment 4: No Quantum Promotion + New Dashboard + Comparative to others
The tribunals in treatment 4 will receive the new dashboard plus another tab or pop-up window that focuses on the tribunal's best performing and worst-performing dimension from the previous month in comparison to the performance of peer tribunals (same competence) in the same month. This comparison leans into social comparison motivation.

Treatment 5: Quantum Promotion + New Dashboard + Comparative to others
The tribunals in treatment 4 will receive the new dashboard plus another tab or pop-up window that focuses on the tribunal's best performing and worst-performing dimension from the previous month in comparison to the performance of peer tribunals (same competence) in the same month. This comparison leans into social comparison motivation.

The court managers' job satisfaction level will be measured with pre and post-surveys to court managers that measure perceptions of their tribunals and their satisfaction with their positions. By informing the court managers about their court's standing in the new dashboard and comparative to others, this could change how empowered or satisfied a court manager is with his or her position and power role.
Overall, the objectives of the survey for court staff are threefold. First, it will measure their knowledge of Quantum statistics. How close or far is their perception of their court's performance from the truth. We can later use this information (prior beliefs) to understand if greater access to Quantum updates the beliefs closer to the truth when we measure their posteriors (survey at the end of intervention). Notice that the measurement of beliefs and opinions is something unique to the survey that the rest of the interventions cannot. Second, the survey will allow us to understand if the court staff find the statistics important and in what order of importance. This is useful for Quantum to know which variables are important for users and make them more salient in the dashboard or in the rest of the Quantum pages. That is, the results from the survey can help tailor the intervention to make it more effective. Third, the survey will promote Quantum as a source of accurate and useful information through the event study results. This may help influence those that are skeptical of Quantum to give it a chance.

Intervention Start Date

2020-12-14

Intervention End Date

2021-09-30

PRIMARY OUTCOMES

Primary Outcomes (end points)

Case clearance rate, average length for filing cases (days), average length for ending cases, the average time the court needs to provide a written submission during the consultation period, percentage of writing resolved with 3-5 days, average number of days to program a hearing, percentage of hearing that started with a delay of 15 minutes, percentage of cases pending for more than 1-2 years, appeal rate, and number of cases appealed.

On the promotion intervention, the main primary outcome is number of logins per court manager to the Quantum platform

Primary Outcomes (explanation)

SECONDARY OUTCOMES

Secondary Outcomes (end points)

Court managers job satisfaction

Secondary Outcomes (explanation)

The court managers' job satisfaction level will be taken from the pre and post-surveys to court managers that measure perceptions of their tribunals and their satisfaction with their positions. By informing the court managers about their court's standing it could change how empowered or satisfied a court manager is with his or her position and power role.

EXPERIMENTAL DESIGN

Experimental Design

The program will have six distinct treatments. The treatments will combine promoting the usage of an electronic platform that contains information on their court performance and providing distinct homepages in this platform that will summarize the courts performance stressing the weaknesses and strengths of the court in comparison to a reference group.

Experimental Design Details

First, we will (randomly) promote the Quantum platform in multiple ways, such as sending court managers a survey that implicitly markets the platform, making phone calls, and sending them emails. Second, we will (also randomly) provide court managers a new dashboard that summarizes the main statistics displayed in Quantum and compares them to themselves in the past or to a reference group of courts. There would be a total of six treatments: (0) no quantum promotion no new dashboard (control) (1) quantum promotion, (2) no quantum promotion and new Quantum dashboard, (3) quantum promotion and a new dashboard, (4) no quantum promotion, new dashboard, and comparative that emphasizes the strongest and weakest indicators for that month in comparison to a similar group of courts in that same month, and (5) quantum promotion, new dashboard, and comparative that emphasizes the strongest and weakest indicators for that month in comparison to a similar group of courts in that same month.

Given that the information in the Quantum platform is updated daily and our dashboards are updated with monthly data, we will have multiple pre-treatment observation and many post-treatment observations.

Randomization Method

Randomization done in office by a computer.

Randomization Unit

The unit of randomization is the court. The randomization was stratified by size (small and big) and court type (7 distinct ones).

Was the treatment clustered?

No

EXPERIMENT CHARACTERISTICS

Sample size: planned number of clusters

346 courts

Sample size: planned number of observations

346 courts

Sample size (or number of clusters) by treatment arms

57 courts stay as control, 57 courts receive T1, 58 courts receive T2, 58 courts receive T3, 58 courts receive T4, and 58 courts receive T5.

