ABSTRACT

Title of Dissertation: ESSAYS ON MACROECONOMICS

Kai Chung Oue
Doctor of Philosophy, 2023

Dissertation Directed by: Professor S. Borağan Aruoba
Department of Economics

This dissertation focuses on two topics in macroeconomics. The first two chapters study the interaction between family and inequality. The first chapter develops a framework for evaluating optimal income tax progressivity with endogenous marriage formation. The second chapter studies how migration frictions interact with parental locational and human capital investment decisions in the context of China. The third chapter revisits the role of strategic complementarity in pricing on monetary non-neutrality.

Chapter 1: The income distribution in the United States has undergone significant transformation over the last four decades. Accompanying the rise in inequality are an increase in positive assortative mating and a steady decline in marriages, especially amongst non-college educated individuals. The literature studying optimal redistribution typically ignores the marital margins, both in terms of who gets married and who marries whom, although they have important welfare consequences. This paper revisits the optimal design of taxation policies using a life-cycle consumption-savings model augmented with an endogenous marriage market. Through the lens

of the model, the optimal income tax features high degree of progressivity for both single and married households combined with large marriage bonuses.

Chapter 2: The *hukou* system is a unique institutional feature in China that restricts internal mobility mainly by blocking access of migrants to public services and programs. In this chapter, I propose that these institutional mobility restrictions play a role in enabling intergenerational transmission of income and economic status, in particular by suppressing human capital investment, both in terms of time and money, by parents on the lower end of the income distribution. Another argument that follows from my hypothesis is that migration frictions can generate a dynamic welfare cost through lower skill formation, in addition to the static losses that are emphasized in the literature. I build a two-region overlapping-generations model with household heterogeneity to show how migration decisions interact with parental human capital investment decisions to exert downward pressure on intergenerational mobility. A counterfactual analysis in which migration frictions are reduced suggests that removal of some *hukou* restrictions can increase the total human capital stock in the economy by 4%, increase aggregate output by 7%, and reduce the intergenerational rank-rank elasticity by at 7 percentage points.

Chapter 3: This chapter proposes a parsimonious framework for real rigidities, in the form of strategic complementarities, that can generate real and nominal dynamics and match key features of the data across several literatures. Existing menu-cost models featuring strategic complementarities require unrealistically volatile shocks to idiosyncratic productivity to be consistent with pricing moments. We develop a simple menu-cost model with strategic complementarities along with idiosyncratic productivity and demand shocks that are disciplined by the data. This approach allows us to overcome previous criticism from analysis of models that employ only an idiosyncratic productivity shock and calibrate solely using data from the price-adjustment literature. Despite its simplicity, the model can generate sizable monetary non-neutrality along with the magnitude of cost pass-through documented in previous studies, while also remaining consistent with micro pricing and markup evidence.

ESSAYS ON MACROECONOMICS

by

Kai Chung Oue

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Professor S. Borağan Aruoba, Chair
Professor Luminita Stevens
Professor Felipe Saffie
Professor Thomas Drechsel
Professor Susan Parker

Dedication

To my family, teachers, and friends.

Acknowledgments

I owe my most sincere gratitude to all the people around me for their continuous support over the years. The completion of my studies would not have been possible without any of you.

First and foremost, I am deeply indebted to my advisor Boragan Aruoba for his extraordinary mentorship, continuous guidance, trust, and patience as I learn to become a better economist everyday. I am equally indebted to Felipe Saffie for the countless conversations and advice about research and life. Both Boragan and Felipe have been extremely supportive and encouraging to me through the ups and downs of this difficult journey. I am forever grateful for the opportunity to work with them as research assistant and later on as co-author. I have learned a lot from them. Words cannot express my gratitude.

I would like to thank Luminita Stevens for her guidance over the years on the many projects that I worked on. She has been very patient and helpful with any ideas that I approached her with. I also benefited tremendously from my conversations with Thomas Drechsel and Pierre De Leo both professionally and personally. I am very happy to have John Shea as my Director of Graduate Studies. And I would like to thank Susan Parker for serving as the Dean's representative for my defense.

I would not be who I am as a researcher today if it was not for the great macro faculty at Maryland. John Shea, Allan Drazen, Boragan Aruoba, and John Haltiwanger taught a wonderful first-year sequence in macroeconomics, which introduced me to this fascinating field and inspired me to pursue a career as a macroeconomist. Borağan Aruoba and Felipe Saffie introduced me to computational methods which has become an irreplaceable tool in my research. John Haltiwanger and Luminita Stevens deepened my understanding of firm dynamics and monetary economics. I would also like to thank Nuno Limão and Eunhee Lee for teaching me international trade, as well as their suggestions on my research.

The brownbag workshops and seminars have been extremely enjoyable. I thank all the presenters for sharing their works and participants for raising excellent comments, which has expanded my knowledge of various topics in economics and sparked countless ideas. I also thank all the members in the macro reading group for the many discussions about cutting-edge research.

I will forever cherish my wonderful classmates and colleagues for their friendship and unconditional help over the last six years. In particular, I would like to thank Yi Liu, Dheeraj Chaudhary, Seho Kim, Edward Olivares, Nathalie Gonzalez-Prieto, Ming Fang, Zu Yao Hong, Panos Dimitrellos, Erika Dömötör, Mrin Chatterjee, Jiahao Tang, Arseniy Braslavskiy, Shihangyin Zhang, Wenbo Yu, Liang-Chieh Chiu, Donggyu Lee, Alvaro Silva, Kanat Isakov, Ben Rodriguez, Cameron Conrad, Chan Kim, Jaehong Choi, Sueyoul Kim, Rosa Lim, Sebastian Montano, Keaton Ellis, Qiyao Zhou, Siqin Ding, and Paul Cheung. I am also grateful to have great friends in Jerry Fan, Kyle Fung, and Jerry Zhang.

The departmental staff members have been very helpful before I had even set foot on campus. I especially thank Vickie Fletcher, Terry Davis, Amanda Statland, Jessica Gray, and Mark Wilkerson. I also greatly appreciate the assistance from Brian McLoughlin and Evan Schroeder.

I am thankful for all my teachers who have taught and nurtured me since pre-school. As a

Chinese proverb says, “knowledge, like the ocean, has no bounds”, but it must start somewhere. Thank you all for every little thing you have taught me, and more importantly for inspiring me to keep pursuing knowledge in life.

Lastly and most importantly, I am grateful for my family for their unconditional support and unwavering encouragement throughout my life. Thanks to my parents, Gary and Clara. Thanks to my brother Justin. Thanks to my grandmother Lam. I would not be me without any of them. And to my grandfathers and aunt, I hope I made you proud if you could be here to witness this.

Table of Contents

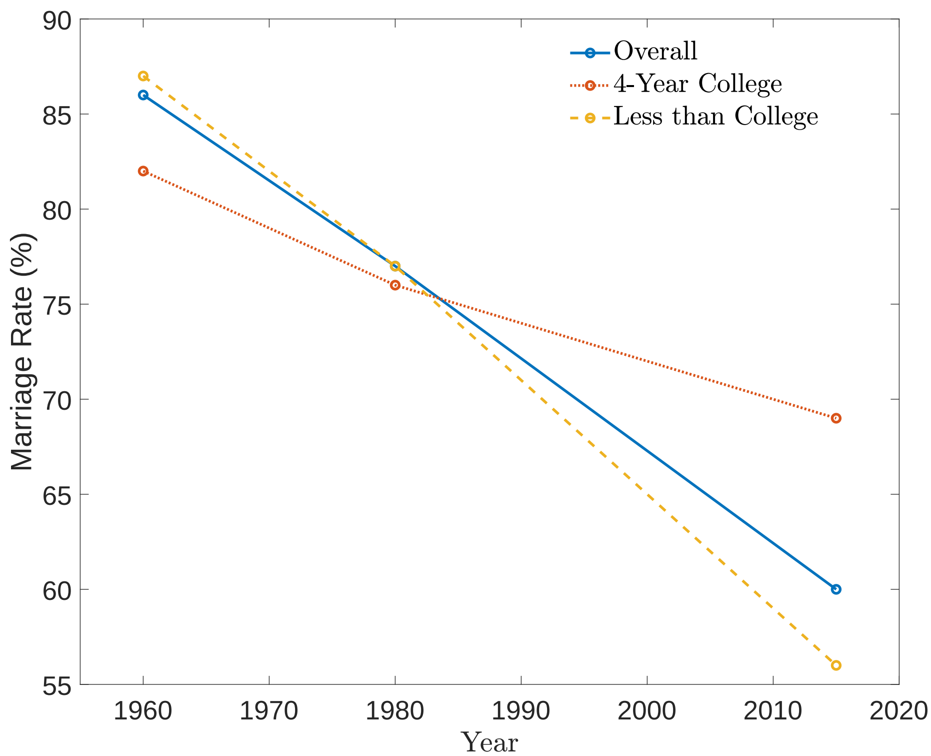
Chapter 1: Optimal Taxation in a Life-Cycle Model with Endogenous Marriage

1.1 Introduction

It is well-known that the income distribution in the United States has undergone substantial transformation over the last few decades, as characterized by rising return to skill (Acemoglu, 2002), declining gender wage gap (Blau and Kahn, 2017), and increasing residual dispersion (Guvenen et al., 2017; Heathcote et al., 2010). The change in the income distribution has sparked countless public and scholarly discussions on the optimal response of taxation policies. Although lesser known, this distributional change has been accompanied by drastic changes in family formation trends. In particular, there has been a sustained decline in the marriage rate and correspondingly an increase in the number of single households. Figure (1.1) shows a steady decline in the marriage rate, from 86% in 1960 to 77% in 1980 and eventually reaching 59% in 2015.[1] Furthermore, the decline in marriage is mostly driven by rising singlehood among individuals with low income and education. In 1960, marriage is more common among individuals with less than a 4-year college degree. Since then, the educational gap saw a sharp reversal, with marriage being much less prevalent among less educated individuals. As marriage becomes less common, the degree of assortative mating by education has also increased substantially

[1] There has also been a rise in cohabitation, but the share of partnered households was declining nonetheless even taking cohabitation into account.

Figure 1.1: Fraction of Population Ever Married, Aged 20-45



Note: Author's calculation based on 1960, 1980 Census microdata, and 2015 American Community Survey. Sample is restricted to adults aged between 20 to 45 that are not currently attending school.

(Greenwood et al., 2014) whereby individuals of the same education background are increasingly likely to marry each other.

Marriages are typically regarded as beneficial. Aside from companionship, marriages allows for consumption and resource sharing, intra-household risk-sharing, and specialization in tasks, just to name a few. As such, changes in marital patterns – who marries and to whom – has important implications on welfare. In fact, the decline of marriages among individuals of low socio-economic status and increased assortative mating suggest that welfare inequality might have risen more so than increasing personal income inequality alone would imply. This is because low income individuals are not able to form stable families and those miss out on gains from marriages. Even among married couple, increase in assortative mating would have led to uneven welfare across households.

Changes in the income distribution and family formation are hardly independent. The structural analysis of Greenwood et al. (2016) and Regalia et al. (2019) show that changes in the skill premium and gender wage gap since the 1980's can explain a sizable fraction of the decline in married households and increase in assortative matching. The quantitative exercise of Santos and Weiss (2016) suggests that more volatile income delays marriage. Empirical studies have also found income distribution to matter for marital outcome. Gould and Paserman (2003) argue that rising inequality reduces marriages as the option value of remaining single and waiting for another match increases. Autor et al. (2019) show that worsening labor market prospects stemming from import competition reduces marriageability of young men.

The intertwinement of the income distribution and marriage prompts a unified framework of income risks and marriage in the study of optimal redistribution policies. Since marriage patterns are shaped by the income distribution and simultaneously augment welfare consequences of inequality, they ought to be studied hand in hand. The goal of this paper is to develop a framework for analysis optimal taxation that takes into account the interaction between fundamental economic environment, policies, and household formation.

I develop a heterogeneous agent life-cycle model with labor income risks, endogenous marriages, and progressive income taxation à la Heathcote et al. (2017). The backbone of the model has close resemblance to a Bewley-Huggett-Aiyagari consumption-saving setup, where workers facing uninsurable income risks work, consume, and save under incomplete markets. The key departure is the addition of an endogenous marriage decision. In the model, single agents have the opportunity to meet other singles with the potential to form marriages. Upon marriage, couples enjoy increasing returns to consumption, the ability to specialize in home production and market work, and may provide each other with spousal insurance against adverse income shocks. The marriage decision – whether to get married and to whom – is endogenous to economic factors encompassing the wage structure and the idiosyncratic states of oneself and the potential partner, non-pecuniary preferences, as well as the tax regime. Following a Ramsey approach, the planner must internalize the effects of various redistribution plans on not only education choice and labor supply decisions of workers, but also their effects on the marriage market. At the same time, the planner is concerned about overall welfare which is contingent on the equilibrium marital patterns.

The model is calibrated to US data in years 1980 and 2015. Feeding in the observed changes in the wage structure, namely an increase in the skill premium and a reduction in the female wage gap, the model generates a relative decline in marriages among low income workers. The residual change in the marital market is attributed to changes in preferences towards marriage. Using the calibrated model as a laboratory, I find that the US tax system in both years feature too little progressivity. The optimal policy through the lens of the model is characterized by a moderately high degree of progressivity and large marriage bonuses accompanied by generous transfers to the poorest singles. Quantitatively, moving to the optimal policy can yield sizable welfare gains of 6% to 8% in consumption-equivalent terms.

Related Literature

There is a large literature on optimal taxation and redistribution, starting with the seminal works of Mirrlees (1971) and Diamond (1998). This literature typically considers a utilitarian planner whose objective is to maximize total social welfare and therefore prefers a more equal distribution of resources all else equal. Optimal policy is characterized by a balance between equality and the distortionary effects from taxation on labor supply and human capital investment. In response to changing economic environment in the US over the last few decades, Heathcote et al. (2020) find that optimal progressivity of income taxation depends crucially on the source of inequality. While the presence of uninsurable income risks call for more pregressive taxes, optimal taxes ought to be less progressive if inequality stems from a rise of the return of skill to avoid distorting skill accumulation. This project deviates from the standard literature by analyzing optimal policy in a framework with single and married households where family formation interacts with other elements of the model.

Along this dimension, this paper is closely linked to a growing literature that studies the role of families in public finance and macroeconomics. Kleven et al. (2009) extend the Mirrleesian approach to study the optimal taxation of couples. Alesina et al. (2011) studies optimal gender-based taxes in a framework with inter-spousal bargaining. This strand of literature typically analyzes optimal policy taking the married unit as given and overlooks the effects polices may have on marital decisions. In fact, this has been the norm in part due to the beliefs that fiscal policies have little effects on marriage formations (Borella et al., 2017). However, the recent paper by Ilin et al. (2022) argue that the tax system, in particular marriage penalties, can explain a sizable portion of cross-state differences in marital patterns in the US. This suggests that the right framework for policy analysis should be one where the marriage market responds endogenously to policy changes. In this regard, Gayle and Shephard (2019) are the first to break from the norm by extending the Choo and Siow (2006) marital matching setup to study the optimal design of taxation in an environment where taxes influence who marries whom and distort intra-household allocations. Despite sharing the same spirit, this paper focuses on a life-cycle model with income risk rather than a static setup, thereby allowing marriages to play an important role in private insurance against income risks. In terms of modeling framework, this paper is closest to Ramakrishnan (2022) who also develops a life-cycle model with endogenous marriage but focuses on social security policy.

This project is also related to a burgeoning literature that studies various marriage-related policies. In response to idiosyncratic labor market shocks, spouses can insure each other by adjusting their labor supply decisions. Because unemployment insurance is a substitute for spousal insurance, this raises the question of whether unemployment insurance should be less generous to minimize crowding out of spousal insurance (Birinci, 2019; Haan and Prowse, 2020; De Nardi et al., 2020; Ortigueira and Siassi, 2013; Wu and Krueger, 2021). There are also questions concerning whether taxation should be marriage neutral. In the US, married couples are taxed differently from single individuals whereby the second earner's marginal tax rate is based on the income of the primary earner. To this end, there are studies that evaluate the effects of eliminating joint tax filing on labor supply and welfare (Bronson and Mazzocco, 2021; Guner et al., 2012). On the design of social security systems, Borella et al. (2023) studies the aggregate implications of marriage-related provisions. The framework developed in this paper can be applied to study the aforementioned questions in a unique setup whereby marriage is endogenous to policy.

Lastly, this paper joins an exciting direction in macroeconomics which incorporates families in the study of the aggregate economy and studies the economics of family from a macroeconomic perspective.[2]. This paper is particularly related to the literature that links family and marriages to income inequality (Fernández and Rogerson, 2001; Fernandez et al., 2005; Greenwood et al., 2014). The model can also be used to evaluate how changing macroeconomic environment may

[2]Doepke and Tertilt (2016) provides a thorough survey of this literature

affect household formation pattners as in Greenwood et al. (2016); Regalia et al. (2019); Santos and Weiss (2016).

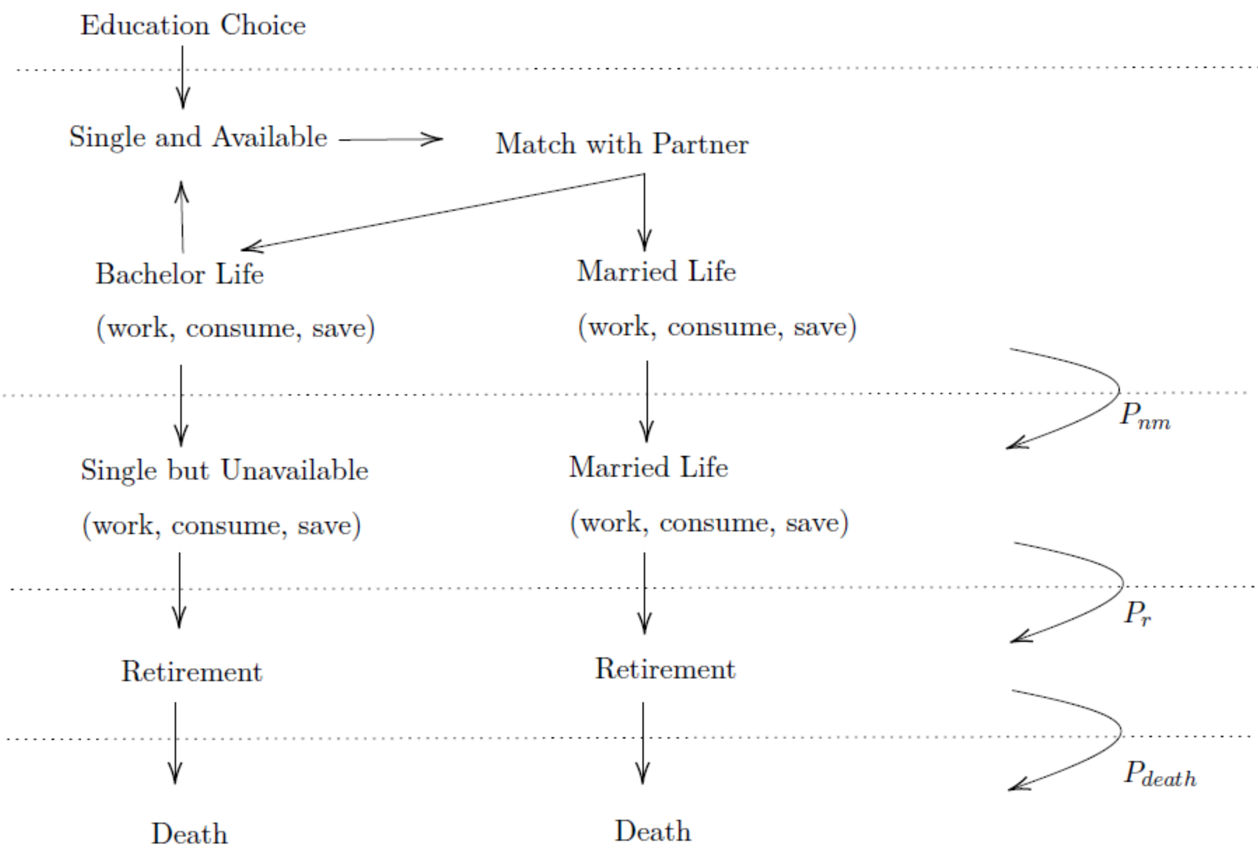
1.2 Model

To address the optimal design of redistribution policy, I build a quantitative model with three key ingredients. The backbone of the model is a life-cycle setup with idiosyncratic labor risks and incomplete markets. Every period, workers work and receive labor income which they then allocate between market consumption and saving. Departing from standard setups, the model features both single and married households. I assume a unitary model for married households in which spouses make decisions jointly and cooperatively. Furthermore, single households decide whether to get married and to whom endogenously in a matching market. Lastly, a government collects labor income taxes to finance redistribution, social security, and unemployment insurance.

1.2.1 Setup

The life-cycle nature is modeled with stochastic aging with four stages: education, working and matching, working and not matching, and retirement. At the beginning of life, agents are born with zero asset. Before entering the labor force, they make an educational choice which determines their skill type for the remainder of lifetime. Thereafter, they enter the working stage initially as single households. In the early phase of their working lives, single agents are randomly matched with another single agent of the opposite sexwhere marriages may occur if both parties agree. In the next stage, agents continue to work but no longer receive matches. In the last stage,

Figure 1.2: Model Timeline



they retire and eventually die. Upon death, a new agent is born so that the total mass of workers in the economy is stationary. Transition between states occur stochastically as depicted in Figure (1.2).[3]

The economy is populated by agents consisting of two genders $g \in \{M, F\}$, men and women, each with equal mass. In this model, gender matters for two reasons. As a simplyfing assumption, marriages are formed between two individuals from opposite genders.[4] Secondly,

[3]The probability of moving from the working and matching stage to the working but no matching state is P_{nm}. The probability of progressing to retirement is P_r and P_{death} is the probability of death.

[4]This is a simplifying assumption to keep modelling marriage easier. In the United States, same-sex marriages and partnerships make up a small fraction of coupled households. Based on the 2019 American Community Survey, same-sex couples including married and unmarried partnerships account for less than 1.5% of all coupled households (Walker and Taylor, 2021).

education cost and wage are gender-specific. The former is needed to match the different propensities to acquire a college education across genders in the data, whereas the latter captures the phenomenon of gender inequality in the labor market.

1.2.2 Education

There are two education levels $e \in \{L, H\}$ in the model, corresponding to less-than-college (L) and college (H). The education level determines the wage rate per labor efficiency unit. It also affects, both directly and indirectly, the marriage prospects of a worker which we discuss in later sections.

In the beginning of life, each agent is born with a utility cost of education ξ, which is drawn exogenously from a normal distribution $N(\mu_\xi^g, \sigma_{g,\xi}^2)$ that is allowed to vary by gender. This utility cost can be interpreted as encompassing financial resources, preferences, aptitude, and psychic costs among others. Agents can attain a college education by paying the utility cost. The education decision is permanent and irreversible, so that an agent that chooses not to go to college remain as non-college educated for the remainder of life.

Let $V_0^g(e)$ denote the expected lifetime value of having education level e, the education investment decision is a discrete choice problem

$$\max_{e \in \{H,L\}} \quad \{V_0^g(H) - \xi, V_0^g(L)\} \tag{1.1}$$

The optimal education acquisition decision is given by

$$h_g^*(\xi) = \begin{cases} H & \text{if } V_0^g(H) - V_0^g(L) > \xi \\ \\ L & \text{otherwise} \end{cases} \tag{1.2}$$

In other words, an agent will acquire a college education if the lifetime value of being a college worker net of the utility cost of education is higher than the lifetime value of not having a college education. Heterogeneity in education cost implies that not all agents of the same gender will acquire a college education. The fraction of agents of gender g who become college graduates is simply Pr $(\xi < V_0^g(H) - V_0^g(L))$ which can be easily computed given the Gaussian distributional assumption on ξ.

1.2.3 Working Stage

After making the education decision, agents draw an idiosyncratic labor productivity level $z_{i,0}$ and enter the labor market. Every period, they supply $h_{i,t}$ of labor and earn labor income. Labor income is subject to a proportional social security tax as well as a general labor income tax via a non-linear tax schedule. Furthermore, employment is not guaranteed as workers face exogenous unemployment risks. When unemployed, they receive an unemployment insurance payment from the government.

Labor Income The pre-tax labor income $y_{i,t}$ of an employed worker with education e and productivity $z_{i,t}$ supplying $h_{i,t}$ units of labor is given by

$$y_{i,t} = (\delta^g W^e)\, z_{i,t} h_{i,t} \tag{1.3}$$

where δ^g is the gender-specific shifter and W^e the wage rate education-specific wage rate per efficiency units of labor.

The gender-specific shifter δ^g captures the differential labor compensation facing male and female workers. Both historically and presently, female workers in the United States earn less than their male counterparts. As reported by Blau and Kahn (2017), female workers in 1980 earn only 79% the their male counterparts controlling for observables such as occupation and experience. The gender wage gap continues to persist although narrowing down to 92% in 2010. Because the reason behind the gender wage gap is beyond the scope of this study, I model the empirical pay inequality as a simple wage wedge across genders.

Workers face income risks through fluctuations in their idiosyncratic labor productivity $z_{i,t}$, which evolves stochastically according to an AR(1) process

$$\log\left(z_{i,t+1}\right) = \rho \log\left(z_{i,t}\right) + \sigma_t^e \epsilon_{i,t+1} \tag{1.4}$$

where $\epsilon \sim N(0,1)$ if employed and $\epsilon \sim N(D^e, 1)$ if unemployed.

The volatility of labor endowment shock depends on education level e to reflect the findings of Meghir and Pistaferri (2004) that high school educated workers face more volatile income risks. When employed, the innovation to productivity has mean zero. When unemployed, the productivity process exhibits a downward drift by D^e, capturing the fact that individuals that are out of work are more likely to experience a deterioration in their earning potential (Braxton et al., 2021).