Minimum detectable effect size for main outcomes (accounting for sample design and clustering)

Supporting Documents and Materials

IRB

INSTITUTIONAL REVIEW BOARDS (IRBS)

IRB Name

Toulouse School of Economics

IRB Approval Date

2020-01-18

IRB Approval Number

N/A

Analysis Plan

Post-Trial

Post Trial Information

Data Publication

Reports, Papers & Other Materials

3A.2 Survey to Court Managers

A baseline survey was rolled out to all 346 court managers to capture managers' perceptions on their roles and the Quantum platform. The survey was launched at the end of January of 2020, and responses were collected until the end of March of 2020. The survey was sent internally by the Department of Institutional Development of the Chilean Judiciary by email to all 346 court administrators. The baseline survey had four sets of questions related to (1) the perception of the performance of the tribunal, (2) their feedback on their job position and obstacles, (3) the knowledge and usage of Quantum, and (4) promoting and disclosing relevant information about Quantum. The questions related to knowledge and usage, and promotion of Quantum were asked to a subset of the population (a survey treatment group).

We received 132 responses, representing a 38.2% response rate, which span all different types of courts and across all seventeen Chilean jurisdictions. The survey's key findings can be divided into the first three main topics. First, regarding court performance and indicators, court managers strongly agree that court indicators are important to know for their job (see Table 8). Yet, court managers cannot accurately recall their courts' performance based on the most common indicator of case clearance. Furthermore, court managers are more likely to over-predict their performance compared to other tribunals. Second, one of the main actions taken by court managers in response to indicators is improving the scheduling of hearings (see Table 9 below for other responses). Court managers find the main barrier to their position is the power dynamics between court managers and judges. Third, Quantum is underutilized. About one-tenth of the court managers do not know about Quantum, and 17% do not use it. Older tribunal administrators are less likely to use Quantum. Of the court managers that are Quantum users, 69% are not using the platform to inform their work decisions. The main comment regarding improvements to Quantum is that it could include additional statistics and information.

Table 8: Descriptive Statistics on Managerial Practices

Statement	Mean	1-Strongly Disagree	2-Disagree	3-Neutral	4-Agree	5-Strongly Agree
Knowledge of indicators	4.67	0	1	2	31	82
Decision based on indicators	4.44	0	0	8	49	59
Satisfied with impact	4.38	0	1	3	63	49
Tracking indicators important	3.91	1	5	28	52	30
Indicators influence decisions	4.18	0	1	13	66	36
Many actions to improve performance	4.40	0	1	5	57	53
Work is relevant for well functioning court	4.69	0	0	0	36	80
Obstacles to improve performance	3.78	3	15	19	47	32

Responses from the court managers survey.

Table 9: Impact of Court Indicators on Court Manager's Actions

Court manager's response actions based on the following indicators			
Number of incoming cases	Number of pending cases	Number of resolved cases	Rate of realized hearings
Courtroom distribution/ agenda (24%)	Review causes of cases on hold (16%)	Avoid cancelling/rescheduling hearings and look over agenda (21%)	Effectiveness in scheduling hearings and preventing cancellation (43%)
Balance of workload/ redistribution of tasks (15%)	Ensure hearings occur (13%)	Strengthen reviewing causes (14%)	Improve usage of courtrooms (13%)
Increase number of judges, employees and/or hearings (9%)	Increase number of hearings/ schedule simultaneous hearings/ improve scheduling (10%)	Improve effectiveness of courtrooms (7%)	Distribute hearings) by topic (8%
	Balance workload (8%)		

Percentages based on text analysis from court managers survey responses.

3A.3 Email Versions

Email on January 26

Dear Court Managers,

Hoping that you are well, we remind you that for a couple of years, the institution has advanced in the construction of the Quantum management tool, which consolidates different jurisdictional indicators by court, in order to provide information for decision-making. This tool has been valued by many courts, since it allows simplifying the consolidation of statistical data to support jurisdictional work.

Did you know ...?

Courts that use Quantum the most generally have a higher term rate. That is, for each case admitted, they tend to solve a greater number of cases in proportion.

Courts that use Quantum the most generally have a smaller inventory of old cases.

For those with feedback treatment

"During the last months, we have worked with a team of court managers in the construction of a new dashboard, which displays key indicators, necessary to support the management of family courts. In this sense, we invite you to log into Quantum and explore this new functionality. This dashboard also incorporates a window that compares the performance of the courts with a court of similar characteristics to yours."

You can access this platform at https://quantum.pjud (through VPN). To access the new dashboard, they must click on the name of their court within their jurisdiction. In case of any observation, requirement and even if you do not have your login credentials, please write to the email quantum@pjud.cl.

Email promotion in February, March, and May

Dear Court [Name],

Please join us in congratulating the family court managers who used the Quantum platform the most during [previous month] 2021:

Name of manager 1 - Name of court 1

Name of manager 2 - Name of court 2

Name of manager 3 - Name of court 3

For those with at least one login on previous month:
"We trust that all of you will continue to use Quantum to track your indicators, evaluate your court's performance, and compare it to other courts."