Unemployment Unemployment risks vary by both education type and labor productivity. Recent studies using administrative micro data on income and employment find that workers with lower income are more likely to become unemployed or experience negatively skewed shocks to income (Braxton et al., 2021; Guvenen et al., 2014). In the model, the probability of becoming unemployed P_u (and the probability of being employed is simply $1 - P_u$) is given by a non-linear function

$$P_u^e(z) = \alpha_1^e \exp\left(-\alpha_2^e z\right) + \alpha_3^e \tag{1.5}$$

When $\alpha_1^e > 0$ and $\alpha_2^e > 0$, unemployment risk is decreasing in labor productivity in a non-linear fashion. In particular, at high productivity levels, unemployment risk is low and close to α_3^e. As productivity declines, the probability of becoming out of work rapidly increases. Note that this assumptions on the labor income process combined with the fact the downward drift in productivity during unemployment generate a scarring effect of unemployment, as job loss induce deterioration of labor productivity which further increases the future likelihood of unemployment. As such, low productivity workers are especially vulnerable to getting caught in a downward spiral of joblessness and low income, which is a key motivation for government redistribution.

During an unemployment spell, workers are entitled to an unemployment benefit that is equal to a fraction of the income of an employed worker of the same type (gender, education, and productivity). In particular, the government stipulates a replacement ratio for single (rep_S) and married households (rep_M) respectively. The income of an unemployed worker $y_{i,t}^u$ is therefore given by

$$y_{i,t}^u = \text{rep}\left(\delta^g W^e\right) z_{i,t} \tilde{h} \tag{1.6}$$

where $\tilde{h}$ is the average labor supply in the economy.[5]

[5] Indexing unemployment payout to the average labor supply in the economy is a simplifying assumption that

In the United States, unemployment insurance benefits have a term limit, with most states granting eligibility up to 26 weeks. To mimic this feature of the system, unemployment insurance in the model also does not last perpetually. A worker is only eligible for the benefit in the first period of an unemployment spell. If a worker is unemployed in consecutive periods, she loses eligibility to unemployment benefits in the second periods onwards.

A recent literature studies the design of UI policies by marital status (De Nardi et al. (2021), Birinci (2019), Bardóczy (2020)). A common theme is that public unemployment insurance is more valuable to single households over married households due to less tools for self-insurance. Furthermore, providing unemployment benefits to married households may crowd out intra-household insurance and thereby undermining the cost effectiveness of UI policy. The inclusion of UI as a policy in this model contributes to this literature by assessing the optimal design in a framework with endogenous marriage formation which is missing in the existing literature.

Social Security Tax and Pension Benefits The government levies a proportional social security tax (τ_{ss}) on labor income to finance a pay-as-you-go pension system. The revenue from the social security tax is used to finance pension payments to retired households, the level of which depends both on marital status

In the United States, social security benefits are based on the workers' lifetime earnings. To avoid the computational complexity of introducing an additional state variable, pension benefits in the model are dependent on educational level (which proxies for lifetime earnings) and are allowed to differ by marital status. Retired singles with education e are eligible to a lump-sum payment of PEN_S^e per period. Retired married households with education levels (e_m, e_f) are entitled to $\text{PEN}_M^{e_m e_f}$.

makes solving the model easier, as it avoids the need to iterate on the type specific labor supply.

Borella et al. (2023) studies the marriage-related provisions of the US social security system and finds that the current rules reduce the participation of married women over the life cycle, the participation of men after age 55, and the savings of married couples. The main reason is that the secondary earner in a marriage is entitled to half of the primary earner's pension benefits and the full amount when the primary earner dies. In our model, workers do not internalize the consequences of their labor supply decisions on the amount of pension payments during retirement. Therefore, the channels of Borella et al. (2023) is muted. In contrast, social security in this model is mostly a potential tool to encourage marriages and manipulate savings behavior of workers.

Labor Income Tax To finance government spending as well as unemployment insurance, the government levies an income tax on labor earnings. Following Heathcote et al. (2017), labor income tax is levied via a non-linear tax schedule. Specifically, given pre-tax income y, after-tax income $T(y)$ is given by the following functional form

$$T(y) = (1 + \chi)\, y^{1-\tau} \tag{1.7}$$

This functional form is adopted because the literature has found that it provides a good empirical fit to the tax and transfer system in the United States (Borella et al., 2022; Heathcote et al., 2017; Wu, 2021). In addition, it allows me to borrow existing estimates of the US tax system for the calibration. With this non-linear tax schedule, the average tax rate is $\frac{y-T(y)}{y} = 1 - (1+\chi)\, y^{-\tau}$ and the marginal tax rate is $1 - (1-\tau)(1+\chi)\, y^{-\tau}$. When $\tau = 0$, labor income tax is flat with a constant marginal tax rate $-\chi$ that is also equal to the average tax rate. When $\tau \in (0, 1]$, the tax system is progressive – such that the average tax rate is larger than the marginal tax rate. In other words, the average tax rate is increasing in the level of income. In contrast, the tax system features regressivity when $\tau < 0$. This can be seen in Figure (1.3).

The parameter χ controls the break-even income level $y^{BR} = (1+\chi)^{\frac{1}{\tau}}$, which is defined as the level of income below at which pre-tax income is equal to post-tax income (equivalently, at which the average tax rate is zero). As depicted in Figure (1.4), when pre-tax income is below y^{BR}, the tax schedule entails a positive net transfer (or negative tax rate). In the other case when pre-tax income is above y^{BR}, tax obligation is positive. Therefore, this tax rule permits the possibility of positive net transfers for workers with sufficiently low income, providing redistribution from the rich to the poor.

Labor income tax is at the household level rather than at the individual level, albeit the tax schedule is specific to marital status. Allowing the policy parameters (χ, μ) to depend on marital status open the door to flexibly vary the marriage "penalty" and "bonus" at different levels of income.

Consumption Workers consume a final consumption bundle that is a CES aggregator of a market good $\mathcal{C}$ as well as a home production good $\mathcal{H}$

$$C = \left[\alpha_{\mathcal{C}}\mathcal{C}^{\theta} + \alpha_{\mathcal{H}}\mathcal{H}^{\theta}\right]^{\frac{1}{\theta}} \tag{1.8}$$

The home production good is produced with a market component (d) that is purchased and time spent on home production (q), also via a CES technology

$$\mathcal{H} = \left[\alpha_d d^{\lambda} + \alpha_q q^{\lambda}\right]^{\frac{1}{\lambda}} \tag{1.9}$$

Figure 1.3: Average Tax Rates for Different Values of τ, $\chi = -0.1$

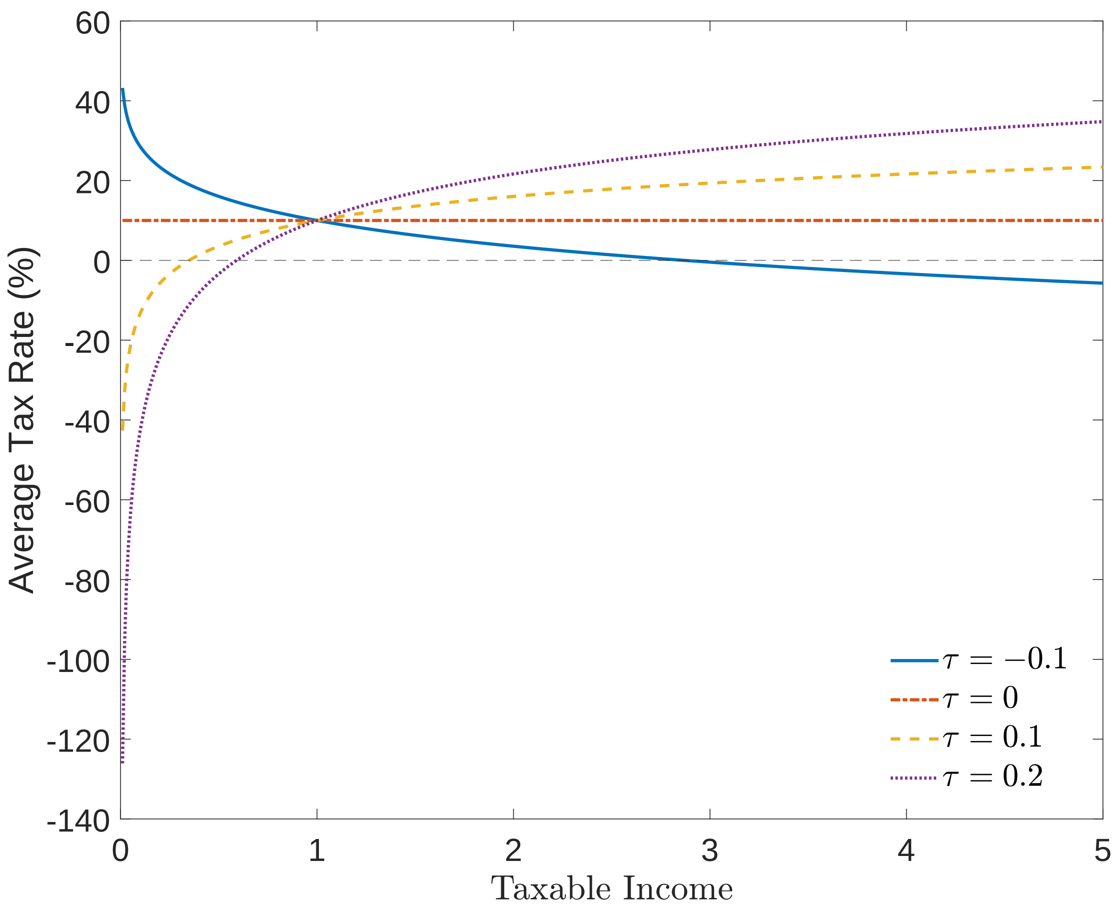
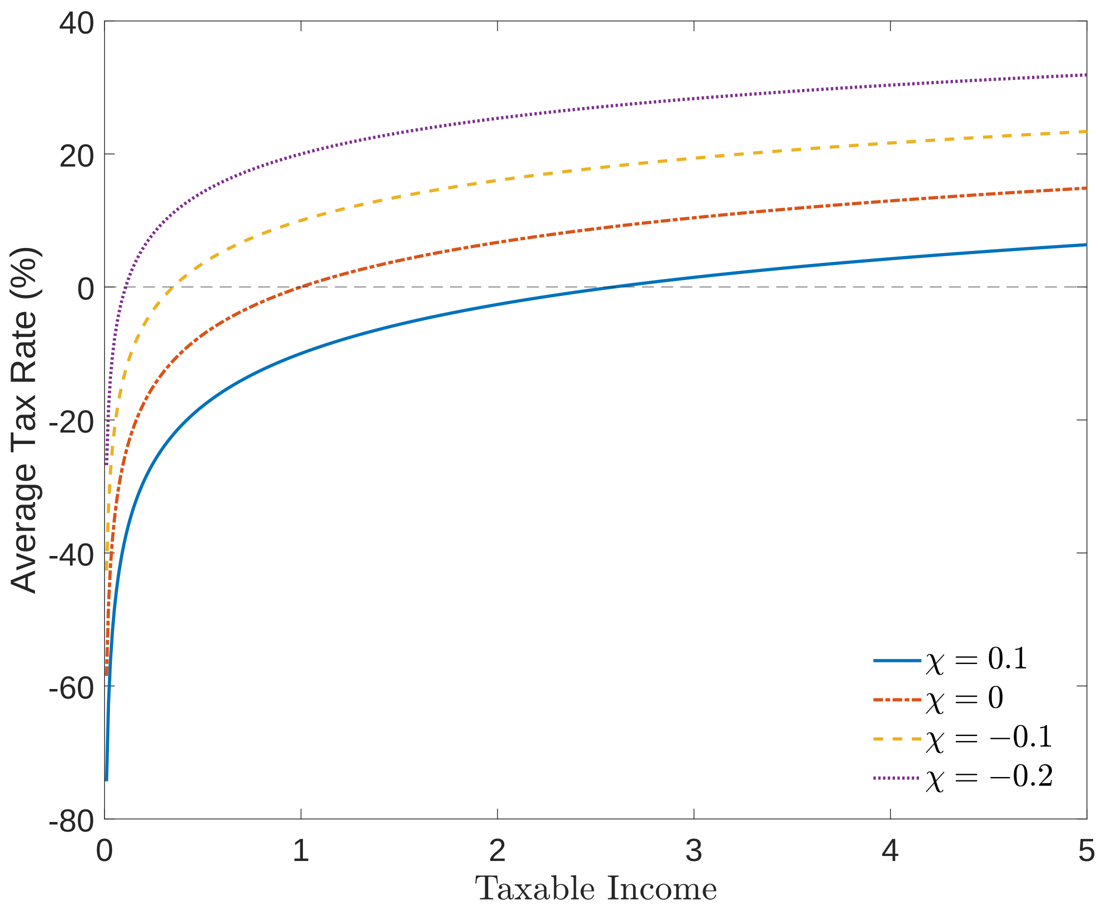


Figure 1.4: Average Tax Rates for Different Values of χ, $\tau = 0.1$



For married couples, consumption is adjusted by an equivalence scale $\zeta < 2$ to capture increasing returns to consumption in a joint household, in the sense that married couples require less than twice the consumption as a single household in order to achieve the same level of utility.

Flow Utility Single workers derive utility from consumption C but dislike market work h and time spent on home production q. The flow utility is separable in consumption and leisure as given below

$$U_S(C,h,q) = \frac{C^{1-\sigma}}{1-\sigma} - f\frac{(h+q)^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} \tag{1.10}$$

Similarly, married couples enjoy consumption and leisure.

$$U_M(C,h^m,h^f,q^m,q^f) = \frac{C^{1-\sigma}}{1-\sigma} - \frac{f}{2}\frac{(h^m+q^m)^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} - \frac{f}{2}\frac{(h^f+q^f)^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} \tag{1.11}$$

The parameter σ governs the intertemporal elasticity of substitution of consumption and also the degree of risk aversion. The parameter γ is the Frisch elasticity of labor supply which determines the responsiveness of labor supply to changes in wages. Workers enjoy leisure and dislike time spent on market work and home production, which is captured by a disutility of work parameter $f > 0$. Notice that to keep the level of the value of leisure comparable across single and married households, the disutility of work f is divided by two in the flow utility function of couples.

In order to match the differential marriage rates by education types, women suffer a utility cost $\mathcal{J}$ when their husbands are out of work. This is consistent with survey evidence that one of the most valued characteristics in a spouse is stable employment (Edin and Reed, 2005) and empirical evidence that deterioration of men labor market opportunities on the marriage rate

(Autor et al., 2019).

Saving Technology Households may save through a one-period risk-free bond b that pays a non-taxable return r. Borrowing is not allowed such that $a \geq 0$ at all times. Households save for two major reasons. Obviously, preparation for retirement is a reason for saving. In addition, the absence of complete markets implies that households have a precautionary motive to save in order to self-insure against adverse income shocks and smooth consumption over the life-cycle. Government redistribution and intra-family insurance represent alternative sources of self-insurance as will be discussed in later sections.

Problem of Single Households The problem of a single worker can be written recursively as followed

$$V_S\left(e, u_{i,t}, z_{i,t}, a_{i,t}\right) = \max_{h_{i,t}, q_{i,t}, l_{i,t}, d_{i,t}, a_{i,t+1}} \quad U_S\left(C_{i,t}, h_{i,t}, q_{i,t}\right) + \beta \mathbb{E}\left[V(e, u_{i,t+1}, z_{i,t+1}, a_{i,t+1})\right] \tag{1.12}$$

subject to

$$
\begin{aligned}
C_{i,t} &= \left[\alpha_{\mathcal{C}}\mathcal{C}_{i,t}^{\theta} + \alpha_{\mathcal{H}}\mathcal{H}_{i,t}^{\theta}\right]^{\frac{1}{\theta}} && (1.13)\\
\mathcal{H}_{i,t} &= \left[\alpha_d \cdot d_{i,t}^{\lambda} + \alpha_q \cdot q_{i,t}^{\lambda}\right]^{\frac{1}{\lambda}} && (1.14)\\
1 &= q_{i,t} + h_{i,t} + l_{i,t} && (1.15)\\
h_{i,t} &\leq \bar{h} && (1.16)\\
a_{i,t+1} &= (1+r)a_{i,t} + (1+\chi_S)\, y_{i,t}^{1-\tau_S} - p_d d_{i,t} - p_{\mathcal{C}}\mathcal{C}_{i,t} && (1.17)\\
a_{i,t+1} &\geq 0 && (1.18)\\
y_{i,t} &= \begin{cases} (1-\tau_{ss})\,(\delta^g W^e)\, z_{i,t}h_{i,t} & \text{if } u_{i,t}=0\\ \text{rep}_S\,(\delta^g W^e)\, z_{i,t}\tilde{h} & \text{if } u_{i,t}=1\\ 0 & \text{if } u_{i,t}=2 \end{cases} && (1.19)
\end{aligned}
$$

In a given period, a single household chooses how to allocate its time endowment (normalized to one) between market work (h), home production (q), and leisure (l), and how much to save and consume to maximize the sum of flow utility and the expected continuation value. An employed worker ($u = 0$) can supply labor up to an upper bound $\bar{h}$. An unemployed worker ($u > 0$) cannot work and is constrained to $h = 0$. Unemployed workers that are eligible to unemployment insurance ($u = 1$) receives a benefit and those that do not ($u = 2$) have no labor income. Labor income, net of social security contribution, is taxed according to the non-linear tax schedule. Note that unemployment benefit is also taxable as is the case in the United States tax code. Simultaneously, the worker allocates cash-on-hand between market consumption ($\mathcal{C}$) which costs $p_{\mathcal{C}}$, the home production good (d) which costs p_d, as well as saving.

For a single household no longer receiving matches, the expected value is written as

$$\mathbb{E}V_{S,nm} = (1 - P_r) \int_{u_{i,t+1}} \int_{z_{i,t+1}} V_{S,nm}\left(e, u_{i,t+1}, z_{i,t+1}, a_{i,t+1}\right)$$

$$dF\left(u_{i,t+1}|e, u_{i,t}, z_{i,t}\right) dP\left(z_{i,t+1}|e, u_{i,t}, z_{i,t}\right)$$

$$+ P_r V_{S,R}\left(e, a_{i,t+1}\right) \quad (1.20)$$

For single households that are still in the matching phase, expectation of future value is taken over marital prospects in additional to future productivity and employment. The role of the marriage market will be discussed shortly and the formal expected continuation value will be deferred until then.

Gains of Marriage and the Problem of Married Households Couples behave cooperatively to maximize joint utility. There are several differences in a couple's problem as compared to the problem of a single household. First of all, couples enjoy increasing returns to consumption and home production, due to the fact that they have twice the time endowments but require less consumption per person compared to a single household. Secondly, couples can engage in specialization of tasks. For example, the spouse with higher earning power can spend more time in market work while the other spouse spends more time on home production. Thirdly, couples have to ability to insure each other against idiosyncratic labor income shocks. Intra-household insurance arise passively as idiosyncratic shocks are uncorrelated between spouses.[6] Workers can actively adjust their labor supply in response to unemployment shocks or adverse labor income

[6] Gorbachev (2016) find that the correlation in income growth between spouses is 0 until the early 2000's, after which it has increased to 0.1. In terms of correlation in unemployment incidence, Shore and Sinai (2010) report that the fraction of couples for which both partners are unemployed at the same time is small, a finding that is echoed by Birinci (2019).

shock of their spouses to mitigate the loss in earnings. Lastly, married households can pool financial resources and save jointly. The aforementioned summarize the gains from marriage in the model.

Written recursively, the problem of a married household is

$$V_M\left(e^m, e^f, u^m_{i,t}, u^f_{i,t}, z^m_{i,t}, z^f_{i,t}, a_{i,t}\right) = \max_{h^j_{i,t}, q^j_{i,t}, l^j_{i,t}, d_{i,t}, a_{i,t+1}} \quad U_M\left(C_{i,t}, h^m_{i,t}, h^f_{i,t}, q^m_{i,t}, q^f_{i,t}\right)$$
$$+ \beta\mathbb{E}\left[V_M\left(e^m, e^f, u^m_{i,t+1}, u^f_{i,t+1}, z^m_{i,t+1}, z^f_{i,t+1}, a_{i,t+1}\right)\right] \quad (1.21)$$

$$C_{i,t} = \left[\alpha_{\mathcal{C}}\left(\frac{\mathcal{C}_{i,t}}{\zeta_{\mathcal{C}}}\right)^{\theta} + \alpha_{\mathcal{H}}\left(\frac{\mathcal{H}_{i,t}}{\zeta_{\mathcal{H}}}\right)^{\theta}\right]^{\frac{1}{\theta}} \quad (1.22)$$

$$\mathcal{H}_{i,t} = \left[\alpha_d \cdot d^{\lambda}_{i,t} + \alpha_q \cdot \left(q^m_{i,t} + q^f_{i,t}\right)^{\lambda}\right]^{\frac{1}{\lambda}} \quad (1.23)$$

$$1 = q^j_{i,t} + h^j_{i,t} + l^j_{i,t} \quad \text{for } j \in \{m, f\} \quad (1.24)$$

$$h^j_{i,t} \leq \bar{h} \quad \text{for } j \in \{m, f\} \quad (1.25)$$

$$a_{i,t+1} = (1+r)a_{i,t} + T(y^m_{i,t} + y^f_{i,t}) - p_d d_{i,t} - p_{\mathcal{C}}\mathcal{C}_{i,t} \quad (1.26)$$

$$a_{i,t+1} \geq 0 \quad (1.27)$$

$$T(y) = (1+\chi_M)\, y^{1-\tau_M} \quad (1.28)$$

$$y^j_{i,t} = \begin{cases} (1-\tau_{ss})\left(\delta^j W^e\right) z_{i,t} h_{i,t} & \text{if } u^j_{i,t} = 0 \\ \text{rep}_M\left(\delta^j W^e\right) z_{i,t}\tilde{h} & \text{if } u^j_{i,t} = 1 \\ 0 & \text{if } u^j_{i,t} = 2 \quad \text{for } j \in \{m, f\} \end{cases} \quad (1.29)$$

The expectation term for a married household in the "matching" phase is

$$\mathbb{E}V_{M,m} = (1 - P_r) \int\int\int\int_{u^m_{i,t+1}, u^f_{i,t+1}, z^m_{i,t+1}, z^m_{i,t+1}} V_{M,m} \left(e^m, e^f, u^m_{i,t+1}, u^f_{i,t+1}, z^m_{i,t+1}, z^m_{i,t+1}, a_{i,t+1}\right)$$
$$dF\left(u^m_{i,t+1}|e^m, u^m_{i,t}, z^m_{i,t}\right) dF\left(u^f_{i,t+1}|e^f, u^f_{i,t}, z^f_{i,t}\right) dP\left(z^m_{i,t+1}|e^m, u^m_{i,t}, z^m_{i,t}\right) dP\left(z^f_{i,t+1}|e^f, u^f_{i,t}, z^f_{i,t}\right)$$
$$+ (P_r) \int\int\int\int_{u^m_{i,t+1}, u^f_{i,t+1}, z^m_{i,t+1}, z^m_{i,t+1}} V_{M,nm} \left(e^m, e^f, u^m_{i,t+1}, u^f_{i,t+1}, z^m_{i,t+1}, z^m_{i,t+1}, a_{i,t+1}\right)$$
$$dF\left(u^m_{i,t+1}|e^m, u^m_{i,t}, z^m_{i,t}\right) dF\left(u^f_{i,t+1}|e^f, u^f_{i,t}, z^f_{i,t}\right) dP\left(z^m_{i,t+1}|e^f, u^f_{i,t}, z^f_{i,t}\right) dP\left(z^f_{i,t+1}|e^f, u^f_{i,t}, z^f_{i,t}\right)$$
$$(1.30)$$

And the expectation term for a married household in the "non-matching" phase is

$$\mathbb{E}V_{M,nm} = (1 - P_r) \int\int\int\int_{u^m_{i,t+1}, u^f_{i,t+1}, z^m_{i,t+1}, z^m_{i,t+1}} V_{M,nm} \left(e^m, e^f, u^m_{i,t+1}, u^f_{i,t+1}, z^m_{i,t+1}, z^m_{i,t+1}, a_{i,t+1}\right)$$
$$dF\left(u^m_{i,t+1}|e^m, u^m_{i,t}, z^m_{i,t}\right) dF\left(u^f_{i,t+1}|e^f, u^f_{i,t}, z^f_{i,t}\right) dP\left(z^m_{i,t+1}|e^m, u^m_{i,t}, z^m_{i,t}\right) dP\left(z^f_{i,t+1}|e^f, u^f_{i,t}, z^f_{i,t}\right)$$
$$+ (P_r) V_{M,r}\left(e^m, e^f, a_{i,t+1}\right) \quad (1.31)$$

where $dP(z'|e, u, z)$ and $dF(u'|e, u, z)$ are the probability densities of future productivity and employment status conditional on current productivity, education, and employment status.[7]

1.2.3.1 Matching

Having described the standard problem pertaining to the household, I now turn to the marriage market, which is largely modeled after macroeconomic models featuring endogenous marriages (Greenwood et al., 2016; Regalia et al., 2019; Santos and Weiss, 2016). At the beginning

[7] Recall that employment status includes three states: employed, unemployed and eligible for UI, and unemployed without UI eligibility. The transition from being eligible for UI to not depends whether or not the worker is currently unemployed.

of each period, single workers still in the matching phase meet another single worker of the opposite sex at random. The matched parties observe the idiosyncratic states of each other and decide if they want to form a joint household. A marriage is formed if and only if both parties agree – that is if both parties independently find marrying each other more attractive than remaining single. If both sides agree to marry, the newlyweds combine their assets and live as a unitary household permanently until death. Because utility is assumed to be non-transferable – meaning that one party cannot convince the other to agree to a marriage by making any transfers at the time of meeting nor during marriage.