For those with no logins on previous month:
"We noticed that you did not log into Quantum during [previous month] 2021. You are missing the opportunity to track your metrics, evaluate your court's performance, and compare it to other courts!"

Did you know that..

Most Quantum users find Quantum easy to use, displays information clearly, contains reliable indicators, and information that is useful and relevant to their work. Family courts that use Quantum multiple times a month have higher rates of termi-

nation of cases pending for more than two years.

You can access this platform at https://quantum.pjud (through VPN). To access the new dashboard, they must click on the name of their court within their jurisdiction. In case of any observation, requirement and even if you do not have your login credentials, please write to the email quantum@pjud.cl.

3A.4 Minimum Detectable Effect

Table 10: Minimum Detectable Effect in Standard Deviations

	Email Promotion vs. No Email	Feedback vs No Feedback
Case Clearance	0.43	0.46
Timely Motion Resolution	0.52	0.55
Realized Hearings	0.42	0.54
Timely Case Resolution	0.52	0.57
Timely Hearing Programming	0.48	0.54

These are the minimum detectable effect in standard deviations computed using the ANCOVA methodology and setting power at 80%, alpha 0.05, 25 observations for control and treatment, 4 pre-treatment observations, and 5 post-treatment observation.

For the five main indicators of interest, the minimum detectable effect, considering power at 80%, alpha at 0.05, and the 49 tribunals and number of monthly observations, is between .42 and 0.57 standard deviations. According to Cohen (1988)who proposes that an effect of 0.2 standard deviation is “small”, 0.5 is “medium” and 0.8 is “large”, even with our small sample of tribunals we have enough power to detected ”medium” size effects.