In addition to economic motives due to the various gains from marriage, workers also marry for non-pecuniary reasons. Upon meeting, each worker draw a love shock x_M for the match, representing how attractive the match is aside from economic factors. The love shock is expressed in terms of utility and is drawn from a Logistic distribution with scale parameter $\mathcal{L}$ which determines the variance of the love shocks. The mean of the love shock distribution is $\mathcal{M}$ for all matches except for matches where both parties have a college education, for which the mean is $\mathcal{M}_{CC}$. This allows for the possibility that college graduates enjoy each other's company more and is essential for matching the observed degree of assortative mating by education in the data.

Formally, a marriage occurs if both conditions below hold

$$V_M^m\left(e^m, e^f, u^m, u^f, z^m, z^f, a^m + a^f\right) + x_M^m \geq V_S^m\left(e^m, u^m, z^m, a^m\right) \tag{1.32}$$

$$V_M^f\left(e^m, e^f, u^m, u^f, z^m, z^f, a^m + a^f\right) + x_M^f \geq V_S^f\left(e^f, u^f, z^f, a^f\right) \tag{1.33}$$

When two singles meet, they can choose to marry this person or wait for the next match.

The option value of waiting for the new match that is potentially more attractive implies that a worker may turn down a current match even if getting married results in static gains. Note that under this setup of search and matching friction alongside non-transferable utility means that the marriage market equilibrium is not necessarily efficient.

With the distributional assumption on x_M, the probability that party i would agree to a marriage in a match $e^m, e^f, u^m, u^f, z^m, z^f, a^m, a^f$

$$\Gamma_i\left(e^m, e^f, u^m, u^f, z^m, z^f, a^m, a^f\right) = \frac{\exp\left(\frac{V_M^i(\cdot)}{\mathcal{L}}\right)}{\exp\left(\frac{V_M^i(\cdot)}{\mathcal{L}}\right) + \exp\left(\frac{V_S^i(\cdot)}{\mathcal{L}}\right)} \tag{1.34}$$

Having described the marriage problem, the expected continuation values of single households can be spelled out explicitly. Let $\mathbb{I}\left(e^i, e^j, u^i, u^j, z^i, z^j, a^i, a^j, x_M^i\right)$ be an indicator function that takes value of one if $V_M^i\left(e^i, e^j, u^i, u^j, z^i, z^j, a^i, a^j\right) + x_M^i \geq V_S^i(e^i, u^i, z^i, a^i)$, and let $\mu(e, u, z, a)$ denote the distribution of available singles. The expected continuation value of a single matching worker is

$$\begin{aligned}
\mathbb{E}V_{S,m} = (1 - P_{nm}) \int_{z^i_{i,t+1}} \int_{u^i_{i,t+1}|} \int_{(e^j,u^j,z^j,a^j)} \int_{x_M} & \left\{\mathbb{I}\left(e^i, e^j, u^i, u^j, z^i, z^j, a^i, a^j, x_M^i\right) \cdot \Gamma_j \right. \\
& \left. \left[V_{M,m}\left(e^i, e^j, u^i_{i,t+1}, u^j_{i,t+1}, z^i_{i,t+1}, z^j_{i,t+1}, a_{i,t+1}\right) + x_M^i\right]\right\} + \\
& \left\{\left(1 - \mathbb{I}\left(e^i, e^j, u^i, u^j, z^i, z^j, a^i, a^j, x_M^i\right)\right) \cdot (1 - \Gamma_j) \ \left[V_{S,m}\left(e^i, u^i_{i,t+1}, z^i_{i,t+1}, a_{i,t+1}\right)\right]\right\} \\
& dG\left(x_M^i\right) d\mu\left(e^j, u^j, z^j, a^j\right) dF\left(u^i_{i,t+1}|e^i, u^i_{i,t}, z^i_{i,t}\right) dP\left(z^i_{i,t+1}|e^i, u^i_{i,t}, z^i_{i,t}\right) \\
+(P_{nm}) \int_{z^i_{i,t+1}} \int_{u^i_{i,t+1}|} & V_{S,nm}\left(e^i, u^i_{i,t+1}, z^i_{i,t+1}, a_{i,t+1}\right) dF\left(u^i_{i,t+1}|e^i, u^i_{i,t}, z^i_{i,t}\right) dP\left(z^i_{i,t+1}|e^i, u^i_{i,t}, z^i_{i,t}\right)
\end{aligned} \tag{1.35}$$

1.2.4 Retirement

Retirees stop working and receive a retirement benefit payments every period until they die. They allocate time between home production and leisure, and chooses how much to consume and save. Retirees survives onto the next period with probability $1 - P_{death}$. Upon death, remaining assets are transferred to the government which implies an effective estate tax rate of 100%. For simplicity, married couples die at the same time.

Formally, the problem of a single retiree is

$$V_{S,R}\left(e, a_{i,t}\right) = \max_{q_{i,t}, l_{i,t}, d_{i,t}, a_{i,t+1}} \quad U_S\left(C_{i,t}, h_t, q_t\right) + \beta P_{\text{death}} V_{S,R}\left(e, a_{i,t+1}\right) \tag{1.36}$$

subject to

$$C_{i,t} = \left[\alpha_{\mathcal{C}} \mathcal{C}_{i,t}^{\theta} + \alpha_{\mathcal{H}} \left(\mathcal{H}_{i,t}\right)^{\theta}\right]^{\frac{1}{\theta}} \tag{1.37}$$

$$1 = q_{i,t} + l_{i,t} \tag{1.38}$$

$$a_{i,t+1} = (1+r)a_{i,t} + \text{PEN}_S^e - p_d d - p_{\mathcal{C}} \mathcal{C} \tag{1.39}$$

$$a_{i,t+1} \geq 0 \tag{1.40}$$

And the problem of a married retiree household is

$$V_{M,R}\left(e^m, e^f, a_{i,t}\right) = \max_{q_{i,t}^m, q_{i,t}^f, l_{i,t}^m, l_{i,t}^f, d_{i,t}, a_{i,t+1}} \quad U\left(C_{i,t}, q_{i,t}^m, q_{i,t}^f\right) + \beta\left(1 - P_{death}\right) V_{M,R}\left(e^m, e^f, a_{i,t+1}\right) \tag{1.41}$$

subject to

$$C = \left[\alpha_{\mathcal{C}}\left(\frac{\mathcal{C}_{i,t}}{\zeta_{\mathcal{C}}}\right)^{\theta} + \alpha_{\mathcal{H}}\left(\frac{\mathcal{H}_{i,t}}{\zeta_{\mathcal{H}}}\right)^{\theta}\right]^{\frac{1}{\theta}} \quad (1.42)$$

$$1 = q^{j}_{i,t} + l^{j}_{i,t} \quad \text{for } j \in \{m, f\} \quad (1.43)$$

$$a_{i,t+1} = (1+r)a_{i,t} + \text{PEN}_{M}^{e^{m}e^{f}} - p_d d - p_{\mathcal{C}}\mathcal{C} \quad (1.44)$$

$$a_{i,t+1} \geq 0 \quad (1.45)$$

1.3 Calibration

The model is calibrated to match the data in 1980 and 2015 in the stationary equilibrium, most notably demographic patterns related to marriage. Most parameters are time invariant, while those related to tax policy and wages may change over time to reflect observed shifts in the data. In this section, I describe the calibration strategy.

1.3.1 External Calibration

Externally calibrated parameters are summarized in Table (3.1). Parameters related to government policies for 1980 and 2015 are separately summarized in Tables (1.2) and (1.3).

Transition Probabilities A model period is one year. The matching phase, during which singles receive matches, lasts 25 periods on average. This corresponds to ages 20 to 45. The non-matching phase where singles no longer receive matches lasts 20 periods, corresponding to ages 45 to 65. Retirement lasts 10 periods on average reflecting a life expectancy of 75 years. This implies setting $P_{nm} = \frac{1}{25}$, $P_r = \frac{1}{20}$, and $P_{death} = \frac{1}{10}$.

Gender Wage Gap In the model, men and women are identical in the labor market except for the wage that they receive. As such, it is appropriate to use the conditional gender wage gap from the data to discipline the size of this wage penalty, rather than the unconditional gap which reflects the outcome of endogenous decisions such as occupational sorting. Blau and Kahn (2017) report that conditional on various factors such as education, occupation, and experience, full-time female workers earn only 79% of their male counterparts in 1980. Although this gap has narrowed down over the last few decades, an unexplained gap of 8% persisted in 2015. Hence, δ^f is set to 0.79 in 1980 and 0.92 in 2010.

Preferences The intertemporal elasticity of substitution (σ) and Frisch elasticity of labor supply γ are 1.894 and 0.53 respectively, taken from estimates of Blundell et al. (2016) who use the same preference specification. The Frisch labor supply elasticity is assumed to be the same across genders as well as marital statuses. Empirical evidence suggests that labor supply of married women is more sensitive to wages, hence this assumption dampens the their responsiveness to changes in wages and taxes.

With respect to the consumption aggregators, θ which pins down the elasticity of substitution between market consumption and the home production good (θ) is set to 0.43. λ governing the substitutability between time spent on home production and market purchases for the home production good, is set to 0.19. Both parameters are borrowed from the estimates of McGrattan et al. (1997). This calibration implies a relatively high degree of substitutability between market and home goods. In contrast, home production time and purchases have a smaller elasticity of substitution and the aggregator to closer to Cobb-Douglas. While the elasticties of substitution are externally calibrated, the weights $(\alpha_c, \alpha_d, \alpha_q, \alpha_d)$ are internally calibrated to match the ratio of time spent on market work to home production and the expenditure share on home production good.

According to the OECD-modified equivalence scale, ζ_C and ζ_H are set to 1.5. This means that a married household only need to consume 50% more in order to achieve the same standard of living as a single household, yielding sizable increasing returns of scale in consumption.

Income Process and Unemployment Risk I follow Krueger and Ludwig (2016) and adopt estimates from Karahan and Ozkan (2013) to discipline the AR(1) income processes for the two education types using the Panel Study of Income Dynamics (PSID) data between 1968 and 1997. The persistence of the autoregressive income process is 0.95, reflecting highly persistent shocks rather than transitory ones. The standard deviations of permanent income shocks are 0.10 and 0.14 for college and non-college educated workers respectively. The calibration of the income process is in line with with more recent estimates based on administrative data (Guvenen et al., 2022). The fact that less skilled workers are subject to larger income risks is consistent with Meghir and Pistaferri (2004).

Parameters governing unemployment risk $(\alpha_1^e, \alpha_2^e, \alpha_3^e)$ are disciplined by the findings of Braxton et al. (2021) who document a non-linear relationship between income and the probability of job loss. In particular, they find that workers with higher permanent earnings have low unemployment risk around 3%. As permanent income decreases, the likelihood of becoming unemployed exponentially increases with the gradient being steeper for workers with less than college education as compared to those with a college degree. Figure (1.5) plots the probability of unemployment as a function of labor productivity and education. Lastly, the drift in labor productivity when out of work (D^e) are taken from the same study, which are -0.142 and -0.11 for workers without and with a college education respectively.

Policy Parameters

The unemployment insurance replacement ratio in 1980 is set to the average replacement ratio of 30% in the United States during 1980-1987 as reported by Nickell et al. (2005), which is computed as the fraction of lost income replaced by UI in the first year of unemployment averaged across three family types (single, single-earner, and dual-earner households). Without major overhaul of the unemployment insurance system, I use the same replacement ratio in 2015.[8]

The Social Security Amendment of 1977 was passed with the goal of anchoring the target replacement at 43%. The implementation has been successful as the effective replacement rate for

[8]The average UI replacement rate reported in the OECD tax-benefit indicators database is 29% in 2015.

Figure 1.5: Unemployment Risk by Labor Productivity and Education

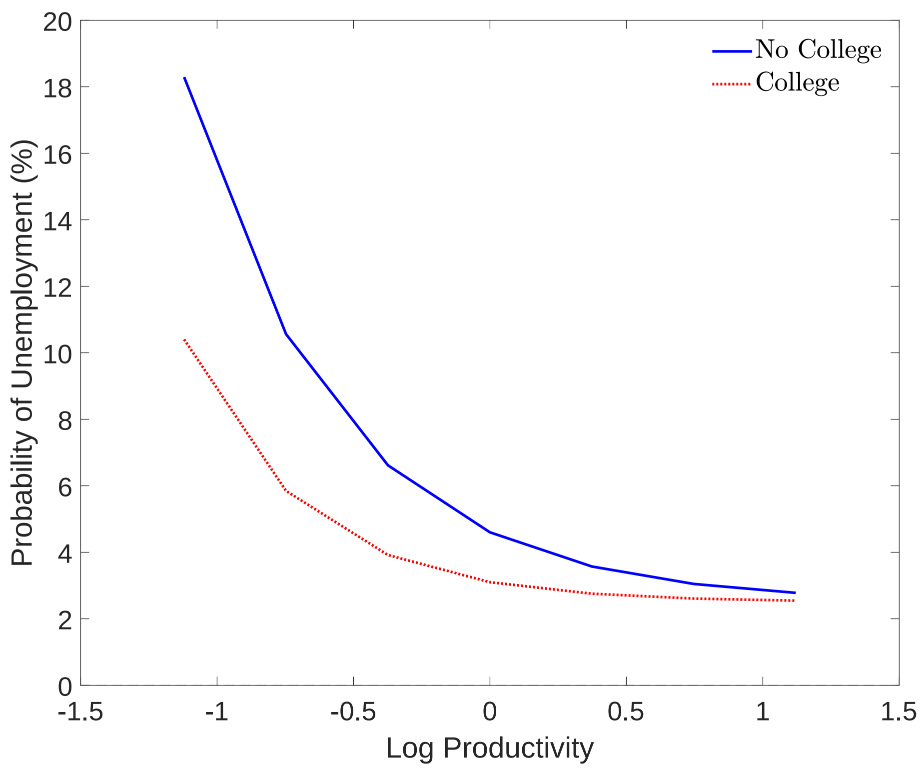


Table 1.1: External Calibration

Parameter	Description	Value	Source
r	Interest Rate	0.02	real risk-free rate
p_c	market good price	1	normalization
p_d	home good price	1	normalization
P_{nm}	transition prob. to non-matching	$1/25$	age 20-45
P_r	transition prob. to retirement	$1/20$	age 45-65
P_{death}	prob. of death	$1/10$	life expectancy of 75
ζ_C	Returns to scale in cons.	1.5	OECD-modified equivalence scale
ζ_H	Returns to scale in cons.	1.5	OECD-modified equivalence scale
ζ	IES consumption	1.894	Wu and Krueger (2021)
γ	Frisch elasticity	0.53	Wu and Krueger (2021)
α_q	weight on home prod. time	0.75	$1 - \alpha_d$
λ	e.o.s btwn mkt & home prod	0.43	McGrattan et al. (1997)
α_h	weight on home good	0.48	$1 - \alpha_c$
θ	e.o.s btwn mkt & home good	0.19	McGrattan et al. (1997)
ρ_z^H	labor prod. persist.	0.96	Krueger and Ludwig (2016)
ρ_z^L	labor prod. persist.	0.96	Krueger and Ludwig (2016)
σ_L	labor prod. shock	0.14	Krueger and Ludwig (2016)
σ_H	labor prod. shock	0.10	Krueger and Ludwig (2016)
δ_f^{1980}	1980 female wage gap	0.79	Blau and Kahn (2017)
δ_f^{2015}	2015 female wage gap	0.92	Blau and Kahn (2017)
α_1^L	unemp. risk (no college)	0.021	Braxton et al. (2021)
α_2^L	unemp. risk (no college)	1.8	Braxton et al. (2021)
α_3^L	unemp. risk (no college)	0.025	Braxton et al. (2021)
α_1^H	unemp. risk (college)	0.006	Braxton et al. (2021)
α_2^H	unemp. risk (college)	2.3	Braxton et al. (2021)
α_3^H	unemp. risk (college)	0.025	Braxton et al. (2021)
D_L	Productivity drift	-0.142	Braxton et al. (2021)
D_H	Productivity drift	-0.11	Braxton et al. (2021)

the median workers has since been kept stable around 40% according to the US Social Security Administration (SSA) Database. In the stylized social security setup of the model, the pension for single household is set to 40% of the average wage of workers with the same education level. The pension benefit of a married couple where both spouses have no college education is set to 40% of the average low-skill wage for each person. For married couples where both spouses have a college education, the pension benefit is set to 40% of the average high-skill wage for each person. For married couples where one spouse is college educated and the other is not, pension benefits is 150% of 40% of the average high-skill wage, reflecting the provision which entitles the secondary earner to the primary earner's level of pension benefits. These replacement rates do not change across 1980 and 2015. The social security tax rate τ_{ss} are chosen such that total revenue from the tax is equal to total pension payouts. In the model, this requires $\tau_{ss} = 0.09$ in both years, which is close to the statutory rate of 10.5% combining employee and employer contributions.

Turning to the non-linear labor income tax schedule, Borella et al. (2022) provide year-to-year estimates of the US federal effective tax function separately for single and married households using PSID data. In 1980, the estimated tax progressivity (τ) is 0.08 for single households and 0.12 for married households. Since then, the progressivity for married household has declined throughout the first half of the 1980's which has then increased until returning to 0.12 in 2015. The single progressivity has been more stable over the years with τ_S reaching 0.075 in 2015. These trends mirrors the findings in Heathcote et al. (2020) who show that progressivity in federal taxes has remained stable between the period 1979-1983 and 2012-2016, though they report a higher value of $\tau = 0.186$.[9] The level parameters (χ_S, χ_M) are chosen in the model to

[9]Other studies document a decline in the US tax progressivity. Using the Current Population Survey Annual

match the average tax rate paid by the median single and married households in the model and the data. In 1980, the median single household had an average tax rate of 8% and it is 12% for the median married household. For both types of household, the average tax rate has declined over the year but more so for married households with the tax codes becoming more favorable to married couples. In 2015, the same numbers decreased to 6% and 8.5% respectively. Although the calibration is chosen to match the average tax rate in the model and the data, the marginal tax rate for the median household implied by the calibrated tax function is very close to what Borella et al. (2022) report. This is indicative of the empirical success of the HSV specification as a tractable approximation of the tax code, especially on the progressivity dimension.

Figure (1.6) plots the average tax rate in 1980 based on the calibrated income tax function. The tax schedule of married households is below that of singles at low income levels, implying a marriage bonus for low earners. However, at higher income levels, being marriage is associated with larger tax obligations. In 2015, as depicted in Figure (1.7), the marriage bonus increases relative to 1980 due to a reduction in average tax rate at every level of income. In addition, the region over which married households face lower tax rates relative to singles has expanded, making marriage more appealing from the perspective of taxes.

1.3.2 Internal Calibration

The remaining parameters of the model are internally calibrated to match model-implied moments with empirical moments.

Skill Premium

Social and Economic Supplements data and the TAXSIM toolkit, Wu (2021) find a decline of τ from 0.15 to 0.08 between 1980 and 2015. Focusing on the tax rates of the very top earners, Piketty and Saez (2007) and Saez and Zucman (2020) argue that the tax code has become less progressive over the last few decades based on taxes paid by the top one percent.

Figure 1.6: Average Tax Rates by Taxable Income: 1980

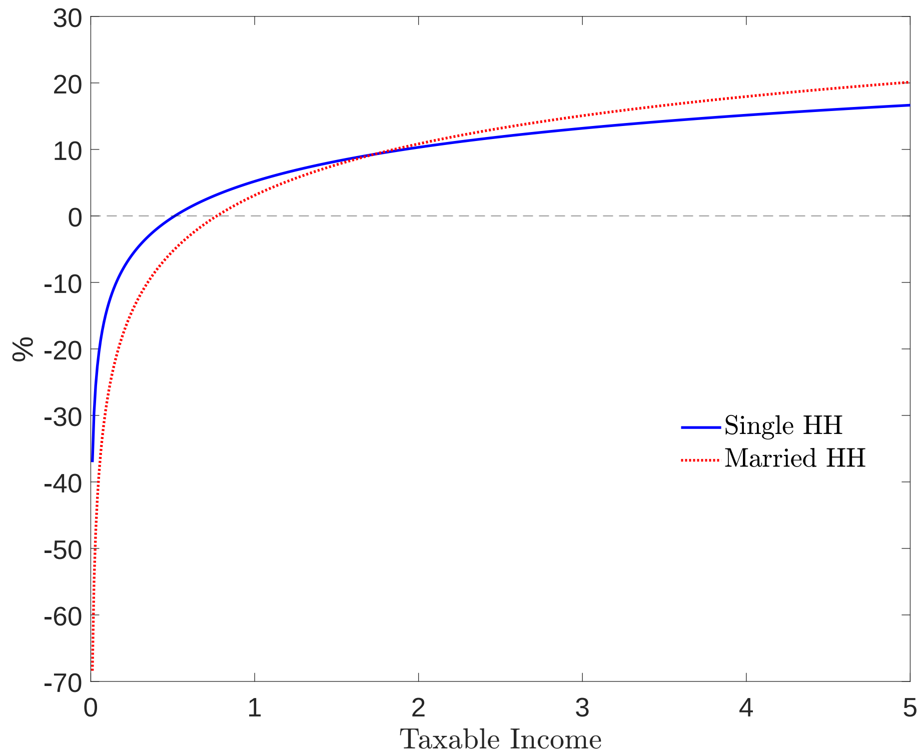


Table 1.2: Policy Parameters: 1980

Parameter	Description	Value	Source
rep_S^L	UI replacement rate (single,L)	0.30	Nickell et al. (2005)
rep_S^H	UI replacement rate (single,H)	0.30	Nickell et al. (2005)
rep_M^L	UI replacement rate (married,L)	0.30	Nickell et al. (2005)
rep_M^H	UI replacement rate (married,H)	0.30	Nickell et al. (2005)
PEN_S^L	SS replacement (single,L)	$0.4 \times \bar{w}_L$	SSA
PEN_S^H	SS replacement (single,H)	$0.4 \times \bar{w}_H$	SSA
PEN_M^{LL}	SS replacement (married,LL)	$2 \times 0.4 \times \bar{w}_L$	SSA
PEN_M^{HH}	SS replacement (married,HH)	$2 \times 0.4 \times \bar{w}_H$	SSA
PEN_M^{HH}	SS replacement (married,HL)	$1.5 \times 0.4 \times \bar{w}_H$	SSA
χ_S	average tax rate for single	-0.052	Borella et al. (2022)
τ_S	progressivity for single	0.075	Borella et al. (2022)
χ_M	average tax rate for married	-0.031	Borella et al. (2022)
τ_M	progressivity for married	0.12	Borella et al. (2022)

Figure 1.7: Average Tax Rates by Taxable Income: 2015

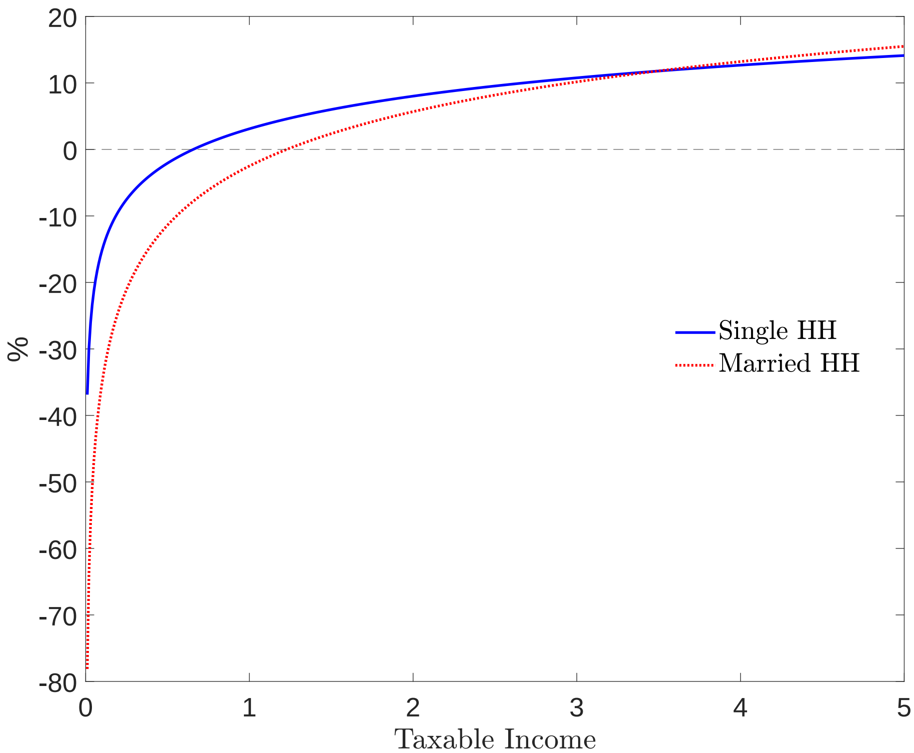


Table 1.3: Policy Parameters: 2015

Parameter	Description	Value	Source
rep_S^L	UI replacement rate (S,L)	0.29	OECD
rep_S^H	UI replacement rate (S,H)	0.29	OECD
rep_M^L	UI replacement rate (M,L)	0.29	OECD
rep_M^H	UI replacement rate (M,H)	0.29	OECD
PEN_S^L	SS replacement (S,L)	$0.4 \times \bar{w}_L$	SSA
PEN_S^H	SS replacement (S,H)	$0.4 \times \bar{w}_H$	SSA
PEN_M^{LL}	SS replacement (M,LL)	$2 \times 0.4 \times \bar{w}_L$	SSA
PEN_M^{HH}	SS replacement (M,HH)	$2 \times 0.4 \times \bar{w}_H$	SSA
PEN_M^{HH}	SS replacement (M,HL)	$1.5 \times 0.4 \times \bar{w}_H$	SSA
χ_S	average tax rate for single	-0.031	Borella et al. (2022)
τ_S	progressivity for single	0.08	Borella et al. (2022)
χ_M	average tax rate for married	0.025	Borella et al. (2022)
τ_M	progressivity for married	0.12	Borella et al. (2022)

Without loss of generality, the low-skill wage W^L is normalized so that the median income in 1980 is roughly one. The skill wage premium, W_H/W_L, is such that the average earnings ratio of college educated workers to non-college educated workers in the model match the data. Using the 1980 Census microdata (Ruggles et al., 2022), the average full-time worker with a four-year college education earned 55% more than the average full-time worker without. In 2015, this measure of the college skill premium has risen to 91% based on calculations using worker level data from the American Community Survey (Ruggles et al., 2022).

Preferences

The relative weight on market consumption in the final consumption basket, α_c, is chosen to match the ratio of time spent on market work to home production of 1.32 for the typical household in the US (Greenwood et al., 1995). Similarly, the disutility of work f is set so that the average leisure constitutes 42% of time (Greenwood et al., 1995). The relative weight on market purchases in the home production good, α_d, is set to match the average share of household-related expenditure of 12% in the Consumer Expenditure Survey.[10] Lastly, the discount factor β is calibrated to match the median wealth-to-income ratio of 1.27 in the 1983 Survey of Consumer Finance.

Marriage Parameters

There are four parameters related to marriage, which are the mean ($\mathcal{M}$) and dispersion ($\mathcal{L}$) of the love shock, the attractiveness of college graduates to other college graduates ($\mathcal{M}_{CC}$), and women's disutility during husband unemployment ($\mathcal{J}$).

In 1980, I normalize $\mathcal{M}$ to zero, and use the remaining three parameters to match the

[10] In the Consumer Expenditure Survey, I define the categories "Utilities, fuel, public services, "household operations", "housekeeping supplies", and "household furnishings and equipment" as expenditure related to home production.

marriage rates by education and the fraction of marriages where both spouses are college graduates. The dispersion of love shocks is directly linked to the relative importance of non-economic factors in marriage. When the dispersion is low, singles consider only the economic aspects when deciding whether or not to enter into a marriage. When the dispersion is high, singles may reject a match with high earnings even if it is economically sound, or similarly accept a match with whom marriage results in small economic gains. Furthermore, more dispersion love shocks imply a higher option value of waiting for the next match due to the upside risk of meeting someone that one really "loves". All in all, the overall marriage rate is lower when the love shock is more dispersed. The parameter $\mathcal{J}$, which governs the disutility of wives when their husbands become jobless, disproportionately affects the marriage prospects of men without a college education. This is because they face higher unemployment risks, which makes them less desirable partners in the marriage market. Therefore, this parameter helps to match the marriage rate among unskilled workers. Lastly, the utility bonus that college graduates get when matched with another college graduate $\mathcal{M}_{CC}$ is chosen to match the fraction of marriages where both spouses are college educated because the higher $\mathcal{M}_{CC}$ the more attractive college graduates find each other.

In 2015, the changes in the wage structure (skill premium and gender wage gap) and tax system are insufficient to account for the observed changes in marriage market patterns. Therefore, I allow the mean of love shocks $\mathcal{M}$ and the disutility for male unemployment $\mathcal{J}$ to vary in order to account for the residual. In particular, the mean of love shocks needs to be negative in 2015, reflecting the fact that there may be preference shifts away from marriage. Also, disdain for male unemployment needs to be larger in 2015 to account for the entire education gap in marriage rate.

Summary statistics on the marriage market are computed using the 1980 Census microdata and 2015 American Community Survey (Ruggles et al., 2022). The sample is restricted to individuals aged between 20 to 45. The marriage rate is calculated as the share of individuals that are ever married.

Education Parameters

The mean and standard deviation of the education cost ξ are time invariant and chosen to match the college attainment rates by gender in both 1980 and 2015. In 1980, 27% of males aged 20-45 have completed a four-year college education compared to 18% of females. In 2015, female college attainment rose rapidly to 38%, while the male attainment rate experienced only a modest growth to 31%. The asymmetrical rise in education across genders is achieved by having a less dispersed cost distribution for women, which induces more college graduates when the skill premium rise. This calibration approach implicitly validates the elasticity of college attendance as the model endogenous match changes in education attainment in response to economic and marital changes. This is useful as it allows for more realistic education responses during the evaluation of policy counterfactuals.

Table (1.4) summarizes the internally calibrated parameters and the fit of the calibrated model moments against target moments from the data.

1.3.3 Changes in Marriage Market: 1980 vs 2015

Decline in Marriage Rate

By construction, the model matches the marriage rates by education in both 1980 and 2015. Several forces are responsible for the decline in marriage.

Table 1.4: Internal Calibration

Parameter	Value	Moment	Data	Model
$\mathcal{L}$	3.8	Col Marriage Rate (1980)	0.76	0.75
$\mathcal{J}^{1980}$	3.7	No Col Marriage Rate (1980)	0.77	0.78
$\mathcal{M}_{CC}$	0.65	Both Spouse with Col	0.10	0.10
$\mathcal{M}^{2015}$	–2.7	Col Marriage Rate (2015)	0.69	0.67
$\mathcal{J}^{2015}$	8.7	No Col Marriage Rate (2015)	0.58	0.58
μ_ξ^m	135.0	Male: fraction with college	0.27	0.26
μ_ξ^f	40.5	Female: fraction with college	0.18	0.18
σ_ξ^m	161.7	Male: change in college rate	0.04	0.06
σ_ξ^f	18.2	Female: change in college rate	0.20	0.20
$(w_H/w_L)_{1980}$	1.40	Skill premium 1980	1.56	1.56
$(w_H/w_L)_{2015}$	1.70	Skill premium 1980	1.91	1.91
α_c	0.55	Ratio of market work to home work	1.32	1.32
α_d	0.28	Expenditure share on home good	0.13	0.12
f	12.0	Average leisure	0.42	0.42
β	0.991	Median wealth-to-income ratio	1.27	1.27

The increase in skill premium reduces the gains from marriage for educated workers. As income increases, educated workers can substitute towards market purchases for the home production good, which reduces the need for marriage for home production and time specialization reasons. Also, they can accumulate more savings which dampens the gains from spousal insurance. These forces reduce the incentives for educated individuals to get married. Meanwhile, higher returns to education also leads to a change in the skill distribution. In particular, the increase in college attainment for women doubled between 1980 and 2015, while men college share rose modestly. This makes matches between college educated workers more prevalent, which increases the marriage rate. The narrowing of the gender wage gap makes singlehood more affordable for women and therefore make them choosier in the marriage market as they are more able to turn down a bad match. At the same time, women become more attractive from the perspective of men due to improved labor market prospect.

On net, the model is able to endogenously generate the rising gender gap in marriage through changes in the wage structure, as it leads to an increase in the college marriage rate (from 75% to 80%) and a decrease in the non-college marriage rate (from 78% to 70%). To match the overall decline in marriage rate observed in the data, there needs to be a reduction in non-pecuniary match quality, which is reflected in a reduction in the mean of love shocks $\mathcal{M}$. The reduction in $\mathcal{M}$ can also be interpreted as changes in social and cultural norms, such as a reduction in stigma related to singlehood (Stevenson and Wolfers, 2007). The remaining portion of the gender gap is attributed to an increase in the distaste towards unemployed men $\mathcal{J}$.

Marital Sorting by Education

Figure (1.5) displays marital sorting patterns by education in both 1980 and 2015.[11] With the exception of the fraction of marriages between college graduates in 1980, all the statistics are untargeted. In 1980, the model is able to match the marital sorting patterns quite well, with the majority of marriages occurring between individuals without a college education. Due to the relative abundance of male college graduates, marriages between a college educated husband and non-college educated wife are common, accounting for 17% of all marriages.

In 2015, the increase in the skill premium and increase in supply of college graduates strengthens motives for assortative matching. This is because for college educated workers, the costs of marrying someone without a college education is higher. The reduction in the gender wage gap also makes women pickier in the marriage market. As a result, the model generates a large increase in homogamy rate of college graduates, defined as the fraction of marriages that occur between the same type. The model is also able to replicate the relative decline in marriages between a male college graduate and female non-college graduate, as well as a relative rise in

[11] The empirical statistics are computed using 1980 Census and 2015 ACS, restricted to individuals aged 20 to 45.

Table 1.5: Marital Sorting: Model vs. Data

	Model		**Data**	
Household type	1980	2015	1980	2015
Husband college, wife college	10%	27%	10%	25%
Husband <college, wife college	8%	9%	5%	17%
Husband college, wife <college	17%	9%	12%	9%
Husband <college, wife <college	66%	56%	73%	49%
Homogamy rate if one partner has college	29%	60%	37%	49%

marriages between educated women and less educated men, though the increase in the latter is dampened by an increased distaste for unemployed men.

The success of the model in matching untargeted marital sorting patterns lends validation to the model and suggests that random-matching in the marriage market is a reasonable assumption. In fact, the model generates a larger increase in assortative sorting compared to the data. If there was directed search in the marriage market, the degree of sorting would be even higher.

1.4 Policy Exercises

This section explores the optimal design of redistribution policies from the viewpoint of a utilitarian planner whose objective is to maximize the expected lifetime utility of a newborn. Because all agents are ex-ante homogeneous, the welfare function that the planner maximizes is

$$W = \int \mathbb{E}V^{m}_{S,m}\left(e,u,z,0\right)dG_m(e,u,z) + \int \mathbb{E}V^{f}_{S,m}\left(e,u,z,0\right)dG_f(e,u,z) - \hat{\xi}_m - \hat{\xi}_f \quad (1.46)$$

where $\hat{\xi}_g$ denote the total education costs paid which is the mean of the education cost distribution truncated above by $V_0^g(H) - V_0^g(L)$ multiplied by the equilibrium measure of college educated workers.

1.4.1 Optimal Tax Progressivity

This section studies the optimal progressivity of income taxes. To start, the progressivity parameter τ is varied for one type of household while holding the tax policy of the other household type fixed. A full optimization is then conducted where $(\tau_S, \tau_M, \chi_S, \chi_M)$ are jointly chosen to maximize welfare. All exercises are revenue neutral. In the first set of experiments, this is done by adjusting χ. In the full optimization, revenue neutrality imposes a restriction on one of the four parameters.

1.4.1.1 Optimal Taxes in 1980

The optimal tax plans from the three exercises are summarized in Table (1.6). First consider the case where taxes for married households are held at the baseline specification. When only τ_S is allowed to vary, the constrained optimal degree of progressivity for single household is 0.45. Comparing with the baseline, as can be seen in panel (a) of Figure (1.8), the optimal plan entails larger transfers to workers with low pre-tax income financed by higher taxes for high earners. The optimal plan also features a marriage bonus for more productive workers and a marriage penalty for less productive workers. This turns out to reduce the marriage rates for both education types, with college and non-college graduates being 4 p.p. and 2 p.p. less likely to be married. In fact, this reduction is driven by workers with lower productivities not wanting to get married. This can be seen in Table (1.7) by comparing the pre-tax income of married households under the baseline and optimal τ_S regimes, which shows an increase in pre-tax incomes of married households across income quantiles. Because taxes for married households are identical in both economies, this reflects purely a change in composition of couples. Overall, this plan yields a consumption-equivalent (CEQ) welfare gain of 3.84%, with the majority of gain coming from larger transfers to poor singles who have the highest marginal utility of consumption.

Similar to the previous exercise, the optimal degree of tax progressivity for married household when holding taxes of singles fixed is much larger than the baseline regime, warranting an increase of τ_M from 0.12 to 0.46. This optimal plan, in contrast, features large marriage bonuses for poorer households and a marriage penalty for high income households. Accordingly, this tax regime leads to a large 8 p.p. decline in marriage rate among college graduates but yields a small increase in marriages among low-skilled workers. A larger bonus at lower ranges of income improves the marriage prospects of less productive workers, who become more likely to be married as evident in the increase in average pre-tax incomes of those who remain single from Figure (1.7). Although this plan encourages marriage, relative to the optimal τ_S plan, it yields a smaller CEQ welfare gain of 2.66%. This is because transfers are targeted at poor married households rather than poor singles who are more vulnerable. Furthermore, it more heavily taxes married household, who constitute the majority of the population, and hence leads to more losses from distortions on labor supply.

The optimal tax plan when all parameters are free to vary is composed of a highly progressive tax schedule on both household types with marriage bonuses across a wide range of income levels as seen in panel (c) of Figure (1.8). This plan, which yields a substantial CEQ welfare gain of 8.51%, involves raising τ_S to 0.50 and τ_M to 0.55, combined with a higher average tax rate on singles and lower average tax rate on couples. Under this regime, marriage becomes more common across both skill types, with the overall marriage rate increasing by 3 p.p. to 80% relative to the baseline. The bottom panel of Table (1.7) sheds light on the merit of this plan relative to the previous two. First of all, it provides transfers to poorer households regardless of

Table 1.6: Optimal Tax Functions: 1980

	Baseline	Optimizing τ_S	Optimizing τ_M	Optimizing All
τ_S	0.08	0.45	0.08	0.50
χ_S	–0.052	–0.059	–0.052	–0.16
τ_M	0.12	0.12	0.46	0.55
χ_M	–0.031	–0.031	0.159	0.309
Marriage Rate	0.76	0.73	0.75	0.80
Col Marriage Rate	0.75	0.71	0.67	0.77
<Col Marriage Rate	0.77	0.75	0.78	0.80
Male college share	0.26	0.26	0.25	0.24
Female college share	0.18	0.14	0.12	0.06
CEQ	0.00%	3.84%	2.66%	8.51%

marital status. More importantly, transfers are more generous towards single households, with more than half of single households facing negative tax rates. Unlike the τ_S exercise, this is achieved without imposing a marriage penalty on the poor. While less productive workers are more likely to get married, they are bailed out by a larger safety net should they remain single due to poor luck.

1.4.1.2 Optimal Taxes in 2015

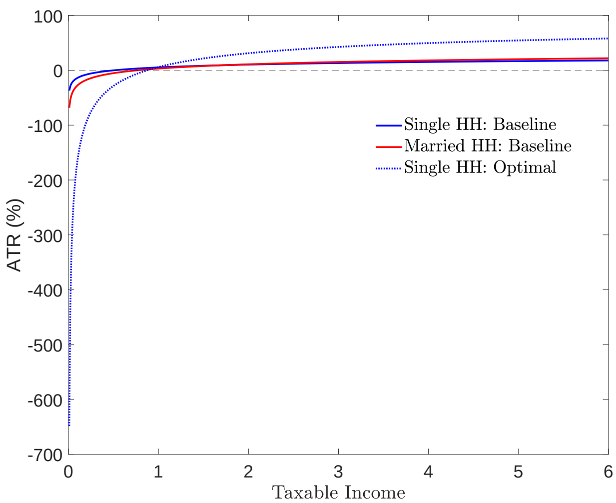
The same set of exercises are repeated for 2015. Relative to the model calibrated to 1980 data, the economy in 2015 differs along two dimensions. The wage structure has undergone significant changes, with the skill premium increasing sharply from 40% to 70% and the female wage gap closing from 79% to 92%. Demographically, marriage has become much less common in 2015 and particularly so for less educated individuals, which is attributed to changes in the wage structure and shifts in marital preferences towards singlehood.

The optimal tax plans under the three exercises are summarized in Table (1.8). In the τ_S exercise, the optimal plan resembles that in 1980 which features an increase in tax progressivity

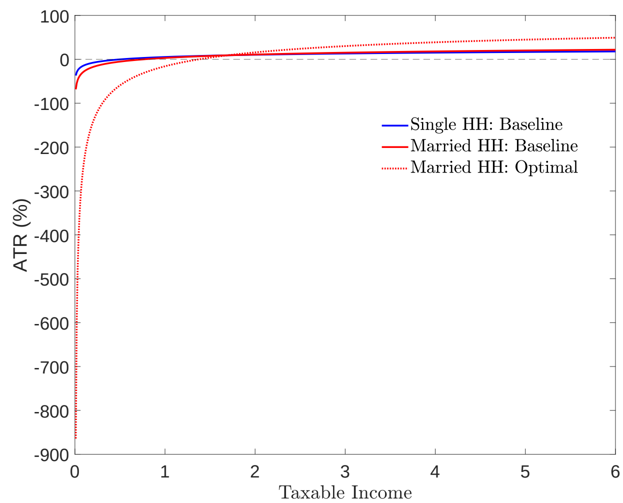
Table 1.7: Pre-tax Income and Average Tax Rate under Different Tax Rules: 1980

		Q1	Q2	Q3	Q4	Q5
Baseline						
Single	Pre-tax Y	0.29	0.70	1.21	2.06	4.10
	ATR	–0.01	0.05	0.08	0.11	0.16
Married	Pre-tax Y	0.83	1.50	2.15	3.04	5.28
	ATR	0.03	0.09	0.13	0.16	0.22
All	Pre-tax Y	0.50	1.20	1.84	2.71	4.79
	ATR	0.01	0.08	0.12	0.15	0.19
Optimal τ_S						
Single	Pre-tax Y	0.20	0.44	0.73	1.16	2.47
	ATR	–0.85	–0.36	–0.07	0.12	0.38
Married	Pre-tax Y	0.94	1.67	2.32	3.23	5.35
	ATR	0.02	0.08	0.11	0.14	0.20
All	Pre-tax Y	0.35	0.95	1.60	2.53	4.61
	ATR	–0.41	0.03	0.14	0.18	0.20
Optimal τ_M						
Single	Pre-tax Y	0.33	0.83	1.48	2.45	4.61
	ATR	–0.05	0.04	0.09	0.12	0.16
Married	Pre-tax Y	0.69	1.26	1.73	2.41	4.03
	ATR	–0.33	–0.04	0.10	0.22	0.39
All	Pre-tax Y	0.50	1.11	1.66	2.46	4.27
	ATR	–0.15	–0.06	0.08	0.21	0.29
Optimal (τ_S, τ_M)						
Single	Pre-tax Y	0.19	0.44	0.67	1.17	2.50
	ATR	–0.80	–0.27	–0.03	0.23	0.47
Married	Pre-tax Y	0.75	1.36	1.92	2.63	4.36
	ATR	–0.46	–0.11	0.08	0.23	0.42
All	Pre-tax Y	0.36	0.92	1.52	2.27	3.94
	ATR	–0.36	–0.13	0.03	0.20	0.40

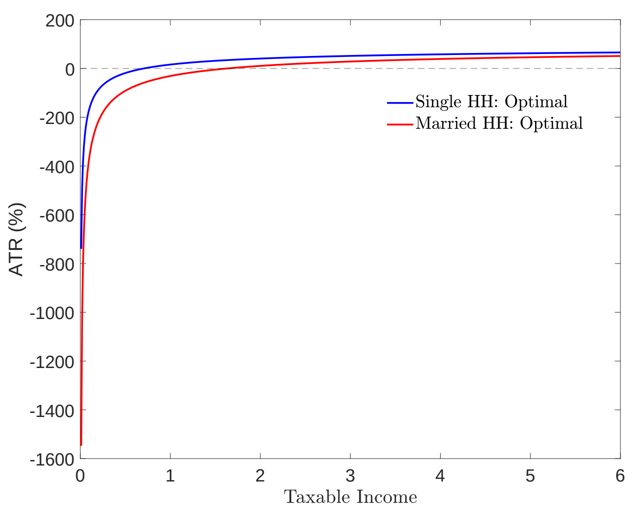
Figure 1.8: Optimal Tax Functions: 1980



(a) Optimal Single Tax Function with Married Tax Fixed



(b) Optimal Married Tax Function with Single Tax Fixed



(c) Optimal Tax Functions

for singles and a sizable CEQ welfare gain of 4.32%. Several forces are at play. In general, women are better off in 2015 as the gender wage gap decreases and college attainment broadens which reduces the need for redistribution. Meanwhile, men income dispersion rises while unskilled men face stagnant real wages and dimmer marital prospects. This creates a counteracting force that calls for more redistribution. Lastly, the increase in skill premium worsens the distortionary effects of progressive taxes which is further amplified by an increase in assortative matching. Netting the three forces, the optimal policy calls for an increase of similar magnitude as in the 1980 plan. In fact, the same τ_S increase in 2015 allows the average tax rate to decrease because of a larger tax base at the top of the distribution (singles in the highest quantile face an ATR of 38% in 1980 but 42% in 2015) and a reduction of the most vulnerable type (the poorest 20% of singles receives an average transfer rate of 57% relative to 85% in 1980).

The optimal τ_M in 2015, on the other hand, can only generate a mere welfare gain of 0.4%. This plan involves a small increase in τ_M to 0.25. Compared to the optimal τ_M plan in 1980, the 2015 plan calls for a marriage bonus over a wider range of income to accommodate an increase in average income. Unlike the 1980 plan, it has minimal effect at promoting marriage for the less desirable individuals. Comparing the pre-tax income distribution of singles between the baseline and optimal τ_M policies in Table (1.9), it can be seen that the average income in the poorest quantiles does not increase when moving to the optimal τ_M plan in contrast to 1980. This is attributed to the fact that poor individuals are much less desirable in 2015 due to increase in reservation values and shift in preference towards singlehood. Under this circumstance, it is more efficient to raise revenue from high earning couples and redistributing to poor singles rather than promoting marriages at the expense of more distortionary taxes.

A joint optimization of taxes for both household types again calls for more progressive

Table 1.8: Optimal Tax Functions: 2015

	Baseline	Optimizing τ_S	Optimizing τ_M	Optimizing All
τ_S	0.075	0.45	0.075	0.35
χ_S	–0.031	0.011	–0.031	–0.05
τ_M	0.12	0.12	0.25	0.45
χ_M	0.025	0.025	0.134	0.41
Marriage Rate	0.61	0.62	0.61	0.64
Col Marriage Rate	0.67	0.72	0.63	0.72
<Col Marriage Rate	0.58	0.54	0.59	0.62
Male col share	0.32	0.30	0.31	0.29
Female col share	0.36	0.26	0.31	0.14
CEQ	0.00%	4.32%	0.41%	5.90%

taxes universally with marriage bonuses across the board, which results in a CEQ welfare gain of 5.9%. Relative to the observed tax schedules in the data, the optimal plan is more favorable to marriages and partially reverts the decline in marriage rate, raising the college and no-college marriage rates by 5 p.p. and 4 p.p. respectively. However, compared to 1980, this plan is less progressive which is mainly due to the fact that revenue losses from more progressive taxes are larger when the skill premium increase. Furthermore, improving conditions for women alleviate the needs for large transfers at the bottom half of the income distribution.

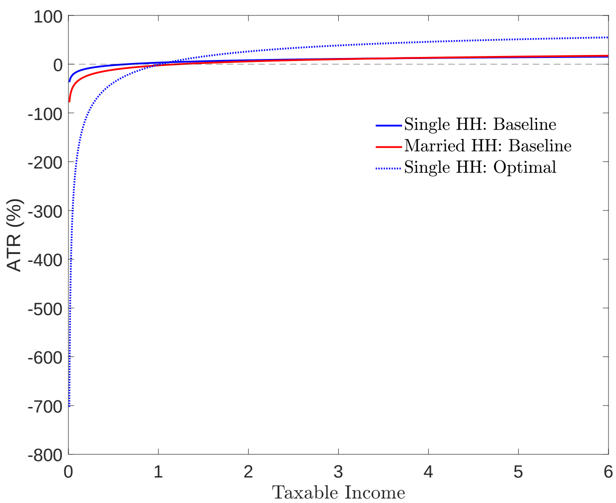
1.5 Conclusion

A central question in public finance and macroeconomics is the optimal design of redistribution policies in the presence of inequality. Large changes in the income distribution over the last few decades have animated wide academic and policy discussions about appropriate policy responses. Meanwhile, there has also been dramatic demographic changes in terms of marriage patterns as characterized by a decline in marriages and rise in assortative matching. Literature that studies the optimal response of redistribution policies often overlook the marriage margin, although it

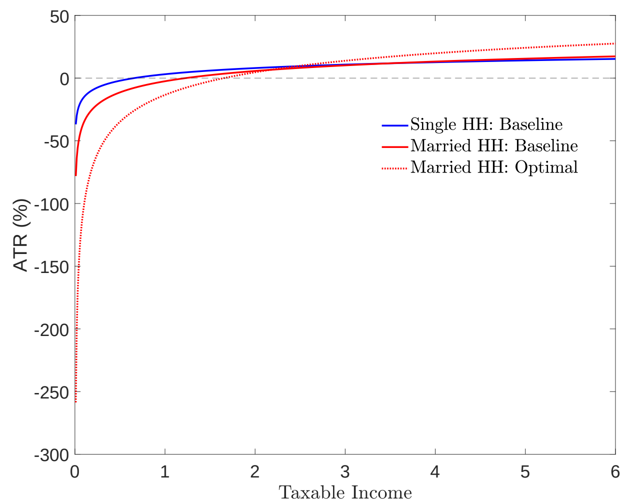
Table 1.9: Pre-tax Income and Average Tax Rate under Different Tax Rules: 2015

		Q1	Q2	Q3	Q4	Q5
Baseline						
Single	Pre-tax Y	0.43	0.91	1.43	2.27	4.47
	ATR	–0.02	0.00	0.04	0.07	0.14
Married	Pre-tax Y	1.15	1.93	2.57	3.49	5.66
	ATR	0.00	0.06	0.09	0.13	0.17
All	Pre-tax Y	0.59	1.27	1.96	2.87	5.02
	ATR	0.01	0.04	0.06	0.10	0.14
Optimal τ_S						
Single	Pre-tax Y	0.29	0.57	0.89	1.38	2.85
	ATR	–0.57	–0.20	0.02	0.19	0.42
Married	Pre-tax Y	1.26	2.09	2.85	3.76	5.89
	ATR	0.02	0.06	0.10	0.12	0.17
All	Pre-tax Y	0.40	0.93	1.64	2.66	4.84
	ATR	–0.39	0.01	0.15	0.18	0.18
Optimal τ_M						
Single	Pre-tax Y	0.42	0.94	1.49	2.37	4.47
	ATR	0.00	0.06	0.10	0.12	0.14
Married	Pre-tax Y	1.06	1.74	2.32	3.12	4.90
	ATR	–0.13	0.00	0.08	0.14	0.23
All	Pre-tax Y	0.59	1.28	1.94	2.81	4.80
	ATR	0.01	0.03	0.07	0.14	0.22
Optimal (τ_S, τ_M)						
Single	Pre-tax Y	0.33	0.69	1.10	1.66	3.40
	ATR	–0.32	–0.07	0.08	0.20	0.39
Married	Pre-tax Y	1.05	1.68	2.23	3.01	4.76
	ATR	–0.36	–0.13	0.01	0.14	0.30
All	Pre-tax Y	0.46	1.05	1.63	2.42	4.22
	ATR	–0.21	–0.02	0.00	0.11	0.30

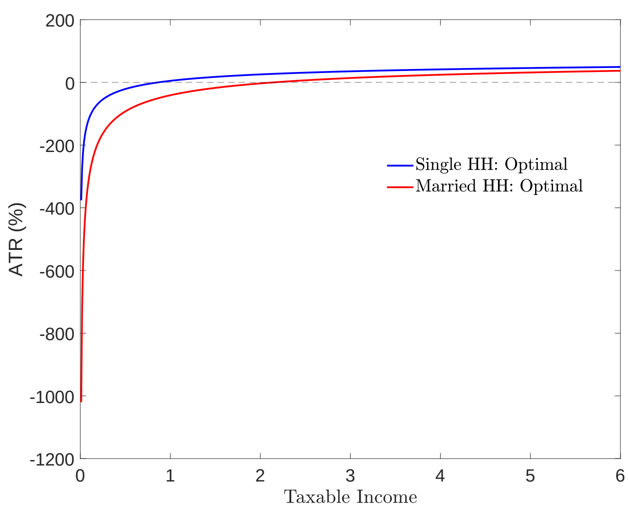
Figure 1.9: Optimal Tax Functions: 2015



(a) Optimal Single Tax Function with Married Tax Fixed



(b) Optimal Married Tax Function with Single Tax Fixed



(c) Optimal Tax Functions

has important consequences the distribution of welfare across the population.