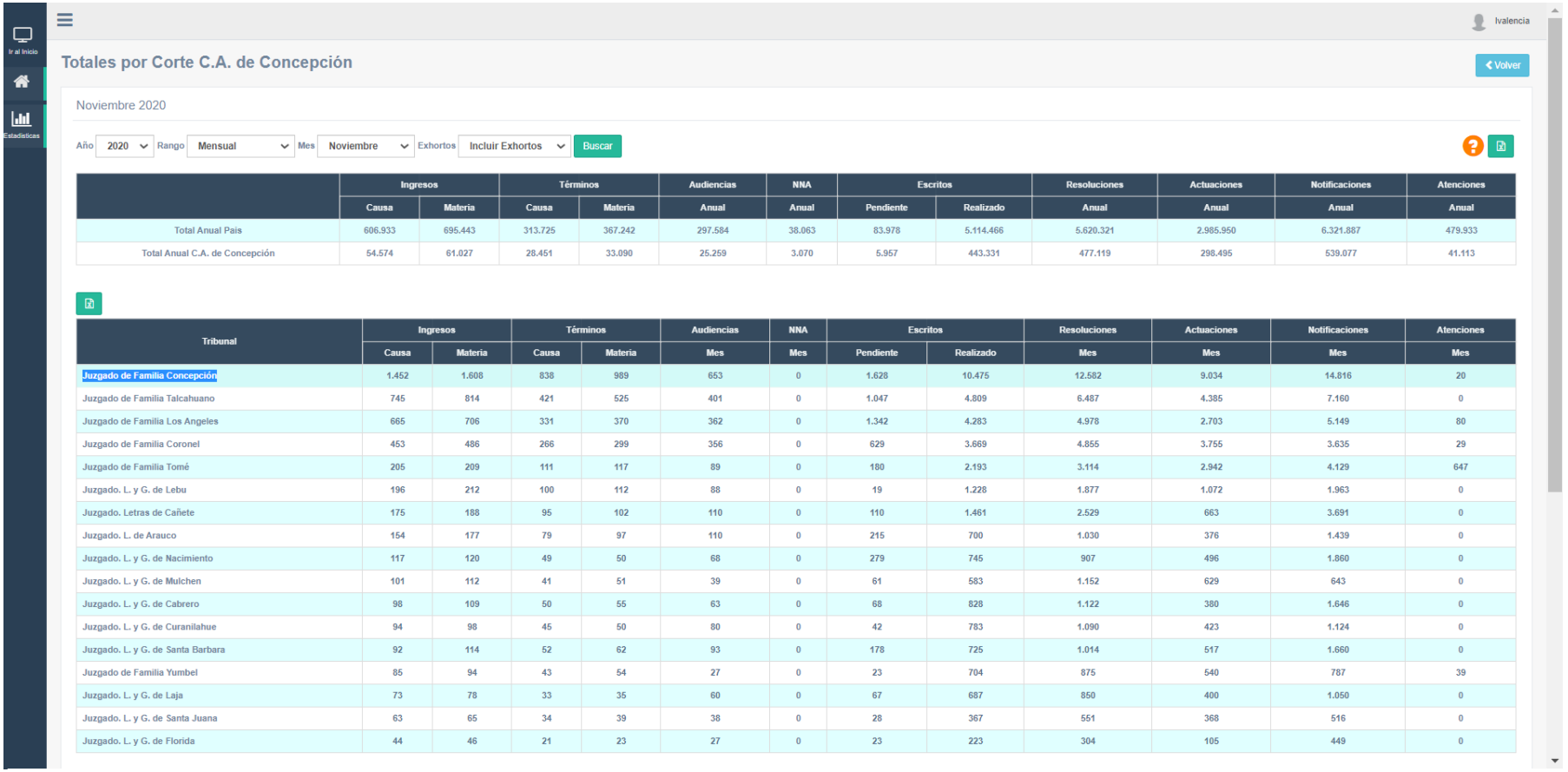
3A.5 Figure Appendix

Figure 1, 2, and 3 show examples of the Quantum platform for the control, dashboard, and pop-up treatments.

3A.6 Regression Appendix

The present Appendix contains regression results that complement those provided in the main text. We first provide the Average Treatment Effects versions of the main

Figure 1: Old Quantum main homepage



	Ingresos		Términos		Audiencias	NNA	Escritos		Resoluciones	Actuaciones	Notificaciones	Atenciones
	Causa	Materia	Causa	Materia	Anual	Anual	Pendiente	Realizado	Anual	Anual	Anual	Anual
Total Anual Pais	606.933	695.443	313.725	367.242	297.584	38.063	83.978	5.114.466	5.620.321	2.985.950	6.321.887	479.933
Total Anual C.A. de Concepción	54.574	61.027	28.451	33.090	25.259	3.070	5.957	443.331	477.119	298.495	539.077	41.113

Tribunal	Ingresos		Términos		Audiencias	NNA	Escritos		Resoluciones	Actuaciones	Notificaciones	Atenciones
	Causa	Materia	Causa	Materia	Mes	Mes	Pendiente	Realizado	Mes	Mes	Mes	Mes
Juzgado de Familia Concepción	1.452	1.608	838	989	653	0	1.628	10.475	12.582	9.034	14.816	20
Juzgado de Familia Talcahuano	745	814	421	525	401	0	1.047	4.809	6.487	4.385	7.160	0
Juzgado de Familia Los Angeles	665	706	331	370	362	0	1.342	4.283	4.978	2.703	5.149	80
Juzgado de Familia Coronel	453	486	266	299	356	0	629	3.669	4.855	3.755	3.635	29
Juzgado de Familia Tomé	205	209	111	117	89	0	180	2.193	3.114	2.942	4.129	647
Juzgado. L. y G. de Lebu	196	212	100	112	88	0	19	1.228	1.877	1.072	1.963	0
Juzgado. Letras de Cañete	175	188	95	102	110	0	110	1.461	2.529	663	3.691	0
Juzgado. L. de Arauco	154	177	79	97	110	0	215	700	1.030	376	1.439	0
Juzgado. L. y G. de Nacimiento	117	120	49	50	68	0	279	745	907	496	1.860	0
Juzgado. L. y G. de Mulchen	101	112	41	51	39	0	61	583	1.152	629	643	0
Juzgado. L. y G. de Cabrero	98	109	50	55	63	0	68	828	1.122	380	1.646	0
Juzgado. L. y G. de Curanilahue	94	98	45	50	80	0	42	783	1.090	423	1.124	0
Juzgado. L. y G. de Santa Barbara	92	114	52	62	93	0	178	725	1.014	517	1.660	0
Juzgado de Familia Yumbel	85	94	43	54	27	0	23	704	875	540	787	39
Juzgado. L. y G. de Laja	73	78	33	35	60	0	67	687	850	400	1.050	0
Juzgado. L. y G. de Santa Juana	63	65	34	39	38	0	28	367	551	368	516	0
Juzgado. L. y G. de Florida	44	46	21	23	27	0	23	223	304	105	449	0