To this end, this paper proposes a novel modeling framework for studying optimal taxation policies that incorporates the interactions between marriage, broad economic environment, and redistributive taxes. Unlike the standard literature, marriage is endogenous to economic environment including policy. It also augments the distributional implications of changing inequality. The model features a heterogeneous agent life-cycle model with uninsurable income risks extended with a marriage market where workers decide whether to marry and to whom.

The calibrated model is used as a laboratory to first study the optimal progressivity of labor income tax in the US in the years 1980. Through the lens of the model, the degree of progressivity in the US tax system is too small relative to the optimal plan. In addition, optimal policy calls for large marriage bonuses alongside generous transfers to the poorest singles. The model is then used to assess how policy should react to a rising skill premium, narrowing gender wage gap, and shift of preference towards singlehood that occurred between 1980 and 2015. The optimal policy in 2015 features less progressivity, though still significantly more than the tax code that is currently in place, as women become better off and the distortionary effects of progressive taxation enlarge with a rise in the return to skill.

The modeling framework can be extended to study related questions. For instance, this paper does not consider the role of changing income risks. There are disagreements in the literature regarding the nature of labor income risks over time. The conventional wisdom is that unexplained wage dispersion has increased since the 1980's which the literature has taken as evidence of more volatile shocks to income over the life-cycle (Blundell et al., 2016; Braxton et al., 2021; Gottschalk et al., 1994; Heathcote et al., 2010, 2014). This view has since been challenged by works using richer administrative earnings data who find no increase in life-cycle risks but an increased dispersion in initial conditions (Guvenen et al., 2017; Moffitt, 2020). Understanding the effects of either story on marriages and how that augments the design of optimal policy is a promising future avenue of research. In principle, more volatile risks call for a larger role of redistribution policy as a insurance tool. If income inequality is instead driven by increased in initial conditions, the insurance motive is absent but may entail a different policy prescription.

Another direction for future work is to incorporate fertility into the modeling framework to study the policy response to increased single parenthood. According to birth data from the National Vital Statistics System in 2016, close to 40% of newborns in the US are born out of wedlock and the majority of them to parents with less education. Given the importance of family structure and parenting on long-term outcomes (Doepke et al., 2019), this is a worrisome trend that may inhibit intergenerational mobility. Elsewhere in the world, some nations are facing demographic crisis due to low fertility rates, a challenge that can be potentially addressed in an extension of this framework.

Chapter 2: Dynamic Consequences of Migration Frictions: How China's Hukou Policy Affects Human Capital Accumulation and Intergenerational Income Persistence

2.1 Introduction

The Chinese economy has been growing at a tremendous pace since its much celebrated market reforms. In 1978, when China first opened itself to the world after decades of economic isolation and political unrest, its GDP per capita stood at a mere $300 (in 2010 USD). Over the last 40 years, that number has increased more than 25 times to $7,800.[1] In line with the past experiences of many developed countries, China's economic development was accompanied by large rural to urban migration flows. The stock of migrant population increased from 7 million in 1982 (0.7% of the total population) to around 250 million (18% of the total population) in 2017.[2] Despite these internal migration flows, which some people call the largest ever in human history (Tombe and Zhu, 2019), migration is in fact not free in China, due to its unique internal migration policies - the *hukou* system. The *hukou* policy is a residency registration system that links each citizen to a particular administrative unit. Eligibility for public services (such as healthcare and

[1] Data from World Bank National Accounts

[2] Data from the National Bureau of Statistics of China. The definition of migrant here is an individual who resides away from her location of hukou registration. An alternative definition in which a migrant is defined as one who lives outside of her county of birth, yields almost identical results.

education) and social benefits (such as social security, retirement pensions, and other government transfers) is tied to the location of one's *hukou* registration. In other words, migrants living away from their *hukou* location do not enjoy the same rights and privileges as natives do.

In this paper, I argue that the *hukou* system of China acts as a barrier to income and educational mobility across generations. First of all, the *hukou* policy deters worker migration from low-income regions to regions offering higher earning opportunities. Not only does the policy suppress upward income mobility by making migration costly, it also leaves those who are stuck in low-income regions with fewer monetary resources to invest in the human capital of their children. Secondly, due to a lack of access to public education and healthcare, many workers who do migrate are forced to leave their children in their hometowns as they move to work in the cities. This is a prevalent phenomenon, commonly referred to as the "left-behind children". The empirical literature has shown that parental time investment is crucial to children's human capital formation and that exposure to parental absence has adverse effects on both health and educational outcomes of children. Since parents who choose to leave their children behind are typically those with lower income and education, the inability to migrate as a family unit is another source of intergenerational immobility.

An important implication that follows from my argument is that the *hukou* policy not only contributes to the persistence of income inequality and economic status across generations, it also reduces the overall accumulation of human capital and consequently the growth rate in the economy. This suggests a dynamic welfare cost of the *hukou* policy and migration frictions in general, a cost that is often overlooked in the literature, which emphasizes mostly the static losses through labor misallocation.

To address my research question, I build a two-region heterogeneous-agent life-cycle model that extends an otherwise standard Huggett-Aiyagari style model to allow for migration between regions as well as parental investments in child human capital. In the model, agents are heterogeneous with respect to their supplies of unskilled and skilled labor. The rural region uses unskilled labor to produce a rural good, while the urban region uses skilled labor to produce an urban good. As a departure from standard assumptions, workers in the model are endowed with both skilled and unskilled labor. The type of labor that a worker supplies is determined by her location.[3] While the existence of an urban wage premium creates incentives for rural-urban migration, not all agents choose to move due to higher child-rearing and education expenses as well as exposure to uninsurable income risk in the urban region. Parents in the model can choose to leave their children behind in the villages while they work in the cities to reduce child-related expenses, though at the expense of lower parental investment in their children's human capital.

The rest of the paper is structured as follows. In Section 2, I provide a more detailed account of the *hukou* policy and document some facts regarding migration, the "left-behind children" phenomenon, and intergenerational mobility in China. In Section 3, I review existing studies relevant to my research question and discuss my planned contribution to the different strands of literature. In Section 4 and 5, I describe the structure of the model and present some quantitative results. In the last section, I will lay out a road map for future work.

[3]This assumption is based on the idea that skilled labor is not more effective at rural production than unskilled labor.

2.2 Institutional Background and Facts

2.2.1 Hukou Policy

The *hukou* system is a residence registration system that classifies each citizen as a resident of a specific administrative unit. In addition to the location of residency, each citizen is categorized into agricultural (rural) or non-agricultural (urban) status. To change one's location or category of registration, approval from local governments in the destination is needed.

In the planned economy era, working outside one's location or sector of registration was strictly prohibited. Due to the nature of the command economy system, each worker was assigned a job according to their *hukou* status. As such, workers with rural registration were assigned agricultural jobs in their location of residency whereas those with urban registration were assigned non-agricultural jobs in the cities. As China began to transition into a market economy starting in 1978, workers started to have the freedom to choose between jobs and occupations. However, working outside one's registration status or residency location remained difficult. Workers would need to apply for temporary residence permits to be able to live and work outside their residency location. Workers with agricultural registration were legally not allowed to be employed in non-agricultural sectors unless approved by the government. Such permits were difficult to obtain and those who did not risked arrest and deportation.

As the market reform deepened, some of these restrictions were relaxed. Beginning in 2003, some provinces and cities started to abolish the distinction between agricultural and non-agricultural registrations (up to 13 provinces out of 31 by the year 2009). Many more areas removed the requirement for migrant workers to obtain temporary permits in order to live and work. Although over time it has become easier for workers to move legally across regions, it remained difficult for individuals to change their official *hukou* registration. It is extremely rare for rural households to be able to obtain an urban *hukou* regardless of how long they have lived in the area (Zhong, 2019). Furthermore, even when given approval to change one's *hukou* registration, many hesitate to do so because those with rural registration automatically lose land ownership in their home townships as soon as their registration is converted to urban. These frictions generate a large cost of migration, as eligibility for public services, healthcare, education, and social security remain tied to one's *hukou* registration. In other words, even though workers are now in general free to migrate and work anywhere in the country, those living away from their *hukou* registration do not have access to public healthcare, cannot send their children to the local public school for free, and cannot participate in unemployment insurance and pension programs. Typically speaking, access to these public services and programs is limited to *hukou* holders at the city or county level, with the exception of pension programs that are typically managed at the province level. For example, a worker who moves from one city to another within a province will not have access to health and education benefits in her new city but will remain eligible to participate in the pension program. The following table shows the participation rates in public programs, comparing hukou holders in the cities and migrant workers without local registration, as calculated by the author using data from the Longitudinal Survey on Rural Urban Migration in China (RUMiC) in 2008.[4]

[4] A note published by the Peterson Institute for International Economics indicate that migrant worker participation rates in health insurance, unemployment insurance, and public pension program are 22%, 17%, 22% respectively in year 2017, citing data from the Ministry of Human Resources and Social Security of China and the National Bureau of Statistics in China. (https://www.piie.com/research/piie-charts/many-migrant-workers-are-excluded-chinas-social-programs)

Table 2.1: Public Program Coverage
(Source: RUMiC 2008)

	Hukou Holders	Migrant Workers
Health Insurance	82%	3%
Unemployment Insurance	65%	15%
Pension	78%	22%

In 2014, the central government of China announced its plan to completely abolish the rural-urban distinction across the entire nation by 2020. It also announced measures to ease the official change of *hukou* registration (including streamlining the application process and allowing rural individuals to retain land ownership or receive fair compensation), to improve access to public services for those without local *hukou*, and to harmonize social security programs across provinces. Despite coordination from the central government, the exact pace and implementation of *hukou* policy reform is left to local governments. Different provinces and cities are allowed to implement their own rules as to the requirements for obtaining local residency. Typically, more developed provinces and cities have established high standards for residency registration in terms of education, wealth, and age.

2.2.2 Facts on Migration in China

As mentioned in the introduction, internal migration in China has surged as the economy took off. Most of these flows are induced by the large spatial dispersion in income levels across provinces. Throughout the course of China's development, growth was not evenly felt across regions, with coastal provinces experiencing faster increases in income relative to inland provinces. As of 2018, the real income gap between the richest and poorest provinces, measured in terms of CPI-adjusted real household disposable income, was about 5 times according to Wen

(2018).[5] Zhu (2002) finds that the income gap is the most important determinant of migration decisions in China, suggesting that most workers move to take advantage of higher earning opportunities.

The literature has found that younger workers in China are more likely to migrate (Li and Zahniser, 2002; Zhao, 1999) because older workers are typically more dependent on public services such as healthcare. Those with more education as well as those with lower landholdings and farm income have a higher propensity to migrate (Li and Zahniser, 2002; Zhao, 1999, 2002; Zhu, 2002). This is because high skilled workers are more likely to find high paying jobs in the cities relative to those with low skills. For workers with high farm income, the rural-urban income gap is smaller, thereby reducing the incentive to move. Workers with large land holdings and high agricultural income are also less likely to migrate because of the risk of land expropriation by the local authorities and the absence of a complete land rental market (Adamopoulos et al., 2017; Ngai et al., 2018).

In terms of migration distance, workers who are younger, single, and more educated are more likely to move across provincial boundaries, whereas those of older age, who are parents, and those with less education tend to stay within their province of origin.[6] Furthermore, migration spells are often temporary, with the mean duration of stay away from home being seven years (Meng, 2012), and 70% of rural migrants leave cities within 10 years of moving (Zhong, 2019).[7] Migrants typically leave their hometowns in their late teens, returning several years later to marry

[5] The same ratio is about 4.5 after adjusting for the purchasing price of housing properties.

[6] According to a report by the National Bureau of Statistics of China in 2017, the average age of inter-provincial migrants is 34 whereas the average age of intra-provincial migrants is 44. About 65% of inter-provincial migrants are married while over 90% of intra-provincial migrants are married.

[7] Using survey data from around 800 households from 6 selected provinces, Zhao (2002) finds that about 40% of migrant families eventually return to their home villages. This is likely to be an underestimate for return migration rate as it considers only migrants who return to their specific home villages.

and have children. Other reasons for return migration include illness and having to take care of family members.

2.2.3 Left-behind Children

Since migrant workers do not have access to public education and healthcare, it is expensive for them to raise their children in the cities. As a result, many workers decide to leave their children in their home villages as they work in urban areas. According to the 2010 National Census, about 61 million children under the age of 18 have at least one of their parents working away from home, which accounts for almost one-third of the rural child population. Of these left-behind children, a majority live without both parents. Based on the 2008 sample of the Longitudinal Survey on Rural Urban Migration (RUMiC), about 30% of left-behind children live with one parent, 59% live with grandparents or other relatives, and 8% are enrolled in boarding schools that are generally regarded as very poor in quality. The RUMiC also asks migrants for the main reason why they decided to leave behind their children, and over 70% cite high tuition, lack of childcare, and the high cost of living as the answer. Chen et al. (2019) study the determinants of parents leaving children behind and find that parents who are less skilled, have lower income, and do not have local *hukou* are more likely to leave their children behind when migrating. Zhong (2019) finds that parents are more likely to leave their children behind when grandparents are available to take care of the children. In addition, she finds that migrants who leave child-rearing to grandparents are more likely to make transfers to grandparents when they become sick, hinting at an informal contract between parents and grandparents.

The remarkably large population of left-behind children leads to the following question:

what are the effects of parental migration on children's human capital development? On the positive side, parental migration might allow for higher human capital investment in children through remittances. On the negative side, parental care has been found to be very important to child outcomes, implying that parental absence due to migration can potentially be detrimental to human capital accumulation. In addition, it is possible that children engage in more household tasks when one or both parents are absent. In the case of China, most empirical evidence suggests that the overall impact of parental migration on children's human capital is negative. Meng and Yamauchi (2017) finds that the length of parental absence due to migration has adverse effects on both health and educational outcomes of children in China, suggesting that the negative effects of parental absence dominate the potential gain from increased financial resources.[8] Their results are in line with previous works that focus on the contemporaneous effect of parental migration on educational attainment (Meyerhoefer and Chen, 2011; Zhang et al., 2014) and health (Mu and De Brauw, 2015). However, several studies that find insignificant effects of parental migration on children outcomes. For instance, Chen (2013) examines rural households in which only the father is absent due to migration and finds no impact on the child's development, as mothers adjust their behaviors and take closer care of the child. Although the effects of the absence of a single parent are perhaps more debatable, the general consensus appears to be that the absence of both parents does have detrimental effects on child outcomes. Given that the majority of left-behind children reside without either parent, it is reasonable to believe that the "left-behind children" phenomenon is in general associated with lower human capital accumulation in the economy.

Although the empirical evidence suggests that parental migration has detrimental impacts

[8]To mitigate endogeneity in parental migration decision, the authors use weather shocks and distance to the provincial capital city as instruments.

on child skill formation in China, it should be noted that there are studies on other countries that suggest otherwise. For instance, Botezat and Pfeiffer (2014) finds that parental migration has positive impacts on educational attainment but negative impacts on mental health for left behind children in Romania.

2.2.4 Intergenerational Mobility

Using the China Family Panel Studies, Fan et al. (Forthcoming) finds that the intergenerational elasticity of income (IGE), which measures the percentage of parental income position transmitted to the next generation, has increased from 0.39 for the 1970-1980 birth cohorts to 0.44 for the 1981-1988 birth cohorts. As a reference, the IGE in the United States is about 0.34 as measured by Chetty et al. (2014). Fan et al. (Forthcoming) also shows that the "Great Gatsby Curve" result (Krueger, 2012), a positive correlation between the intergenerational elasticity and income inequality, holds in time series data from China. Consistent with my hypothesis, they find that a rising outflow of migrants is negatively correlated with intergenerational income persistence at the provincial level, suggesting that relaxing internal migration frictions may help improve intergenerational mobility. They also document that government expenditure on education has reduced mobility across generations. This result could also be coming from the *hukou* policy, since households with different *hukou* status have differential access to these government resources.

Altogether, the results of Fan et al. (Forthcoming) suggest that intergenerational mobility is a pressing policy issue for the Chinese economy. The fact that there has been an increase over time in intergenerational persistence in China but not in other countries such as the United States suggests that differences in institutions and policies could potentially explain both the level and trend of intergenerational mobility. The *hukou* system is potentially a contributing factor.

2.3 Related Literature

My work is closely related to a growing literature that studies endogenous child human capital investment by parents and its implications for intergenerational persistence. The idea that parental investments and genetic inheritance of ability are linked to intergenerational transmission of income started with the seminal work of Becker and Tomes (1979), who used a simple two-period overlapping generations model to demonstrate how various factors drive the transmission of earnings and wealth across generations. Since then, many studies have extended the Becker-Tomes framework to derive richer implications. One prominent theme is the role of credit constraints facing both parents and young adults in skill investment (Carneiro and Heckman, 2002; Keane and Wolpin, 1997; Keane and Wolpin, 2001; Caucutt and Lochner, 2019a; Caucutt and Lochner, 2019b). Restuccia and Urrutia (2004) expands the Becker-Tomes model to distinguish between early-age and late human capital investment. The idea is that the returns to education depend crucially on prior investments at the earlier stages of childhood. Agents have less incentive to make human capital investments in later stages of life if they have not invested in the earlier years. This dynamic complementarity in skill formation is emphasized in Cunha and Heckman (2007), who stress that early childhood human capital investment is especially important. Yum (2016) and Lee and Seshadri (2019) develop models with even longer horizons to investigate more closely how earnings over the life-cycle interact with parental human capital investment decisions at various stages of childhood. Lastly, the literature has also studied how income taxation and education policies can mitigate inequality persistence across generations (Abbott et al., 2019; Lee and Seshadri, 2019; Daruich, 2019). In particular, Daruich (2019) studies the interactions between educational policies, taxation, endogenous human capital investment, inequality, and growth in a general equilibrium framework.

My work extends previous studies to study how inter-regional migration decisions interact with other forces affecting human capital investment. Migration introduces opportunities for upward income mobility as agents can relocate for jobs offering higher earnings. At the same time, the decision to migrate may affect parental time investments in children. In the extreme case when migrants are separated from their children, the loss in time investments may be very costly, especially in light of Agostinelli and Sorrenti (2018), who find that time and monetary investments in children are complementary.

Incorporating migration decisions into the Becker-Tomes framework resembles the strand of literature that studies neighborhood choices, human capital accumulation, and intergenerational mobility. Fogli and Guerrieri (2019) argues that in the presence of local skill spillovers, residential segregation exerts downward pressure on social mobility, as households with low income tend to be located together, leading to lower skill formation. Zheng and Graham (2019) uses a model with an endogenous housing market in which public education expenditure is funded by local property taxes, and shows that neighborhoods with higher housing prices also tend to provide higher quality schooling that is only affordable to households with high income, thereby generating intergenerational income persistence. My work differs from these two papers in the sense that they focus on residential location decisions within a region, whereas I consider the decision to migrate across regions with different wages and income processes. Furthermore, I focus explicitly on the restrictions imposed by the *hukou* policy, whereas previous work implicitly models housing costs as a barrier to migration, as pointed out by Garriga et al. (2017). The closest study to mine is the concurring work of Sieg et al. (2020), who estimate a structural spatial model to quantify the impact of *hukou* restrictions on intergeneraional mobility and education investment, with a particular focus on fiscal externalities. Their mechanism is similar to mine, but the modelling framework is different as I consider a richer life-cycle setup with general equilibrium. My model also allows for circular migration whereas theirs feature a one-time migration decision.

My work also contributes to the literature that studies the implications of migration restrictions for productivity, inequality, and welfare. Also in the context of China, Fan (2019) finds using a spatial trade model that eliminating the mobility frictions stipulated by the *hukou* system could greatly reduce aggregate inequality by reducing income inequality across regions. Tombe and Zhu (2019) finds that a complete removal of mobility costs in China would raise real GDP per capita by 13% and welfare by 46%. Their paper considers a feature of the *hukou* system that is overlooked by Fan (2019), which is the lack of private land ownership and functioning land rental markets in rural areas. As mentioned, under the *hukou* system, rural households lose ownership of their land when they migrate permanently, and the absence of a complete rental market makes it costly for households to migrate even temporarily. This aspect of the *hukou* system is also studied in Ngai et al. (2018) and Adamopoulos et al. (2017), who study the labor misallocation effects stemming from this dimension of the policy. Using a dynamic spatial equilibrium framework, You and Wu (2019) find that removing *hukou*-related restrictions would increase GDP and welfare by 15%-30%, consistent with the estimates of Tombe and Zhu (2019). In addition to evaluating economy-wide welfare changes, You and Wu (2019) also considers the distributional impacts between natives and migrants as well as across regions stemming from congestion externalities and fiscal effects. My contribution to this literature is that

I emphasize a dynamic cost that works through human capital accumulation across generations, instead of focusing on the static costs of mobility frictions. To the extent that mobility frictions suppress human capital accumulation, they can affect the growth rate of the economy by lowering aggregate skill.

Instead of modelling migration frictions as a sunk cost of moving, I incorporate more subtle mobility costs such as adverse impacts on child development and uninsurable idiosyncratic risks after moving. Also, I allow for return or circular migration, in which agents can move across regions at every stage in life. In this regard, my work relates to the literature on migration decisions and migrant behavior. My model of migration is similar to Lagakos et al. (2018), who study the determinants of rural-urban migration and the effects of migration subsidies using a model of migration calibrated using RCT evidence from Bangladesh. In their model, rural workers can either migrate temporarily to insure against bad income shocks at home or migrate permanently to take advantage of higher paying jobs in urban areas. However, their focus is on the role of experience for migration decisions as well as how subsidies can help overcome migration barriers. My focus on return migration links my work to the literature studying how return migration interacts with the initial migration choice as well as migrant behavior (Dustmann and Görlach (2016) provides a thorough review). Some findings from this literature that relate to my work concern the implications of migration for saving behavior and human capital accumulation. Firstly, when migration is expected to be temporary and when skill is more valued in the destination region, migrants may have a lower incentive to accumulate human capital, since the lifetime return is diminished, considering that the migrant will eventually return to her original location where the return to skill is lower (Navarro and Zhou, 2018,Dustmann and Glitz, 2011). Also, uncertainty and lack of risk insurance may induce precautionary saving motives among migrants, particularly in the context of China, due to the social insurance features of the *hukou* system. Dustmann (1997) shows that migration and saving decisions are determined jointly in the presence of income uncertainty. Consistent with Dustmann's predictions, Tan, Yu and Rao (2014) shows that Chinese migrant households exhibit a higher savings rate than non-migrant households and that the saving rate is positively correlated with the intention to return home in order to smooth lifetime consumption. My model features borrowing constraints and uninsurable idiosyncratic income risk à la Huggett (1996) and Aiyagari (1994), thereby generating precautionary saving motives for migrants.

Finally, my work is related to empirical studies on the effects of parental migration on child outcomes (for instance: Chen, 2013, Meyerhoefer and Chen, 2011, Meng and Yamauchi, 2017, Zhang et al., 2014) as well as the determinants of left-behind children (Zhong, 2019 and Chen et al., 2019). It is also related to empirical studies on skill formation (Agostinelli and Sorrenti, 2018, Cunha and Heckman, 2007, Cunha et al., 2010). I will rely on these estimates to discipline my model.

2.4 Model

2.4.1 General Environment

There are two regions in the economy: urban (U) and rural (R). The rural region hires unskilled labor to produce an agricultural good. The urban region hires skilled labor to produce a non-agricultural good. The two goods are then aggregated into a final consumption bundle consumed by all agents in the economy. In the absence of trade costs, agents consume the same bundle and face identical good prices regardless of location. The details of the production structure will be presented after the discussion of the household problem.

The household side of the economy follows a life-cycle overlapping-generations structure with 4 periods. There are two types of agents: urban hukou holders and rural hukou holders, each with unit mass. Rural hukou holders have the option to move between the two regions in the economy. Meanwhile, urban hukou holders always stay in region U and do not make migration decisions. This simplifying assumption reflects the fact that most internal migration flows in China are from rural areas to cities, and that farmland is difficult to acquire due to lack of private ownership and an incomplete rental market.

Rural hukou holders begin their adult life in region R. They work and choose how much to consume/save, in addition to whether or not to migrate in the next period. In the second period, they continue to receive income and make consumption-saving decisions. In addition, they choose where to locate themselves in the next period when they will have a child deterministically. In the third period, agents who are now parents choose how much to consume, save, and invest in their children's human capital. Furthermore, parents decide how much inter-vivos transfers to give to their children. The newborn generation becomes independent immediately after the third period of the parents' lives. Therefore, any parent-child interaction occurs only in the third period of life. In the final period, agents cease working and consume all their remaining wealth.

Urban hukou holders face the same timeline and make the same decisions as rural hukou holders, with the only difference being the absence of locational decisions.

The income that agents receive from working depends their location. The income of an agent working in region R is the product of a common component w_R and an idiosyncratic component f_i that varies across individuals. This idiosyncratic component can be interpreted as an endowment of unskilled (agricultural) labor. The value of f_i is fixed throughout an individual's life but is imperfectly passed through generations according to an AR(1) process:

$$log(f_{i,t+1}) = \rho_f \cdot log(f_{i,t}) + \epsilon^f_{i,t+1} \tag{2.1}$$

where $\epsilon^f_{i,t}$ is drawn from a normal distribution $N(\mu_f, \sigma_f)$ and $\rho_f \in (0, 1)$.

An agent working in the urban area earns $w_U \cdot h_i \cdot s_{it}$. The first term w_U is the urban wage per unit of skilled labor supplied. The second term h_i is the skilled labor endowment of the worker, which is determined by parental investments in their children's human capital. The human capital endowment is known to agents prior to any decisions being made and is fixed over the lifetime. The last term s_{it} represents an idiosyncratic stochastic labor endowment shock which follows an AR(1) process:

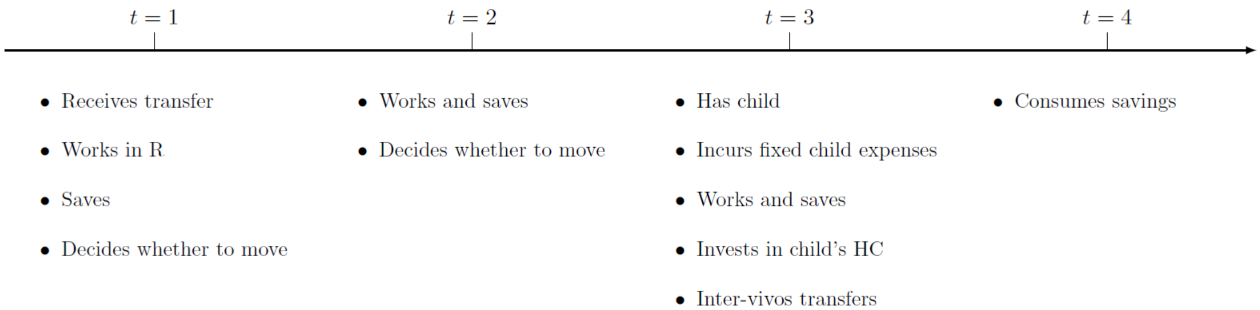
$$log(s_{i,t+1}) = \rho_s \cdot log(s_{i,t}) + \epsilon^s_{i,t+1} \tag{2.2}$$

where $\epsilon^s_{i,t}$ is drawn from a normal distribution $N(\mu_s, \sigma_s)$ and $\rho_s \in (0, 1)$.

The realization of s_{it} is not known to a rural individual when she makes her migration decision. If she decides to move from R to U, she draws her s_{it} from its stationary distribution. The evolution of the shock thereafter follows the AR(1) process. The same goes for urban hukou holders. Upon beginning their working lives, they draw their values of s_1 from the unconditional distribution, after which the evolution follows the stochastic process.

Finally, there is a local government in each region which collects proportional labor income tax (τ) to fund public education. Local governments tax all workers working in their region regardless of hukou status. However, public education is only provided to those with the local hukou. In other words, rural migrant workers in region U pay taxes to the urban government but

Figure 2.1: Model Timeline: Rural Hukou Holder



do not have access to public education in return. There are no fiscal transfers between the local governments and both governments run a balanced budget.

2.4.2 Problem of a Rural Hukou Holder

2.4.2.1 Period 1

A rural hukou holder begins her adult working life in the rural region. The agent has initial assets a_1 transferred inter-vivos from the parent. She observes her values of h_i and f_i, earns labor income, and makes two decisions: (1) how much to consume/save and (2) whether or not to move to U in the next period.

Agents have the option to save in a risk-free asset that pays an interest rate of r, issued by a body external to this economy with a perfectly elastic demand. In addition, agents are not allowed to borrow, i.e. $a_2 \geq 0$. If the agent wishes to move to U at the beginning of the next period, a migration cost M_{RU} must be paid in the current period.

Formally, the problem of a rural agent in period 1 can be formulated as follows (where $L_{t+1} = \{U, R\}$ denotes the locational choice for next period):

$$V_{R1}(a_i, f_i, h_i) = \max_{c_1, a_2, L_2} \frac{c_1^{1-\nu} - 1}{1-\nu} + \beta \cdot \max\Big\{\mathbb{E}\big[V_{U2}(s_2, a_2, h_i, f_i)\big], V_{R2}(a_2, h_i, f_i)\Big\} \tag{2.3}$$

subject to the budget constraint (where $\mathbb{I}$ denotes an indicator function):

$$c_1 + a_2 - M_{RU} \cdot \mathbb{I}_{\{L_2=U\}} = (1-\tau_R) \cdot w_R \cdot f_i + a_1 \tag{2.4}$$

and a borrowing constraint $a_2 \geq 0$.

If the agent chooses $L_2 = U$, the idiosyncratic income shock s_2 is drawn from the stationary distribution of the process $f_0(s)$.

2.4.2.2 Period 2

Agents begin the second period in the location chosen in the previous period. They work and earn income and again make two decisions: (1) how much to consume/save and (2) where to locate next period.

The migration decision in period 2 is different from that in period 1. This is because the agents will have a child deterministically in the third period, and they must decide in the current period the joint locational problem of themselves and the child. In this period, there are three possible options. Firstly, they can choose to live in the urban region with the child ($L = (U, U)$). Alternatively, they can choose to live in the rural region with the child ($L = (R, R)$). The last option is to leave the child behind in R while the agent works in U ($L = (U, R)$). Different arrangements have different implications for child-rearing expenses and the cost of human capital investment, as well as the skill formation of the child, which will be discussed in more detail in the next subsection. In addition, parents incur a utility cost Θ from leaving the child behind.

Again, we assume that migration is subject to a fixed migration cost (M_{RU} for moving from the rural region to the urban region, M_{UR} for the other way around) and that there is a borrowing constraint ($a_3 \geq 0$).

The formal problem of an agent living in U in period 2 is then:

$$V_{U2}(s_2, a_2, h_i, f_i) = \max_{c_2, a_3, L_3} \frac{c_2^{1-\nu} - 1}{1 - \nu} + \beta \cdot \max\Big\{\mathbb{E}\big[V_{UU3}(s_3, a_3, h_i, f_i)\big], \mathbb{E}\big[V_{UR3}(s_3, a_3, h_i, f_i)\big], V_{RR3}(a_3, h_i, f_i)\Big\} \quad (2.5)$$

subject to:

$$c_2 + a_3 - M_{UR} \cdot \mathbb{I}_{\{L_2=(R,R)\}} = (1 - \tau_U) \cdot w_U \cdot h_i \cdot s_2 + (1 + r) \cdot a_2 \quad (2.6)$$

The formal problem of an agent living in R in period 2 is:

$$V_{R2}(s_2, a_2, h_i, f_i) = \max_{c_2, a_3, L_3} \frac{c_2^{1-\nu} - 1}{1 - \nu} + \beta \cdot \max\Big\{\mathbb{E}\big[V_{UU3}(s_3, a_3, h_i, f_i)\big], \mathbb{E}\big[V_{UR3}(s_3, a_3, h_i, f_i)\big], V_{RR3}(a_3, h_i, f_i)\Big\} \quad (2.7)$$

subject to:

$$c_2 + a_3 - M_{RU} \cdot \mathbb{I}_{\{L_2=(U,U)\cup(U,R)\}} = (w_R \cdot f_i) + (1+r) \cdot a_2 \tag{2.8}$$

2.4.2.3 Period 3

In period 3, agents have a child deterministically. Child-rearing expenses are modeled as fixed expenses (E) that differ across regions for rural *hukou* workers. More specifically, child-rearing expenses are higher in the urban region than in the rural region ($E_U > E_R$), to reflect the fact that rural migrants lack access to health services and schooling in the urban region.[9] In this period, in addition to making consumption-saving and migration decisions, parents also choose the amount of investment e in the child's human capital. The cost per unit of investment (p_e) also differs by region, with human capital investments being more expensive in region U, also reflecting the features of the *hukou* policy.

Following the skill formation literature (Agostinelli and Sorrenti (2018)), skill formation requires two inputs: time and monetary investments. More specifically, monetary investment is composed of a public component G and a private component e. Similar to this literature, I assume that the two inputs enter a CES human capital production function:

$$h_{child} = \bar{h} \cdot (T^\rho + (G+e)^\rho)^{\frac{\gamma}{\rho}} \tag{2.9}$$

Also following the literature, I assume public and private investments to be perfect substitutes.

Access to G depends on hukou status. Rural hukou workers who migrate with the child

[9] Urban *hukou* workers face a fixed expense of $E_{URB} = E_R$

to region U are barred from local urban public education, and as a result $G = 0$. Children that are left behind in region R can enroll in rural public education, therefore $G = G_R$. Similarly, workers who remain in region R with their children also have $G = G_R$.

$T = \{T_l, T_h\}$ represents parental time investment, which depends on the locational choice made by the parent. If a parent resides in the same location as the child, $T = T_h > T_l$. Otherwise if the parent leaves the child behind in region R, $T = T_l$.[10]

To generate incentives for parental investment in children, I assume that parents are altruistic in the sense that the lifetime value function of the child $V_1(f_i, h_i)$ enters in the parents' objective function with a weight $\psi \in (0, 1]$.

The formal problem pertaining to an agent living together with the child in region U is:

$$V_{UU3} = \max_{c_3, a_4, e, L_4, a_{child}} \frac{c_3^{1-\nu} - 1}{1 - \nu} + \beta \cdot \max\Big\{V_{U4}(a_4), V_{R4}(a_4)\Big\} + \psi \cdot V_{R1}(a_{child}, f_{child}, h_{child}) \quad (2.10)$$

subject to:

$$c_3 + a_4 + M_{UR} \cdot \mathbb{I}_{\{L_4 = R\}} = (1 - \tau_U) \cdot (w_U \cdot h_i \cdot s_3) + (1 + r) \cdot a_3 - E_U - p_{eU} \cdot e - a_{child} \quad (2.11)$$

$$h_{child} = \bar{h} \cdot (T_h^{\rho} + (G + e)^{\rho})^{\frac{\gamma}{\rho}} \quad (2.12)$$

[10] It is assumed that the amount of time investment depends only on the living arrangement made by the parent as the model does not feature endogenous labor supply decisions. To assess this assumption, I check if time spent on children differ for migrant workers who moved with their children from the general population. Using the RUMiC 2008 sample, I found that migrants who live in the cities with their children spend on average 4.4 hours per day with their children (include care-taking, playing, and studying). Because this variable does not feature in the non-migrant sample, I checked the 2018 China Time Use Survey published by the Chinese Bureau of Statistics and found that on average, parents spend 188 minutes taking care of their children and 92 minutes accompanying the child's study.

$$f_{child} = \rho_f \cdot f_i + \epsilon^f \tag{2.13}$$

The formal problem pertaining to an agent living together with the child in region R is:

$$V_{RR3} = \max_{c_3,a_4,e,L_4,a_{child}} \frac{c_3^{1-\nu} - 1}{1-\nu} + \beta \cdot \max\Big\{V_{U4}(a_4), V_{R4}(a_4)\Big\} + \psi \cdot V_{R1}(a_{child}, f_{child}, h_{child}) \tag{2.14}$$

subject to:

$$c_3 + a_4 + M_{RU} \cdot \mathbb{I}_{\{L_4=U\}} = (1-\tau_R) \cdot (w_R \cdot f_i) + (1+r) \cdot a_3 - E_R - p_{eR} \cdot e - a_{child} \tag{2.15}$$

$$h_{child} = \bar{h} \cdot (T_h^\rho + (G_R + e)^\rho)^{\frac{\gamma}{\rho}} \tag{2.16}$$

$$f_{child} = \rho_f \cdot f_i + \epsilon^f \tag{2.17}$$

And lastly, for an agent working in U while leaving the child behind in R:

$$V_{UR3} = \max_{c_3,a_4,e,L_4,,a_{child}} \frac{c_3^{1-\nu} - 1}{1-\nu} - \Theta + \beta \cdot \max\Big\{V_{U4}(a_4), V_{R4}(a_4)\Big\} + \psi \cdot V_{R1}(a_{child}, f_{child}, h_{child}) \tag{2.18}$$

subject to:

$$c_3 + a_4 + M_{UR} \cdot \mathbb{I}_{\{L_4=R\}} = (1-\tau_U) \cdot (w_U \cdot h_i \cdot s_3) + (1+r) \cdot a_3 - E_R - p_{eR} \cdot e - a_{child} \tag{2.19}$$

$$h_{child} = \bar{h} \cdot (T_l^\rho + (G + e)^\rho)^{\frac{\gamma}{\rho}} \tag{2.20}$$

$$f_{child} = \rho_f \cdot f_i + \epsilon^f \tag{2.21}$$

2.4.2.4 Period 4

Agents stop working in the fourth period, which is also the final period of life. In this period, agents receive a pension P_l from their location of hukou registration $\{R, U\}$ funded by the respective local government, regardless of where they choose to live, which the agents consume in entirety alongside all remaining assets.

As a result, the value functions in the last period are:

$$V_{R4}\left(a_4\right) = \frac{c_4^{1-\nu} - 1}{1-\nu} \tag{2.22}$$

where

$$c_4 = P_R + (1+r) \cdot a_4 \tag{2.23}$$

$$V_{U4}\left(a_4\right) = \frac{c_4^{1-\nu} - 1}{1-\nu} \tag{2.24}$$

where

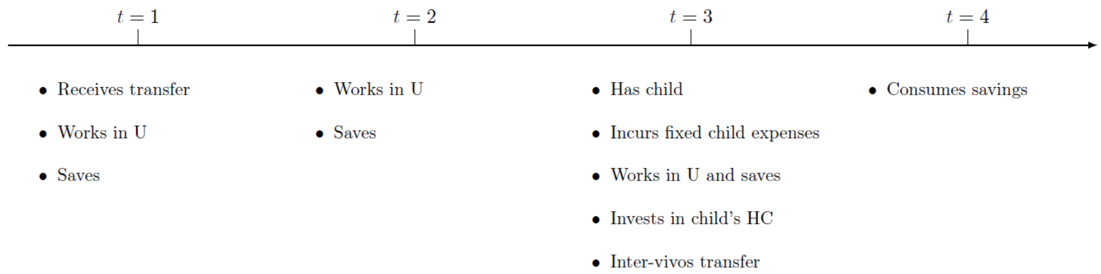
$$c_4 = P_R + (1+r) \cdot a_4 \tag{2.25}$$

2.4.3 Problem of Urban Hukou Holders

Urban hukou holders also live for four periods. Unlike rural hukou holders, they spend their entire lives in region U and do not make migration decisions.

In the first period of life, an urban hukou holder solves the following problem. Upon

Figure 2.2: Model Timeline: Urban Hukou Holder



drawing her value of s_1 from the unconditional distribution of the AR(1) process, she works, earns labor income, and chooses how much to consume/save.

$$V_1(s_1, a_1, h_i) = \max_{c_1, a_2} \frac{c_1^{1-\nu} - 1}{1-\nu} + \beta \cdot \mathbb{E}V_2(s_2, a_2, h_i) \tag{2.26}$$

subject to:

$$c_1 + a_2 = (1 - \tau_U) \cdot w_U \cdot h_i \cdot s_1 + a_1 \tag{2.27}$$

$$a_2 \geq 0 \tag{2.28}$$

$$s_2 = \rho_s \cdot s_1 + \epsilon_2^s \tag{2.29}$$

In the second period, she continues to work and make consumption-saving decisions:

$$V_2(s_2, a_2, h_i) = \max_{c_2, a_3} \frac{c_2^{1-\nu} - 1}{1-\nu} + \beta \cdot \mathbb{E}V_3(s_3, a_3, h_i) \tag{2.30}$$

subject to:

$$c_1 + a_3 = (1 - \tau_U) \cdot w_U \cdot h_i \cdot s_2 + (1 + r) \cdot a_2 \tag{2.31}$$

$$a_3 \geq 0 \tag{2.32}$$

$$s_3 = \rho_s \cdot s_2 + \epsilon_3^s \tag{2.33}$$

In period 3, the worker has a child deterministically. In addition to consumption-saving decision, she chooses how much to invest in the child's human capital and the amount of inter-vivos transfers to the child. As in the case with rural hukou holders, the parent incurs a fixed child-rearing expense of $E_{URB} = E_R$ and faces price $p_{e,URB} = p_{eR}$ per unit of education investment.

$$V_3(s_3, a_3, h_i) = \max_{c_3, a_4} \frac{c_1^{1-\nu} - 1}{1-\nu} + \beta \cdot \mathbb{E}V_4(a_4) + \psi \cdot \mathbb{E}\big[V_1(s_{1,child}, a_{child}, h_{child})\big] \tag{2.34}$$

subject to:

$$c_3 + a_4 + a_{child}\cdot = (1-\tau_U) \cdot w_U \cdot h_i \cdot s_3 + (1+r) \cdot a_3 - p_{e,URB} \cdot e - E_{URB} \tag{2.35}$$

$$a_4 \geq 0 \text{ , } a_{child} \geq 0 \tag{2.36}$$

$$h_{child} = \bar{h} \cdot (T_h^{\rho} + (G_U + e)^{\rho})^{\frac{\gamma}{\rho}} \tag{2.37}$$

And in the last period, urban agents receive a pension from the urban government and consumes all remaining assets.

$$V_{U4}\left(a_4\right) = \frac{c_4^{1-\nu} - 1}{1-\nu} \tag{2.38}$$

where

$$c_4 = P_U + (1+r) \cdot a_4 \tag{2.39}$$

2.4.4 Production

There is a representative firm in the rural region that hires unskilled labor H_R to produce an agricultural good Y_R using a linear production technology:

$$Y_R = H_R \tag{2.40}$$

Similarly, a representative firm in the urban region hires skilled labor H_U to produce a non-agricultural good Y_U according to:

$$Y_U = H_U \tag{2.41}$$

The two goods are aggregated into a final consumption bundle Y according to the following CES function:

$$Y = A \cdot [SY_R^{\alpha} + (1 - S)Y_U^{\alpha}]^{\frac{1}{\alpha}} \tag{2.42}$$

The simple structure above is identical to a setting where a representative firm combines rural and urban labor to produce the final good. This representative firm solves the following problem:

$$\max_{H_R, H_U} \qquad Y = [SH_R^{\alpha} + (1 - S)H_U^{\alpha}]^{\frac{1}{\alpha}} - w_R \cdot H_R - w_U \cdot H_U \tag{2.43}$$

where the price of the final good is normalized to 1.

Optimality conditions require that the firm sets the marginal product of each input equal to the marginal cost:

$$A \cdot S \cdot H_R^{\alpha-1} \cdot [sH_R^{\alpha} + (1 - s)H_U^{\alpha}]^{\frac{1-\alpha}{\alpha}} = w_R \tag{2.44}$$

$$A \cdot (1-S) \cdot H_U^{\alpha-1} \cdot [sH_R^{\alpha} + (1-s)H_U^{\alpha}]^{\frac{1-\alpha}{\alpha}} = w_U \tag{2.45}$$

Note that the parameter s as well as the labor allocation across regions govern the urban-rural wage premium in this economy, since the following condition must hold in equilibrium:

$$\frac{w_U}{w_R} = \frac{1-S}{S}\left(\frac{H_R}{H_U}\right)^{1-\alpha} \tag{2.46}$$

2.4.5 Government

In each region, there is a local government that collects proportional labor income tax at rate τ from all workers working in the region, where the urban tax rate τ_U may differ from the rural tax rate τ_R. Government tax revenue is then used to provide public education and pension benefits. There is no revenue sharing between the governments and each runs a balanced budget.

The public education production function exhibits decreasing returns to scale and takes the form:

$$\bar{G}_i = \ln{(1 + x_i)} \tag{2.47}$$

where x_i is the total tax revenue net of pension expenses. To be clear, the local governments choose the tax rate and the level of pension benefits, with the restriction that total pension expense must be less than total tax revenue. The remaining surplus is spent on public education provision.

$\bar{G}_i$ is split equally among all children residing in the region who have hukou registration. The value of G_R that enters the skill formation function of parents with locational choice $\{RR, UR\}$ is then equal to $\bar{G}_R$ divided by the total mass of left-behind children and children residing in region R with their parents. Since children of rural migrant workers in region U do not have access to public education, the value of G_U for children with urban hukou is equal to $\bar{G}_U$, as there is a unit mass of native urban workers.

Incorporating governments and public education generates a potential downside to migration. First of all, rural-urban migration shrinks the tax base of the rural region, which reduces public education provision in the rural region. Although the tax revenue of the urban government would increase, it faces capacity constraints and overcrowding of schools due to the decreasing returns property of the public education production function.

Furthermore, granting migrant workers access to public schooling in the urban region would hurt urban hukou holders. This is because the public education resources would now be shared among a larger pool of children, whereas with the hukou restrictions, migrant workers pay taxes to the urban government but do not receive benefits in return.

2.4.6 Definition of Equilibrium

The stationary equilibrium of this model is a collection of:

1. Wages: $\left\{w_R, w_U\right\}$

2. Value functions: $\left\{V_{R1}, V_{R2}, V_{U2}, V_{UU3}\ V_{RR3}, V_{UR3}, V_{U4}, V_{R4}\right\}$

3. Saving decision rules : $\left\{a^*_{U1}, a^*_{R1}, a^*_{U2}, a^*_{R2}, a^*_{UU3}, a^*_{RR3}, a^*_{UR3}\right\}$

4. Inter-vivos transfers: $\left\{B^*\right\}$

5. Human capital investment decision rules: $\left\{e^*_{UU3}, e^*_{RR3}, e^*_{UR3}\right\}$

6. Locational decision rules: $\left\{L^*_{R1}, L^*_{U2}, L^*_{R2}, L^*_{UU3}, L^*_{RR3}, L^*_{UR3}\right\}$

7. Distributions of rural hukou holders:

$$\Big\{\lambda^*_{R1}(h,f,a),\lambda^*_{R2}(h,f,a),\lambda^*_{U2}(h,f,a,s),\lambda^*_{UU3}(h,f,a),\lambda^*_{UR3}(h,f,a,s),$$
$$\lambda^*_{RR3}(h,f,a),\lambda^*_{R4}(h,f,a),\lambda^*_{U4}(h,f,a)\Big\}$$

8. Distributions of urban hukou holders: $\Big\{\lambda^*_1(s,a,h),\lambda^*_2(s,a,h),\lambda^*_3(s,a,h),\lambda^*_4(s,a,h)\Big\}$

9. Public education provisions: $\big\{G_U, G_R\big\}$

such that:

1. Given the external parameters, agents solve their optimization problems in all periods, and the policy functions are based on optimal choices

2. Given wages, the two labor markets clear. In other words, the firm's labor demands are equal to the agents' supply of labor, as implied by their locational choices.

   - In particular, at any point in time the total labor supply in the rural region is:

$$H_R = \int f\cdot\Big(d\lambda^*_{R1}(h,f,a) + d\lambda^*_{R2}(h,f,a) + d\lambda^*_{RR3}(h,f,a)\Big) \tag{2.48}$$

   - The total labor supply in the urban region is:

$$H_U = \int (s\cdot h)\cdot d\lambda^*_{U2}(h,f,a,s) + \int (s\cdot h)\cdot d\lambda^*_{UU3}(h,f,a,s) + \int (s\cdot h)\cdot d\lambda^*_{UR3}(h,f,a,s)$$
$$+ \int (s\cdot h)\cdot d\lambda^*_1(h,a,s) + \int (s\cdot h)\cdot d\lambda^*_2(h,a,s) + \int (s\cdot h)\cdot d\lambda^*_3(h,a,s) \tag{2.49}$$

3. The economy-wide resource constraint holds, i.e. the sum of total consumption, expenditures on education investments and migration is equal to the sum of total output and income on assets.

4. The distributions of rural hukou holders over state variables are invariant over time, with transition functions $\Gamma(\cdot)$; in particular:

$$\lambda^*_{R1}(a,h,f) = \int \Gamma \cdot d\lambda^*_{R1}(a,h,f) \tag{2.50}$$

   Since the distributions over (s,a,h,f) in the later periods of life are determined by the initial distribution over (a,h,f) at birth, stationarity of $\lambda^*_{R1}(a,h,f)$ implies that all other distributions are invariant.