Figure 2: New simplified homepage

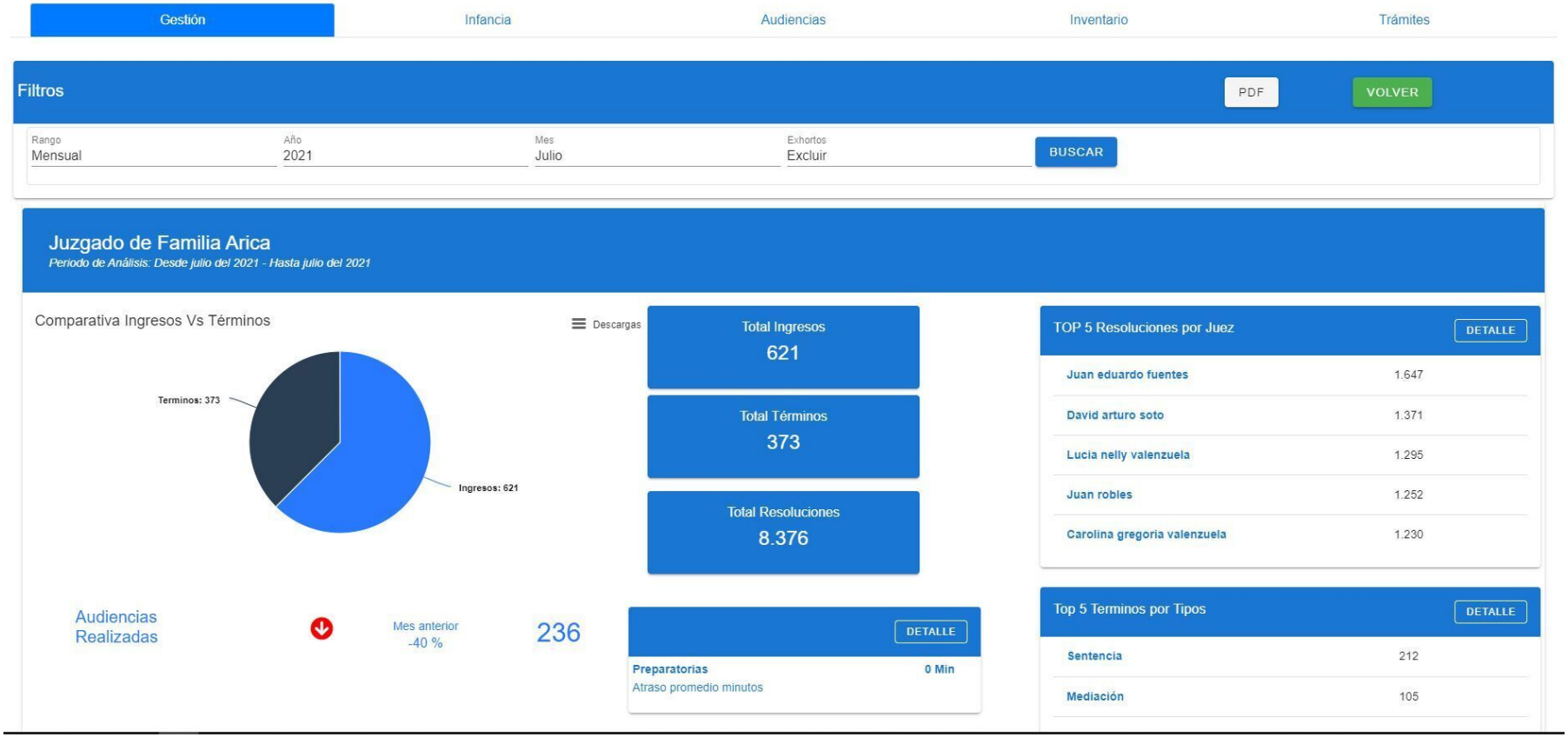
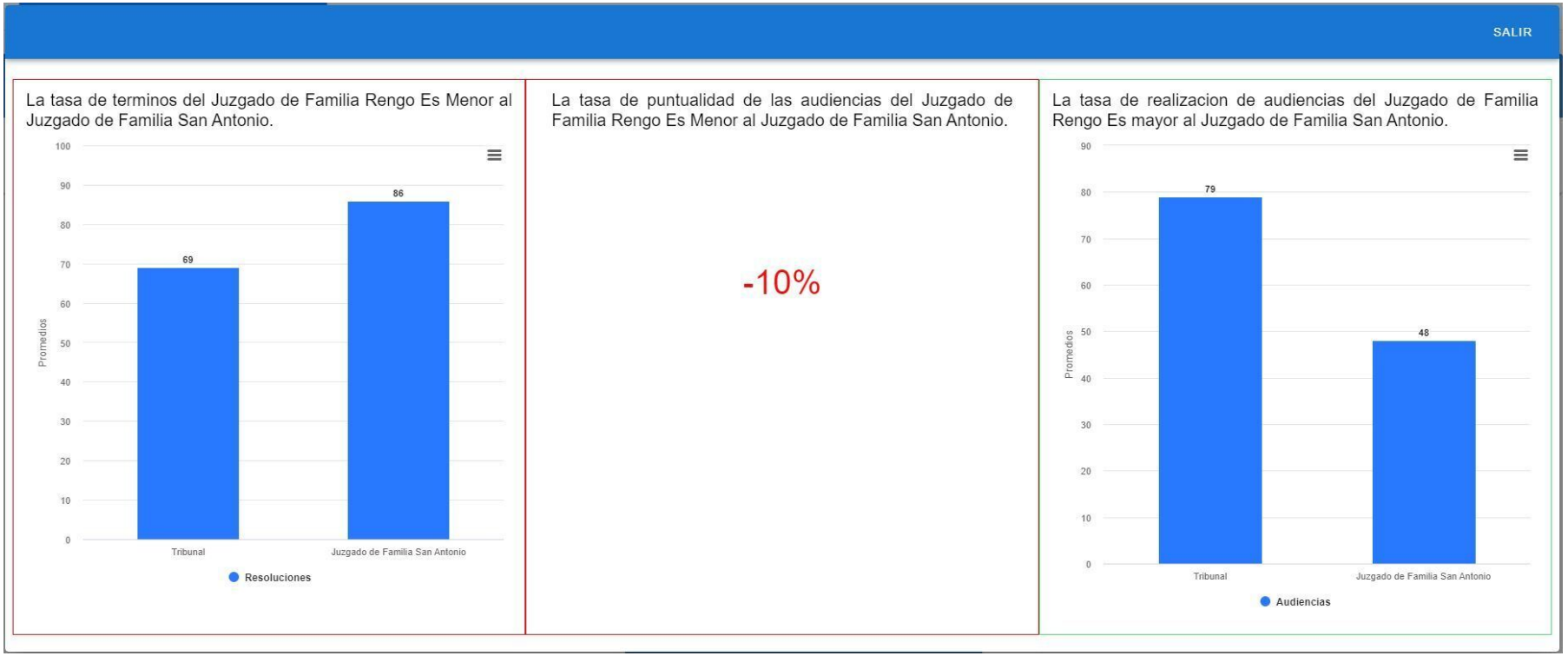