5. The distributions of urban hukou holders over state variables are invariant over time, with transition functions $\Gamma(\cdot)$; in particular:

$$\lambda^*_{1}(s,a,h) = \int \Gamma \cdot d\lambda^*_{1}(s,a,h) \tag{2.51}$$

6. Government budget constraints hold:

$$\bar{G}_U = \ln\left(1 + \tau_U \cdot w_U \cdot H_U - P_U\right) \tag{2.52}$$

$$\bar{G}_R = \ln\left(1 + \tau_R \cdot w_R \cdot H_R - P_R\right) \tag{2.53}$$

2.4.7 Solution Algorithm

This model can be solved using the following algorithm:

1. Discretize the AR(1) process for $s_{i,t}$ using the method of Tauchen (1986)

2. **Outer Loop.** Guess a vector of wages $(\tilde{w_R}, \tilde{w_U})$ and a vector of public education provisions (G_U, G_R)

3. **Inner Loop I (Rural Hukou Holders).** Make an initial guess for the value functions in period 1: $\tilde{V}_{R1}(a, h, f)$.

   (a) Using the fact that the agents consume all their pension income and wealth in period 4, compute the value functions V_{R4} and V_{U4}.

   (b) Given V_{R4}, V_{U4} and the guess $\tilde{V}_{R1}$, solve the agent's problem in period 3 for each locational pair $\{(U, U), (U, R), (R, R)\}$, which yields the value functions V_{UU3}, V_{RR3}, V_{UR3}, as well as the decision rules for savings, human capital investment, inter-vivos transfers, and migration.

   (c) Solve the agent's maximization problem from period 2 given the period 3 value functions and obtain the value functions V_{U2} and V_{R2}

   (d) Solve the agent's maximization problem in period 1 yielding V_{R1}.

   (e) **Update Value Function Guess.** Calculate the differences between the solution V_{R1} and the guess $\tilde{V}_{R1}$: $\left|\tilde{V} - V\right|$. If the difference is larger than a pre-set tolerance level, use the solution V_{R1} as the new guess and repeat until the value function guess and implied value function are sufficiently close.

4. **Inner Loop I (Urban Hukou Holders).** Make an initial guess for the value functions in period 1: $\tilde{V}_1(s, a, h)$.

   (a) Using the fact that the agents consume all their pension income and wealth in period 4, compute the value functions V_4.

   (b) Given V_4 and the guess $\tilde{V}_1$, solve the agent's problem in period 3, which yields the value functions V_3, as well as the decision rules for savings, human capital investment, and inter-vivos transfers.

   (c) Solve the agent's maximization problem from period 2 given the period 3 value functions and obtain the value functions V_2.

   (d) Solve the agent's maximization problem in period 1 yielding V_1.

   (e) **Update Value Function Guess.** Calculate the differences between the solution V_1 and the guess $\tilde{V}_1$: $\left|\tilde{V} - V\right|$. If the difference is larger than a pre-set tolerance level, use the solution V_{R1} as the new guess and repeat until convergence is reached.

5. **Inner Loop II.** Solve for the stationary distribution. First, initialize with a guess of initial distributions $\tilde{mu}_R(f, a, h)$ for rural hukou holders and $\tilde{mu}_U(s, a, h)$ for urban hukou holders.

   (a) Use the decision rules from period 1 to compute the distribution of assets, locations, and productivities in the second period.

   (b) Using the distribution over state variables in the second period, along with the decision rules, compute the distributions for period 3.

(c) Using the decision rules and distributions over states in period 3, compute the distributions of the new cohort of rural and urban hukou holders.

(d) **Update Distribution Guess.** Compute the difference between the new distributions obtained from the previous step with the initial guess. If the difference $|\tilde{\mu} - \mu|$ is larger than a pre-set tolerance level, repeat using the implied distribution as the new guess until convergence.

6. **Update Wage Guess.** Check if the labor market clears for the two regions. If not, pick a new guess for wages and repeat the steps above. To do this, plug in the total supply of each type of labor into the firm's optimality conditions and recover the implied wages that clear the market. If the implied wages differ from the initial guess from step (2), update the new guess as a convex combination of the last guess and the implied values until the difference between the two are sufficiently small.

7. **Update Public Education Guess.** Using the governments' budget constraints, compute the implied values of $\bar{G}_U$ and $\bar{G}_R$. Then, use the mass of children in each region to recover G_U and G_R. If the implied values disagree with the initial guess, update the new guess as a convex combination of the last guess and the implied values until the difference between the two are sufficiently close.

2.5 Quantitative Analysis

2.5.1 Calibration

2.5.1.1 Externally Calibrated and Policy Parameters

Table 2.2: Externally Calibrated Parameters

Parameter	Description	Value	Source
r	Interest Rate	0.07	Chamon et al. (2013)
β	Discount Rate	0.935	$\beta = \frac{1}{1+r}$
ν	Risk Aversion	2	Literature
ρ	Subs b/w time and money	-0.35	Daruich (2019)
γ	Skill production function elasticity	0.7	Zheng and Graham (2019)
$\bar{h}$	Skill production function shifter	4.55	Normalization
ρ_s	Urban income persistence	0.5	Yu and Zhu (2013)
μ_s	Urban income shock mean	-0.125	$\mu_s = -\sigma_s^2/(2 \cdot (1-\rho_s^2))$
σ_s	Urban income shock S.D.	0.5	2008 RUMiC
μ_f	Rural income shock mean	-0.245	$\mu_f = -\sigma_f^2/(2 \cdot (1-\rho_f^2))$
σ_f	Rural income shock S.D.	0.3	2008 RUMiC
α	Subs b/w rural and urban labor	0.30	Katz and Murphy (1992)
A	Final good production function shifter	1.25	Normalization
M_{RU}	Mig. Cost	0.2	10% of rural income
M_{UR}	Mig. Cost	0.1	$0.5 \times M_{RU}$
p_e^R	Edu. Cost	1.5	Normalization
p_e^U	Edu. Cost	2.2	Migrant child tuition (RUMiC 2008)
E_R	Fixed Expenses	0	Normalization
τ_R	Labor Tax	0.05	Gov. edu spending (% of GDP)
τ_U	Labor Tax	0.05	Gov. edu spending (% of GDP)
P_R	Pension in R	0	
P_U	Pension in U	0	

When calibrating the model, I take each model period to be 5 years. The annual real interest rate in China over the period 1990-2006 is 1.4% (Chamon et al., 2013), which maps to an interest rate of about 7% in the model. The risk aversion parameter in the household utility function is set to 2, within the common range of values used in the literature.

The parameter ρ, which governs the substitution between parental time and monetary investments in skill formation, is set to -0.35 following Daruich (2019), so that time and monetary investments are complementary. The elasticity of children's human capital with respect to total monetary investment, γ, is 0.7 following Zheng and Graham (2019). The constant shifter $\bar{h}$ normalizes the mean human capital level to 1. The substitution between urban and rural labor α in the final goods production function is set to 0.3, which translates to the estimated elasticity from Katz and Murphy (1992) between skilled and unskilled labor.

To calibrate the processes of s (urban idiosyncratic labor endowment) and f (agricultural productivity), I rely on Chinese household survey data. Using survey data in China covering 1990-2009, Yu and Zhu (2013) estimate the annual household income persistence to be 0.8. In the model, I choose ρ_s to be 0.6, which translates to an annual persistence of 0.9, slightly higher than the empirical estimate. The standard deviation of the urban idiosyncratic income shock ϵ^s is chosen to match the cross-sectional income distribution in the urban region. To do this, I fit a log-normal distribution to the cross-sectional urban household income distribution from the 1995 and 2007 samples of the Chinese Household Income Project (CHIP). In both years, I find the standard deviation to be around 0.7.[11] As a result, I set the standard deviation of ϵ^s (σ^s) to be 0.5 so that the unconditional distribution of s matches the cross-sectional data. The mean of ϵ^s is chosen such that the mean of the unconditional distribution is 1.

I use the same procedure to calibrate the standard deviation of ϵ^f. Again, I fit a log-normal distribution to the cross-sectional distribution of household income from the rural samples of the 1995 and 2008 waves of CHIP. In both cases, I find the standard deviation to be around 0.7.[12] With the appropriate conversion, the standard deviation of ϵ^f (σ^f) is set to 0.3.

[11] Similar results are obtained when conditioning on the level of education.

[12] For robustness, I repeat the exercise for size of landholdings in 1995 CHIP and again find the cross-sectional standard deviation to be around 0.7.

The use of cross-sectional data to calibrate the income processes is a limitation of the current calibration. Access to life-cycle earning profiles would allow me to discipline the processes better such as the China Family Panel Study.

The fixed migration cost for moving from R to U, M_{RU}, is set to match the estimates of internal migration costs in the literature. Using Indonesian data, Bryan and Morten (2019) finds that rural-urban migration cost to be 10% of rural income. This translates to a value of 0.2 in the model. Migration costs for flows in the reverse direction M_{UR} is set to be half of M_{RU} to reflect the fact that it is less costly for migrants to return to their hometowns, a common feature in the domestic migration literature (Lagakos et al., 2018).

The cost per unit of monetary investment e in human capital for local hukou holders is normalized to 1.5. In the Longitudinal Survey on Rural Urban Migration in China (RUMiC), respondents are asked about total schooling expenses as well as the fraction of additional fees they had to pay due to lack of hukou registration.[13] Using the data from 2008, I find that on average, migrant parents whose children are enrolled in schools in the city had to pay 80% more than parents with local hukou. Therefore, I set the value of p_{eU} to be 2.2. The fixed child expense for children living in their hukou locations is normalized to one and the expense pertaining to migrant workers is to be internally calibrated.

Lastly, the current analysis abstracts from pension benefits so that P_R and P_U are both set to zero. As such, government tax revenues are spent entirely on providing public education. As a result, I set the labor tax rates in both regions to be 0.05, which is the government expenditure share on education as a percentage of GDP according to the Ministry of Education of China.

[13]The RUMiC consists of three surveys, one for rural households, one for urban households, and one for migrant workers in the cities without legal hukou status. In each survey, respondents are surveyed on information relating to demographics, health, education, income and expenditure. The migrant worker survey includes additional questions, for example migration history, future migration plans, and more detailed information regarding their children.

2.5.1.2 Policy Parameters and Internally Calibrated Parameters

Table 2.3: Internally Calibrated Parameters

Parameter	Description	Value	Target	Data	Model
ρ^f	Rural Inc. Persist.	0.9	IGE (rank-rank)	0.32	0.34
S	Prod. Func.	0.193	Wage Premium	2.25	2.25
Θ	Separation Cost	1.9	Fraction left-behind	0.33	0.34
ψ	Altruism	0.56	Education Exp. Share	0.18	0.38
E_U	Fixed Expenses	2.8	Fraction of child migrants	0.20	0.20

This leaves 5 parameters to be internally calibrated: the intergenerational persistence of rural income (ρ^f), the weight of the rural good in the aggregate production function (S), the utility cost from being separated from child (Θ), the strength of altruistic preferences (ψ), and the fixed child expenses faced by migrant workers in region U (E_U). For each parameter, I assign a data moment for which the parameter is closely linked to. The intergenerational persistence of f (ρ^f) is calibrated to match the intergenerational income elasticity in China, which Fan et al. (Forthcoming) estimate to be 0.44. S is chosen to match the observed urban-rural wage premium of 2.25 (Sicular et al. (2007)). The utility cost of separation Θ is chosen to match the incidence of left-behind children, who account for one-third of all rural children as of 2010. The strength of altruism ψ is used to match the share of parental income spent on children's educational expenses, which I find to be around 0.18 in the 2008 RUMiC.[14] Finally, I normalize the fixed child expenses for hukou holders to be 0 and internally calibrate the fixed expenses E_U faced by migrant workers in region U to match the fraction of migrant workers who move to the cities with their children. Among migrant workers who have a school-age child in the 2008 RUMiC sample, about 38%

[14] In each of the three samples (rural, urban, migrant), I keep only households with children currently in school. I then compute educational expenses on the child (including tuition fees, kindergarten fees, extra-curricular classes, registration fees, board fees) as a share of total household income. The mean values are 0.18, 0.21, and 0.16 for the rural, urban, and migrant samples respectively.

live with their children in the cities while the remaining 62% have left their children behind. Given that left-behind children account for 33% of all rural children, the implied fraction of rural families with children who live with their children in the city is 0.20 ($0.33 \times \frac{0.38}{0.62}$).

Table (2.3) shows the comparisons between the model moments and target data moments. In general, the data moments generated under the current calibration match the target moments quite well, with the exception of the share of income spent on education - the model implied value of 0.38 is more than twice the value of 0.18 observed in the data. This can be attributed to the fact educational expenses are concentrated in one five year period in the model, where in reality they are spread out over more than ten years. Taking this into account, one could argue that the data moment should be adjusted upward by at least a factor of two. In that case, the model moment is not too far away from the data.

2.5.2 Quantitative Results

2.5.2.1 Incidence of Migration

Table (2.4) below shows the migration decisions of rural hukou holders in the baseline economy.

Table 2.4: Migration Patterns in the Baseline Economy

Locational Choice	Fraction of Rural Hukou Holders
$U2$	86%
$R2$	14%
$UU3$	20%
$RR3$	47%
$UR3$	33%

Although the incidence of migration in period 3 is used in the internal calibration, the fraction of rural workers moving in period 2 is untargeted and therefore can be used for model validation. In the model, close to 86% of workers choose to migrate to the urban region in period 2. Using the rural sample of the 2009 RUMiC, I find that 56% of respondents under the age of 35 - who were 15 years or older when internal migration restrictions began to be relaxed in 1990 - have migrated at least once in their lives. This number is a lower bound, as only workers who have migrated and returned to their hometown are included in the sample. However, judging by other metrics, the model still predicts too much rural-urban migration. For instance, aggregate statistics reported by the Chinese National Bureau of Statistics show that only 27% of rural hukou holders are working in the cities, while the model predicts that 47% of rural agents are working in the urban region at any given point in time.

There are two major reasons for the excess migration produced in the model relative to the data. First of all, there may be additional costs to migration that are not captured in the model, such as labor market discrimination faced by migrant workers. One way to incorporate such effects in the model is to introduce an urban wage penalty for migrant workers, where the size of the penalty can be estimated with a regression of labor income on observable characteristics alongside a dummy indicator for migrant workers, although more sophiscated methods are needed to circumvent selection bias and unobserved ability. The second reason for excessive migration in the model is that the model abstracts from migration decisions in later periods of life when the propensity to migrate is lower and return migration is more common, as the focus of the paper is to study parental investment decisions. These return migration flows, due to old age and other reasons, are not captured well in the model.

Large return migration flows are another feature observed in Chinese internal migration. In the baseline economy, 38% of workers who move in period 2 return in the next period. Because of the prevalence of circular and temporary migration, as well as poor documentation, it is difficult to precisely measure the amount of return migration flow. To get a rough sense of the size of return migration in the data, I turn to the migrant sample of the 2008 RUMiC. I find that among migrant worker respondents, the average duration of a migration spell is 7.6 years, while 75% of respondents stay away from home for than 10 years.[15]

2.5.2.2 Intergenerational Persistence

To measure the extent of intergenerational income persistence in the model, I simulate two generations from the model and examine what fraction of the variation in an agent's lifetime income can be accounted for by variation in the parent's lifetime income. Specifically, I estimate the following regression using simulated data from the model:

$$\ln\left(Y_{i,p}\right) = \beta_0 + \beta_1 \cdot \ln\left(Y_{i,c}\right) + \epsilon_i \tag{2.54}$$

where $Y_{i,p}$ and $Y_{i,c}$ represent the lifetime labor earnings of the parent and child in household i respectively.

In addition, I estimate a rank-rank specification:

$$\text{Rank}_{i,c} = \gamma_0 + \gamma_1 \cdot \left(\text{Rank}_{i,p}\right) + \epsilon_i \tag{2.55}$$

where $\text{Rank}_{i,p}$ and $\text{Rank}_{i,c}$ represent the rank in the income distribution of the parent and child in household i respectively.

[15] These findings are consistent with those reported in Meng (2012) and Zhong (2019).

The coefficient β_1, commonly referred to as the intergenerational income elasticity, measures the fraction of lifetime income transmitted from one generation to another. The coefficient γ_1, referred to as the rank-rank slope, is an alternative measure of intergenerational mobility that identifies the correlation between parent's and children's positions in their respective income distributions. Chetty et al. (2014) has found the log-log specification to yield unstable estimates of mobility due to the nonlinear relationship between log parent and log child income and sensitivity to data treatment. As a result, in the calibration, I use the rank-rank estimate for China from Fan et al. (Forthcoming) as the target.

	Model (Rural Hukou Holders)	Data (Fan et al., Forthcoming)
β_1	0.35	0.44
γ_1	0.34	0.32

Table 2.5: Measures of Intergenerational Mobility: Model and Data

As shown in the table above, the estimated coefficients from model simulations match well the empirical estimates for China. In the counterfactual analysis, I will show how intergenerational mobility varies with migration restrictions in the model.

2.5.2.3 Decisions of Households

Migration Decisions in Period 1.

The following diagram depicts the migration decision rule at the end of period 1 for a rural worker with no initial assets, as a function of (f, h). Agents in the blue (dark) region choose to remain in region R in the next period and those in the yellow (light) region choose to migrate to the city.

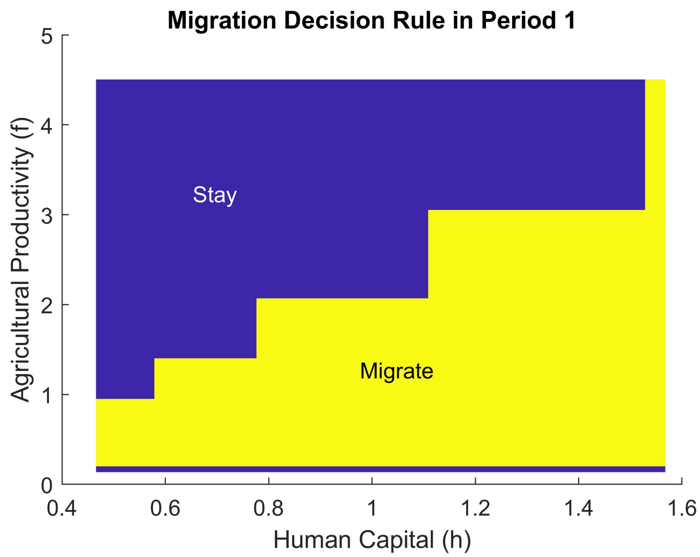


Figure 2.3: Migration Decision in Period 1, $a_1 = 0$

For each level of f, there exists a cutoff $\bar{h}$ above which the agent finds it optimal to migrate to the urban region, with this cutoff increasing in the value of f. This result is due to the fact that h and f control the effective income gap between the two regions.[16] Conditional on human capital endowment, an agent with higher f has less to gain from moving to region U relative to another agent with a lower f.

Notice that agents with the lowest values of f are forced to stay in the rural region as they cannot afford to pay the fixed migration cost. If migration were costless, these agents would find it optimal to move to the city for higher earning potential.

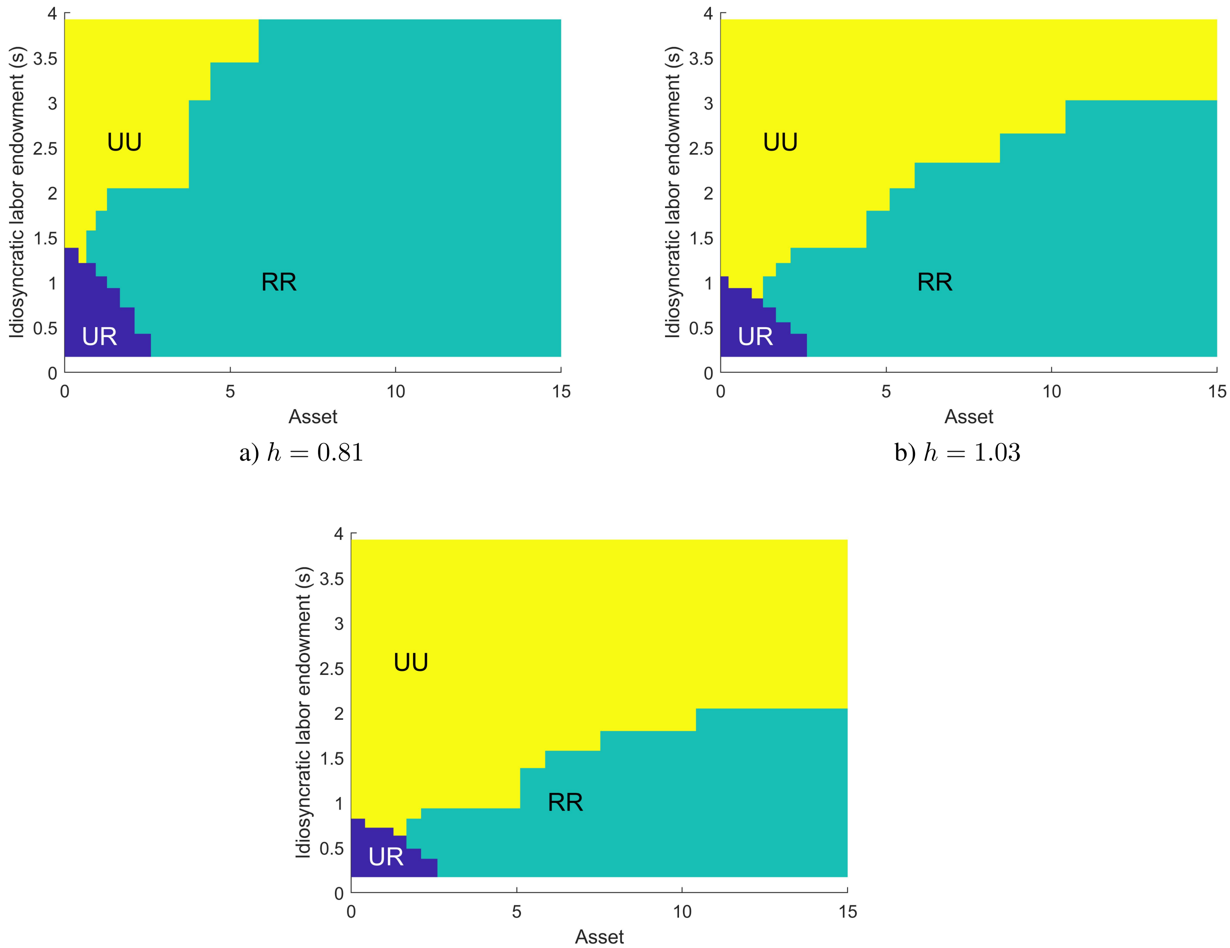
This prediction of the model is consistent with the empirical evidence that high skilled workers are more likely to migrate to cities whereas individuals with high farm income and large land holdings are less likely to migrate.

Migration Decision in Period 2.

[16]In region R, agents earn $(1 - \tau_R) \cdot w_R \cdot f$ with certainty. The expected earning from migrating to region U is $(1 - \tau_U) \cdot w_U \cdot h$.

Under the current model setup, all agents who decide to stay in region R in period 2 will also choose to work and raise their children in the rural area in period 3. This is because the expected benefit of migrating to the urban area with the child is lower than the expected benefit of migrating in the first period due to higher child-rearing expenses in cities. These agents also do not choose to work in the urban area while leaving their children behind because the expected gain in income is low for them, and the possibility of higher earnings to use for education investment does not offset the impact of parental absence on time investments in the children.

The following figures show the migration decision rules over (s_2, a_2) of an agent located in region U at $t = 2$ with $f = 0.3$, for 3 different values of h.



a) $h = 0.81$

b) $h = 1.03$

c) $h = 1.24$

Figure 2.4: Migration Decisions of Agents Conditional on $U2$

In the yellow (lightest) region, the optimal decision is to work in U with the child. In the green (less light) region, the optimal decision is to return to R and raise the child there. In the blue (darkest) region, the optimal choice is to work in the city and leave the child behind in the rural region.

Observe that agents with high realizations of the idiosyncratic labor endowment shock s_2 are more likely to raise their children in the urban region. This is because they can expect to continue earning high income in the urban region in the next period. Meanwhile, the cutoff in s

for choosing UU is decreasing in the level of human capital h, as workers with higher skill have higher earning ability in the urban region and a stronger incentive to remain there.

Workers with high levels of assets are more inclined to return to the rural area to raise their chilren. This is because agents with large wealth can continue to consume more despite the income loss associated with return migration. Additionally, migrant workers do not have access to urban public education. By migrating back to the rural region, workers can invest more in the children's skill.

The poorest workers - both in terms of income and wealth - are the ones who choose to leave their children behind. Although leaving children behind leads to lower parental time investment in children's skill, these workers find it worthwhile to do so because their own marginal consumption needs exceed the marginal benefit of investment in their children. Note that with sufficient amount of wealth, leaving children behind is no longer the optimal option, because such workers have enough resources for their own consumption as well as investing in their children. Since time and money are complementary inputs in skill formation, given sufficient monetary resources, it is not optimal to be separated from the children, as parental absence reduce the returns to monetary investments.

The fact that the poorest migrant workers are the most inclined to leave their children behind is consistent with migration patterns in China. Also, it is the core driver of intergenerational income persistence in the model, as these left behind children accumulate less skill and grow up to be poor. These children are also more likely to find themselves in a position to leave their own offspring behind, thereby generating a poverty trap across generations.

2.6 Counterfactual Analysis

To study the impact of the *hukou* policy on intergenerational income persistence and human capital accumulation. I conduct two counterfactual exercises in which hukou restrictions are relaxed or removed.

In the first exercise, the differential in education cost between urban hukou holders and migrant workers in region U is eliminated ($p_{eU} = p_{eR} = p_{e,URB} = 1.5$). In the second exercise, the education cost differential is completely eliminated and the additional fixed child-rearing expenses for rural migrants are cut in half ($E_U = 1.4$). In both exercises, migrant workers have access to public education G_U in the urban region.