Figure 3: Pop-up window with social comparison



results of the paper. Then we present an analysis of the impact of our interventions on the login behavior of the court managers.

Average Treatment Effects and Local Average Treatment Effects

This section presents Average Treatment Effects (ATE) associated with the and Local Average Treatment Effects (LATE) from Section 3.5, Tables 3.4 and 3.5 in the main text. For ease of comparison, Tables 11 and 12 shows both the ATE and LATE estimates.

Quantum access

In this section, we evaluate the impact of the email promotion and feedback interventions on the managers' usage of Quantum. We consider the following specification:

$$y_{it} = \beta_0 + \beta_1 Email_i + \beta_2 \mathbf{X_i} + \beta_3 l_{it_0} + \beta_4 \gamma_t + \epsilon_{it}, \tag{2}$$

where y_{it} is the number of logins by court i in month $t \in$\{February, March, April, May, June\}; $\mathbf{X_i}$ is a vector of characteristics of court i, which includes the court's manager's sex, age, and tenure, and the stratified variable of a dummy for large courts computed using the 2019 number of incoming cases; l_{it_0} is a dummy indicating whether the manager logged at least once into Quantum in the pre-treatment month of January; and γ_t is a month fixed effect.

We can adapt (2) to a probit specification, in which the dependent variable is a dummy indicating whether the manager of court i logged at least once in month t. We can also substitute indicators of our other treatments for $Email_i$ in the specification, to assess whether the simplified or social comparison feedback treatments impacted Quantum usage.

During the intervention, from January 26 2021 to June 30 2021, the volume of logins seems to follow a seasonality pattern. Figure 4 shows the logins per week for the 50 courts in our intervention, beginning in January 1. In the figure, the vertical line shows the beginning of the treatment. The weeks following the intervention saw

Table 11: Average Treatment Effects (ATE) and Local Average Treatment Effects (LATE)

Case Clearance	ATE (1)	LATE (2)	ATE (3)	LATE (4)	ATE (5)	LATE (6)
Email promotion	0.0995	0.222			0.152	0.0270
	(0.127)	(0.284)			(0.240)	(0.607)
Simplified feedback			0.368**	0.957**		
			(0.180)	(0.475)		
Social comparison feedback			0.107	0.500		
			(0.174)	(0.656)		
Feedback					0.285	0.807
					(0.184)	(0.591)
Email * Feedback					-0.0617	-0.0675
					(0.288)	(0.805)
Timely Motion Resolution	ATE (1)	LATE (2)	ATE (3)	LATE (4)	ATE (5)	LATE (6)
Email promotion	0.0933***	0.208***			0.154**	0.0909
	(0.0315)	(0.0711)			(0.0767)	(0.186)
Simplified feedback			0.149***	0.360***		
			(0.0415)	(0.112)		
Social comparison feedback			0.170***	0.587***		
			(0.0378)	(0.149)		
Feedback					0.197***	0.492***
					(0.0491)	(0.167)
Email * Feedback					-0.0799	-0.0743
					(0.0944)	(0.255)
Realized Hearings	ATE (1)	LATE (2)	ATE (3)	LATE (4)	ATE (5)	LATE (6)
Email promotion	-0.335***	-0.794***			-0.786***	-1.361***
	(0.0863)	(0.217)			(0.130)	(0.401)
Simplified feedback			-0.180	0.360***		
			(0.119)	(0.112)		
Social comparison feedback			-0.223*	0.587***		
			(0.123)	(0.149)		
Feedback					-0.557***	-1.011***
					(0.121)	(0.308)
Email * Feedback					0.630***	1.148**
					(0.153)	(0.558)
Control Group	No email promotion		No feedback		No email and no feedback	

All regression have as covariates the strata variable, three manager characteristics, four pre-treatment lagged values of the dependent variable, pre-treatment login dummy, and month fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

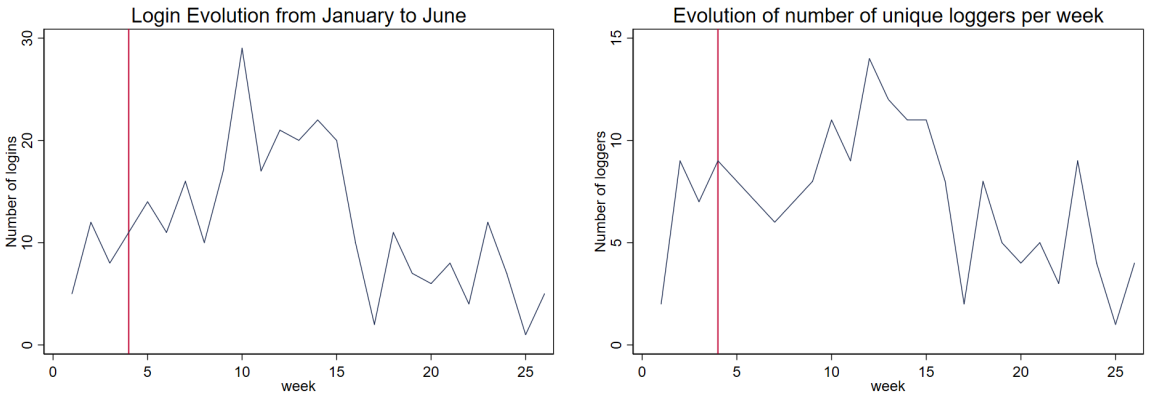
Table 12: Average Treatment Effects (ATE) and Local Average Treatment Effects (LATE)

Timely case resolution	ATE (1)	LATE (2)	ATE (3)	LATE (4)	ATE (5)	LATE (6)
Email promotion	0.0564*	0.118*			0.0233	0.0266
	(0.0323)	(0.0695)			(0.0665)	(0.188)
Simplified feedback			0.0393	0.0880		
			(0.0409)	(0.0927)		
Social comparison feedback			-0.0243	-0.0891		
			(0.0469)	(0.157)		
Feedback					-0.0119	-0.0607
					(0.0328)	(0.109)
Email * Feedback					0.0503	0.153
					(0.0832)	(0.245)
Hearing Programming	ATE (1)	LATE (2)	ATE (3)	LATE (4)	ATE (5)	LATE (6)
Email promotion	-0.0565	-0.127			0.385***	0.737**
	(0.0563)	(0.133)			(0.0864)	(0.310)
Simplified feedback			0.200***	0.563**		
			(0.0692)	(0.236)		
Social comparison feedback			-0.0170	-0.0212		
			(0.0704)	(0.320)		
Feedback					0.383***	1.330***
					(0.0691)	(0.353)
Email * Feedback					-0.635***	-1.656***
					(0.110)	(0.441)
Control Group	No email promotion		No feedback		No email and no feedback	

All regression have as covariates the strata variable, three manager characteristics, four pre-treatment lagged values of the dependent variable, pre-treatment login dummy, and month fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

a general upward trend in the number of logins, peaking on week 10 (March 8-14), and then a downward trend. We see a similar pattern if we measure login activity as the number of court managers who logged in at least once in the week—although the peak in this series happens somewhat later, on week 12.

Figure 4: Evolution of Quantum Logins and Loggers



A) Number of logins B) Unique loggers

Note: Red x-line at week 4 marks the beginning of the intervention

In the first part of Table 13, we explore the impact of the email promotion on Quantum logins through two distinct econometric models: random effects (columns 1 and 2) and random-effects probit (columns 3 and 4). For each of them, the reference control group is those not receiving email promotions. All specifications show a positive effect of email promotion on logins. For the specifications that include our most extensive set of controls (month fixed effects, characteristics of the managers and pre-treatment outcomes), the effects are significant. The random-effects regression in column (2) suggests that the email promotion caused the number of logins per month in a court to increase by 0.5 logins compared to those that did not receive the email promotion. The average number of logins over the 49 family courts was approximately 1.1 logins per month; thus, the relative impact of the email promotion is large.

In the second part of Table 13, we explore the impact of the simplified feedback and social comparison feedback on Quantum logins through the same two distinct econometric models as before. In each of them, the reference control group is those

with no feedback. The estimates for both simplified feedback and social comparison feedback are quite imprecise. We cannot conclude that either had any differential impact on the number of logins. To further explore the impacts of the feedback and email promotions, the third part of Table 13 shows results of specifications that include the feedback and email treatments, and their interaction. The results indicate that the email promotion caused an average increase of 1.5 logins per month.[12] The impact of the feedback is again not statistically significant on logins but the point estimates are now positive. The results for the random effects probit in column (3) and (4) provide similar evidence of an increase in logins for the email promotion but with less precision. We find that the email promotion makes the likelihood of login into Quantum at least once in a month increase by 12.5% at the ten percent level significance with respect to those that did not get the email promotion.

To summarize, Table 13 shows strong evidence that the email promotion increased the monthly number of logins. We also obtain suggestive evidence that the feedback intervention increases the number of logins.

Finally, Table 14 reproduces the results from the email promotion regressions from Table 13—this time reporting the estimated coefficients for select control variables. In particular, the results indicate that more experienced and older managers log into Quantum less than their younger and less experienced colleagues.

[12]The coefficients for email in the first and third parts of the table do not need to be equal to each other, as the reference control groups in the two specifications are different. Particularly, if the feedback treatment has any positive impact on logins, then the coefficient for email in Panel C should be larger, as the reference control group for Panel A includes courts receiving feedback without email promotion.

Table 13: The Average Treatment Effects of Email and Feedback on Quantum Logins

	Random-effects		Random-effects probit	
	Number of logins per month		At least one login per month	
	(1)	(2)	(3)	(4)
Email	0.272	0.469**	0.0887	0.125*
	(0.200)	(0.221)	(0.0659)	(0.0710)
Control group	No email			
Simplified feedback	-0.140	-0.277	0.0137	-0.00383
	(0.265)	(0.282)	(0.0839)	(0.0880)
Social comparison feedback	-0.170	-0.274	-0.124	-0.134
	(0.237)	(0.273)	(0.0836)	(0.0862)
Control group	No feedback			
Feedback	-0.254	0.429	-0.0902	0.0168
	(0.323)	(0.354)	(0.101)	(0.0926)
Email	0.116	1.453**	0.0314	0.218
	(0.435)	(0.571)	(0.139)	(0.158)
Feedback*Email	0.224	-1.461**	0.0818	-0.144
	(0.498)	(0.688)	(0.164)	(0.189)
Control group	No email and no feedback			
N	246	246	246	246
DV Mean	1.07	1.07	0.35	0.35
Strata	✓	✓	✓	✓
Month FE	✓	✓	✓	✓
Admin. Characteristics		✓		✓
Pre-treatment Login		✓		✓

RE Probit shows the average marginal effects. The reference control group for each is those that did not receive an email promotion. Manager controls include male, three age brackets dummies, and the number of years in position. Standard errors are clustered at the court level and bootstraped.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: The Average Treatment Effects of the Email Promotion on Quantum Logins

	Random-effects		Random-effects probit	
	Number of logins per month		At least one login per month	
	(1)	(2)	(3)	(4)
Email Promotion	0.272	0.469**	0.0887	0.125*
	(0.200)	(0.221)	(0.0659)	(0.0710)
Strata (size)	1.130***	0.748***	0.244***	0.172**
	(0.202)	(0.206)	(0.0690)	(0.0736)
March	1.108***	1.126***	0.174*	0.176*
	(0.398)	(0.395)	(0.106)	(0.106)
April	0.231	0.248	0.0614	0.0630
	(0.289)	(0.276)	(0.0945)	(0.0994)
May	-0.279	-0.262	-0.0692	-0.0690
	(0.281)	(0.255)	(0.0897)	(0.0986)
June	-0.361	-0.344	-0.0429	-0.0445
	(0.296)	(0.272)	(0.0962)	(0.115)
Age (51-60)		-0.574**		-0.0390
		(0.263)		(0.0826)
Age (61-67)		-0.163		-0.107
		(0.256)		(0.109)
Experience		-0.0982***		-0.0145**
		(0.0236)		(0.00656)
January Logger		1.616***		0.230***
		(0.316)		(0.0863)
Constant	-0.861***	0.737*		
	(0.296)	(0.416)		
N	246	246	246	246
DV Mean				

RE Probit shows the average marginal effects. The reference control group for each is those that did not receive an email promotion. Standard errors are clustered at the court level and bootstraped.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$