Comparisons between the baseline economy and the two counterfactual economies are shown in the following table:

Table 2.6: Comparison: Baseline vs. Counterfactual

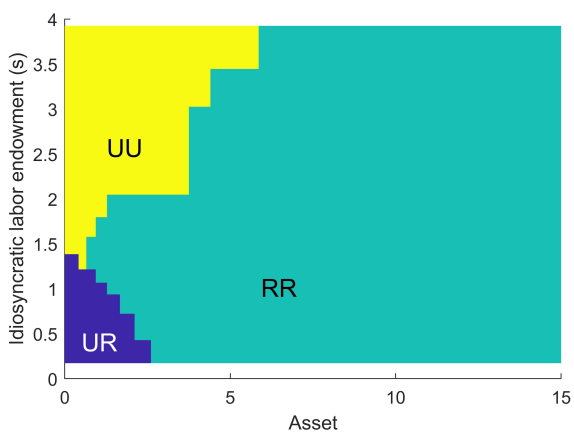
Object	Baseline	CF(1)	CF(2)
Fraction $U2$	85%	88%	87%
Fraction $R2$	15%	12%	13%
Fraction $UU3$	20%	34%	71%
Fraction $RR3$	47%	36%	20%
Fraction $UR3$	33%	30%	9%
w_U/w_R	2.25	2.09	1.91
IGE (rank-rank)	0.34	0.29	0.27
Mean h: Rural Hukou Holders	0.97	1.05	1.12
Mean h: Urban Hukou Holders	1.28	1.25	1.23
Aggregate h	2.25	2.30	2.35
Aggregate Output	5.13	5.3	5.51

In counterfactual economy (1), migrant workers gain access to public education G_U and do not face larger costs for additional education relative to hukou holders. As a result, it is more attractive for rural hukou holders to work in the cities. Compared to the baseline economy, there is more rural-urban migration in period 2 and less return migration in the following period. The fraction of workers choosing to migrate with their children increases drastically from 20% to 34%, as migrants are now eligible to tap into urban public education. However, the high fixed child-rearing costs in region U continue to force poorer migrant workers to leave their children behind, though there are fewer left-behind children relative to the baseline. Improved access to public education, lower education investment costs, and a reduction in the share of left-behind children raise the average human capital level of rural hukou holders from 0.97 to 1.05. The aggregate stock of human capital in counterfactual economy (1) increases slightly by 2.25%, and aggregate output increases by 3%.

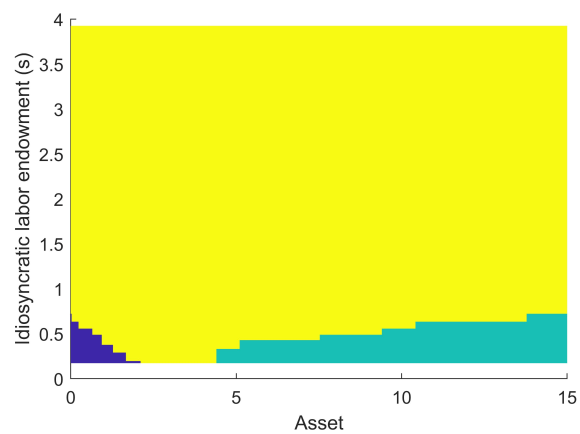
In counterfactual economy (2), the fixed child-rearing expenses faced by migrant workers are cut in half, which makes it less costly for workers to bring their children to the urban region. As a result, the incidence of left-behind children is reduced from 34% in the baseline economy to 9% in the counterfactual economy. There is also much less return migration. Relative to the baseline, the average human capital level among rural hukou holders increases by 8% to 1.12. The aggregate stock of human capital increases by 4% and aggregate output increases by 7%.

There are several reasons why human capital stock among rural hukou holders increases with relaxation of migration frictions. First, relaxing migration frictions directly favors human capital investment as it is less costly for migrant workers to live in the cities and invest in their children's human capital. In addition, a reduction in the probability of return migration increases incentives to accumulate human capital because the benefits can be reaped over a longer period of time (recall that the rural region only hires unskilled labor). Reduction in the incidence of left-behind children also plays an important role. Not only do left-behind children develop less skill due to lack of parental time investment, parents of left-behind children also have less incentive to invest money in their children's human capital, because the complementarity between time and monetary investment in the skill formation function means that low time investment reduces the marginal return to monetary investment. Since workers who leave behind their children tend to have low income, their marginal utility of consumption is high, making education investment less attractive.

In terms of intergenerational mobility, the counterfactual exercise shows that relaxing migration frictions leads to a reduction in the transmission of income from parents to children - with the rank-rank slope dropping from 0.34 in the baseline to 0.27 in counterfactual economy (2). This is mainly due to fewer children being left behind by migrant workers. Table (2.5) compares the migration decision rules, between the benchmark economy and counterfactual (2), of a rural hukou holder working in region U in period 2. As seen, the state space over which leaving children behind is optimal shrinks dramatically across the two scenarios.



a) Baseline Economy



b) Counterfactual Economy (2)

Figure 2.5: Migration Decisions of Agents Conditional on $U2$; baseline vs counterfactual

However, the gains from relaxing migration frictions are not universal. Although there is an improvement in human capital among rural hukou holders, the average level of human capital among urban hukou holders decreases. This is because the influx of migrant children crowds out public education resources that were previously exclusively enjoyed by natives. Even though migrant workers contribute to the urban government's tax revenue, decreasing returns to scale in public education production imply that the increase in revenue does not compensate for the inflow of migrant children. As for children growing up in region R, the average human capital level is actually higher in the counterfactual economies, although the rural tax base shrinks from larger out-migration. This is because there are also fewer left-behind children who compete for local rural resources, and also because there is a better selection of migrants. When there are migration frictions, some workers who would otherwise earn more in the urban region might be discouraged from moving. Reduction in frictions allows these workers to move, thereby increasing the average agricultural productivity among workers who choose to stay.

2.7 Discussion and Future Avenues

The model in this paper is able to generate migration and human capital investment decisions that largely align with the empirical evidence. First of all, agents' propensity to migrate is increasing in the urban-rural income gap. In particular, those with low agricultural productivity or land holdings and with high skill are more likely to migrate from the rural region to the urban region. Secondly, the differential child-related expenses across regions stipulated by the *hukou* system generate return migration flows as well as a large incidence of "left-behind" children, which are key features of Chinese internal migration. As in the standard literature on parental human capital investments, credit constraints and income histories are key determinants of the monetary investments made by parents in their children. Due to complementarity of time and money in human capital accumulation, conditional on disposable income, absent parents invest fewer monetary resources in their children's human capital, which further magnifies the impact of parental absence on the development of children.

The model lends support to the argument that the migration frictions resulting from the *hukou* policy may have a negative impact on intergenerational mobility as well as human capital accumulation. First, on the margin the migration frictions reduce rural-urban migration, which suppresses upward mobility by reducing the opportunity for agents to take advantage of regional income gaps. Even for those who decide to move, the high costs of child-rearing in the urban area for migrant workers lead to return migration flows, again taking away the ability to earn the higher urban wage. As a result, rural agents, who are on average poor, do not have as many resources to invest in their children's human capital relative to the richer urban agents, thereby reducing intergenerational income mobility. More crucially, the model predicts that "left-behind" children accumulate lower levels of human capital due to the absence of parental time investment and reduced monetary inputs. Since parents of "left-behind" children are generally poor, this is a key driving force of intergenerational immobility.

One limitation of the current model is that the expenditure share of educational expenses and the size of inter-vivos transfers from parents to children are far from what is observed in the data. The average worker in the model spends close to 40% of their income on skill investment in their children and the mean size of inter-vivos transfer is 9% of period 3 wealth. In the data, education investment is only about 18% of total expenditure, whereas average net inter-vivos transfer from parents to their children are negative as observed in the RUMiC survey. As such, the model needs to be modified and disciplined better to match the patterns of education investment and inter-vivos transfers. One extension in this direction is to allow for additional working periods following parenthood rather than going straight from parenthood into retirement, so that education investments are made in early life and monetary bequests are left in later stages of life. This is more consistent with the literature on parental skill investment and should bring the model closer to the data. Furthermore, adding further periods would make it possible to study whether migration frictions are related to lifetime saving patterns.

Another direction to pursue is to validate the model regarding the elasticity of migration with respect to geographic frictions. As mentioned in earlier sections, different parts of China have relaxed migration restrictions by varying degrees over the past two decades. Fan (2019) constructed a regional index for hukou reforms between the years 2000 and 2010, which is a score based on the level of difficulty in obtaining local hukou. Regression analysis shows that the hukou reforms, as measured by his index, have increased migration inflows in the destination region. Therefore, one way to validate the model is to conduct counterfactual exercises that can be mapped to the hukou reform index and to compare the migration elasticity with respect to frictions in the data with the elasticity implied by the model.

Lastly, the model framework can be used to investigate the optimal design of policy. For instance, in the counterfactual exercise, urban hukou holders suffer from the lifting of hukou restrictions due to overcrowding of public resources. Therefore, it is an interesting to see the effects of different policy mixes on both aggregate welfare and distributional consequences.

2.8 Conclusion

Masked behind the rapid economic growth of China is ever increasing income inequality both across households and regions. In addition, empirical estimates of intergenerational income transmission show that social mobility across generations in China is lower than in many developed countries and declining over time. In this paper, I offer a potential contributing factor to inequality and its intergenerational persistence – the *hukou* system. The *hukou* system is a unique institutional feature in China that poses restrictions on internal mobility, mainly by blocking access of migrants to public services and programs. For instance, workers are not eligible for services such as healthcare and education outside their location of residency registration.

Not only do these mobility frictions reduce overall internal migration flows, they also influence the decisions taken by households with regard to saving, return migration, and human capital investment. A remarkable phenomenon caused by the *hukou* system is the prevalence of "left-behind children" – children that get left behind in the villages while the parents go work in the cities. In 2010, 61 million children were left behind in villages, about a third of rural children. Inspired by the empirical literature showing negative impacts of parental absence on child outcomes, both in terms of educational attainment and health, I propose that migration frictions instituted by the *hukou* system contribute to the high intergenerational income persistence observed in China. The core idea is that poor households have fewer resources to invest in their children and are more likely to leave their children behind in villages while they seek employment in the cities. In turn these children grow up to earn low incomes and continue the vicious cycle of poverty.

To investigate my hypothesis, I present a two-region heterogeneous-agent overlapping generations model in which agents endogenously choose where to work, whether to migrate with their children, and how much to invest in their children. My model predicts that the institutional features of the *hukou* system make it difficult for rural households to take advantage of the large rural-urban income gap, thereby reducing the resources available for human capital investments.

Furthermore, households on the left end of the income distribution are more likely to leave their children behind. At the same time, parents of left-behind children are less able and have less incentive to make monetary investments in their children's skill, hence contributing to low income mobility across generations. A counterfactual analysis, in which regional differences in child-rearing expenses and education costs are reduced suggests that removal of some *hukou* restrictions can increase the total human capital stock by up to 4% and reduce the intergenerational elasticity of income by up to 7 percentage points.

Chapter 3: Real Rigidities, Firm Dynamics, and Monetary Non-Neutrality: Role of Demand Shocks (Co-authored with Borağan Aruoba, Felipe Saffie, and Jonathan Willis)

3.1 Introduction

Modeling the response of output and prices to monetary policy shocks has been a long-term quest in the macroeconomics literature. Quantitative estimates consistently indicate that monetary policy shocks have persistent impacts on real output (Christiano et al., 1999; Ramey, 2016). When it comes to modeling, early studies (Akerlof and Yellen, 1985; Mankiw, 1985) propose that nominal rigidities, like a menu cost, can lead to real fluctuations. However, subsequent studies such as Golosov and Lucas (2007) show that a menu cost alone is insufficient to generate the observed non-neutrality from quantitative estimates. To this end, Ball and Romer (1990) demonstrate that real rigidities in combination with nominal rigidities can lead to sizable real effects of monetary policy.

Strategic complementarities in pricing represent one promising form of real rigidity used in the money non-neutrality literature. Kimball (1995) introduces a demand system where price elasticity of demand varies with relative price and quantity. Eichenbaum and Fisher (2007) and Smets and Wouters (2007) add strategic complementarities in the form of a Kimball demand system into dynamic stochastic general equilibrium models in order to match the estimated responses of real output to a nominal shock, while also being consistent with micro evidence on the frequency of price adjustment. However, Klenow and Willis (2016) find that a menu-cost model with a Kimball demand system that matches firm-level pricing moments requires idiosyncratic productivity shocks that are much larger than observed in the data.

Elsewhere in the economics literature, the use of strategic complementarities in models has been shown to be important for generating firm-level and aggregate dynamics needed to match the key features of the data. It is used extensively to study exchange-rate pass-through in international macroeconomics (Amiti et al., 2019; Berger and Vavra, 2019; Gopinath and Itskhoki, 2010), variable markups in firm dynamics and international trade (Arkolakis et al., 2019; Edmond et al., 2018), and inflationary dynamics and optimal monetary policy (Blanco, 2021; Coibion et al., 2012; Harding et al., 2022a,b).

In this paper, we propose a simple model featuring strategic complementarities that can generate monetary non-neutrality, while also remaining consistent with the micro and macro evidence. The baseline specification is a menu-cost model with a real rigidity in the form of a Kimball demand system, in which the elasticity of substitution between a given variety and other varieties is decreasing in the relative quantity consumed. The final important element of this model is the specification of two forms of idiosyncratic shocks: productivity and demand. Under a Kimball demand system, idiosyncratic demand influences the desired markups of firms through its effect on the demand elasticity. Previous studies, including Klenow and Willis (2016), exclusively focus on the role of idiosyncratic productivity shocks.[1] Despite its parsimony, the

[1] Burstein and Hellwig (2007) explore supply and demand shocks under a CES demand system with decreasing returns to scale technology.

model generates sizable non-neutrality of monetary policy with realistic shocks, while remaining consistent with micro pricing facts. Thus our model opens the door to consistently modeling real (investment, labor, and sales), nominal (pricing and non-neutrality), and competition (markup) dynamics in a parsimonious framework amenable to policy analysis.

The model is calibrated to match the frequency of price adjustment and the serial correlation and standard deviation of idiosyncratic productivity and demand shocks. For the latter productivity and demand moments, we use results from Foster et al. (2008) who point to important, and separate, roles for these shocks using data from Census of Manufactures.[2] We investigate the role of real rigidities by calibrating the model to two additional moments from Foster et al. (2008), that help disentangle demand from supply and discipline the degree of strategic complementarities. The first moment is the correlation between revenue-based total factor productivity (TFPR) and total factor productivity (TFP), or as it is sometimes refer to as, quantity-based TFP (TFPQ). The second moment is the correlation between firm price and TFPQ.

To illustrate the key ingredients of our baseline model relative to a model with CES preferences and TFP shocks, we calibrate four different specifications. The first version is a standard CES model without idiosyncratic demand shocks that is calibrated to match three pricing moments: the frequency of price changes, the fraction of positive changes, and the average size of price changes. The second version is the same CES model, but calibrated to match the autocorrelation and variance of idiosyncratic TFP from Foster et al. (2008) along with the frequency of price changes. The third version goes a step further by adding and calibrating idiosyncratic demand shocks to match the autocorrelation and variance of demand from Foster et al. (2008). The

[2]Recent papers that jointly study and model both productivity and supply shocks are Aruoba et al. (2022) and Carlsson et al. (2022)

first CES model is consistent with the standard procedure in the pricing literature, i.e., target pricing moments using exogenous shock processes. This first model matches pricing dynamics by construction, but it is inconsistent with firm-level shocks (Klenow and Willis, 2016). The two other CES models are by construction consistent with the idiosyncratic shock processes documented by Foster et al. (2008), but fail at replicating non-targeted pricing moments. The fourth model is our baseline specification featuring a Kimball demand system in combination with idiosyncratic demand and productivity shocks with the addition of Corr(TFPQ,TFPR) and Corr(TFPQ,P) as targeted moments. In addition to matching a broader set of targeted moments, our baseline model effortlessly matches three non-targeted pricing moments: the average size of a price change conditional on a change, the fraction of adjustments that are positive, and the dispersion of non-zero price changes. Furthermore, the model delivers a downward-sloping pricing hazard consistent with the data.

The incorporation of a Kimball demand system also allows our model to generate an untargeted cross-sectional markup distribution that closely resembles its empirical counterpart, which is impossible to achieve with CES demand. This is because the desired markup under CES is constant. Although nominal pricing rigidities can generate a non-degenerate distribution around the desired markup, such a model cannot replicate the large cross-sectional variance in the empirical markup distribution without unreasonably large nominal rigidities. In contrast, a Kimball demand system induces more dispersion in firm markups due to variable elasticities as well as the incorporation of demand shocks which affect the desired markup. Additionally, by the virtue of Kimball demand system, the calibrated model produces incomplete cost pass-through (pass-through of 38 percent), in line with the empirical literature.

Our framework overcomes the challenge of using a Kimball demand system in pricing models arising from inconsistency with micro evidence on price setting behavior by firms. The inclusion of both supply and demand shocks is instrumental to this success, through the expanded set of shocks that triggers firms to adjust prices and the magnitude of pass-through from underlying shocks into prices. In fact, a Kimball demand system generates a trade-off between the strength of strategic complementary and the pass-though of productivity shocks. Previous studies using a Kimball demand system included only idiosyncratic productivity shocks. In a typical calibration of the degree of strategic complementary, firms would only pass through 20 to 40 percent of productivity shocks into prices. Therefore, in order to generate price changes consistent with the micro evidence, extremely volatile productive shocks are needed to produce price changes that are aligned with the empirical literature. Our calibrated model implies that firms pass through approximately 62 percent of idiosyncratic demand shocks and 38 percent of productivity shocks into prices. With demand shocks also playing a role in the pricing decisions of firms, our model can generate a distribution of price changes that is similar to the data remaining consistent with micro evidence on the volatility of idiosyncratic shocks.

To study the degree of monetary non-neutrality generated by the model, we expand the framework to incorporate nominal expenditure shocks. On impact, approximately 82 percent of an increase in nominal expenditure is reflected in an increase in real output. The real output response decays with a half-life of 5 months, leaving a cumulative response of 0.40 before eventually fully dissipating after 20 months. The total response of real output is 4 times as large as a simple CES menu cost model without real rigidities calibrated to pricing moments. It is in the upper range of richer models that explore alternative sources of real rigidities without using micro-estimates for their shock processes.[3] The two key features generating this non-

[3]See Golosov and Lucas (2007), Gertler and Leahy (2008), Burstein and Hellwig (2007), Nakamura and

neutrality are the Kimball demand system and the presence of both idiosyncratic productivity and demand shocks. The former generates a strategic pricing complementarity under which the firm will temper its price adjustment in response to a shock because of an endogenous change in its desired markup, whereas the latter dampens the selection effect in price adjustments to an aggregate shock and results in larger real output responses.

Overall, our model, calibrated to micro evidence from both the firm-dynamics and the price-adjustment literatures, is able to produce results consistent with the evidence on monetary non-neutrality, cost pass-through, and markups. The combination of menu costs, a Kimball demand system, and idiosyncratic shocks for demand and productivity in the model produces cost pass-through in line with previous studies along with a distribution of markups that is similar to its non-targeted empirical counterpart. Regarding monetary non-neutrality, the findings demonstrate that a model featuring strategic complementarities can produce persistent real output responses to nominal expenditure shocks while also remaining consistent with the micro evidence. Thus, our parsimonious framework proposes a promising path to jointly model pricing and firm level dynamics.

The remainder of the paper is structured as follows. Section 3.2 introduces a quantitative menu-cost model with idiosyncratic productivity and demand augmented with a Kimball demand system and explores its theoretical properties. Section 3.3 presents the calibration of the model. Section 3.4 discusses the model's implications for non-targeted pricing moments, markup distribution, and non-neutrality of monetary shocks. Finally, Section 3.5 concludes.

Steinsson (2010), Vavra (2014), and Mongey (2021).

3.2 Menu Cost Model

We build a quantitative menu-cost model following Golosov and Lucas (2007). The model features a representative household, a representative final-good producer, and a continuum of monopolistically competitive intermediate-variety producers who face nominal pricing frictions.

3.2.1 Households

A representative household supplies labor to firms in exchange for wage payments, purchases a complete set of Arrow-Debreu securities, $\mathbf{B_{t+1}}$, and consumes a final good, C_t. It also owns all firms in the economy and receives all accrued profits. The representative household solves the following problem

$$\max_{C_t, h_t, \mathbf{B_{t+1}}} \quad \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\log\left(C_t\right) - \chi h_t\right] \tag{3.1}$$

subject to the budget constraint

$$P_t C_t + \mathbf{Q_t} \cdot \mathbf{B_{t+1}} \leq B_t + W_t h_t + \Pi_t, \tag{3.2}$$

where $\mathbf{Q_t}$ is a vector that contains the prices of the state-contingent securities, $\mathbf{B_{t+1}}$. B_t represents the payoff of the state-contingent security purchased in period $t-1$ that had a non-zero payoff in period t. P_t and W_t are the price of the final good and nominal wage, respectively, both of which are taken as given by the households. Π_t denotes the net dividends the household receives from the producers.

Household optimality requires

$$\frac{W_t}{P_t} = \chi C_t, \tag{3.3}$$

and we can also define the household's stochastic discount factor as

$$\Xi_{t,t+1} \equiv \beta \frac{C_t}{C_{t+1}}. \tag{3.4}$$

3.2.2 Producers

Production is carried out by a continuum of perfectly-competitive final-good producers, who purchase varieties of intermediate goods and sell a combined final good to the households. The intermediate-good producers are monopolistically competitive as the varieties they produce are not perfect substitutes.

3.2.2.1 Final Good Producers

A representative final-good firm combines intermediate varieties, y_t^i, to produce the final good, Y_t, using the Kimball (1995) aggregator. This aggregator is defined implicitly as

$$\int_0^1 G\left(\frac{n_t^i y_t^i}{Y_t}\right) di = 1 \tag{3.5}$$

where n_t^i represents an idiosyncratic variety-specific demand shifter. Following Dotsey and King (2005), we use the following specification for $G(.)$

$$G\left(\frac{n_t^i y_t^i}{Y_t}\right) = \frac{\omega}{1+\omega\psi}\left[(1+\psi)\frac{n_t^i y_t^i}{Y_t} - \psi\right]^{\frac{1+\omega\psi}{\omega(1+\psi)}} + 1 - \frac{\omega}{1+\omega\psi} \tag{3.6}$$

where (3.5) and (3.6) show the only two deviations from a textbook menu cost model: the introduction of the variety-specific demand shifters and the Kimball aggregator. This specification nests the familiar constant elasticity of substitution (CES) Dixit-Stiglitz aggregator when $\psi = 0$. When this is the case, the final good Y_t can be expressed explicitly as

$$Y_t = \left[\int_0^1 \left(n_t^i y_t^i \right)^{\frac{1}{\omega}} \right]^{\omega}, \tag{3.7}$$

where the price elasticity of demand is given by $\frac{\omega}{1-\omega}$, elasticity of substitution is given by $\frac{\omega}{\omega-1}$ and the gross markup is given by ω. When $\psi \neq 0$, price elasticity of demand, elasticity of substitution and desired markup are no longer constant. We discuss how Kimball aggregation affects firms' pricing decisions in Section 3.2.3.

Taking as given variety prices, p_t^i, as well as P_t, n_t^i and aggregate demand, Y_t, the representative final-good producer chooses y_t^i to maximize profits

$$\max_{y_t^i} \quad 1 - \int_0^1 \frac{p_t^i y_t^i}{P_t Y_t} di \quad \text{subject to} \int_0^1 G\left(\frac{n_t^i y_t^i}{Y_t} \right) di = 1. \tag{3.8}$$

The optimality condition of the final-good producer's maximization problem implicitly defines the demand function for each variety i

$$\frac{n_t^i y_t^i}{Y_t} = \frac{1}{1+\psi} \left[\left(\frac{p_t^i}{\lambda_t n_t^i P_t} \right)^{\frac{\omega(1+\psi)}{1-\omega}} + \psi \right], \tag{3.9}$$

where λ_t is the Lagrangian multiplier on the constraint in the optimization problem, which can be obtained by substituting (3.9) into (3.5) as

$$\lambda_t = \left[\int_0^1 \left(\frac{p_t^i}{n_t^i P_t} \right)^{\frac{1+\omega\psi}{1-\omega}} di \right]^{\frac{1-\omega}{1+\omega\psi}}. \tag{3.10}$$

The aggregate price index is derived from the zero-profit condition for the final-good producer as

$$P_t = \frac{1}{1+\psi} \left[\int_0^1 \left(\frac{p_t^i}{n_t^i} \right)^{\frac{1+\omega\psi}{1-\omega}} di \right]^{\frac{1-\omega}{1+\omega\psi}} + \frac{\psi}{1+\psi} \int_0^1 \frac{p_t^i}{n_t^i} di \tag{3.11}$$

3.2.2.2 Intermediate Variety Producers

There is a continuum of intermediate-good producers indexed by i, each producing a differentiated variety y_t^i. Intermediate producers are heterogeneous in their physical productivity, z_t^i, and face demand shocks for their variety, n_t^i. Production technology is linear with labor as the only input

$$y_t^i = z_t^i l_t^i \tag{3.12}$$

Idiosyncratic productivity, z_t^i, and idiosyncratic demand, n_t^i, evolve according to a VAR(1) process

$$\begin{pmatrix} \log(z_t^i) \\ \log(n_t^i) \end{pmatrix} = \begin{bmatrix} \rho_z & 0 \\ 0 & \rho_n \end{bmatrix} \begin{pmatrix} \log(z_{t-1}^i) \\ \log(n_{t-1}^i) \end{pmatrix} + u_t^i \text{ where } u_t^i \sim N \left(0, \begin{bmatrix} \sigma_z^2 & \sigma_{zn} \\ \sigma_{zn} & \sigma_n^2 \end{bmatrix} \right) \tag{3.13}$$

At the beginning of each period, intermediate-good producers decide whether or not to adjust their nominal prices and if so, by how much. Nominal price adjustments are subject to a fixed cost f in terms of labor. Given the demand schedule for individual varieties, the intermediate producers' gross profit when adjusting their price p is

$$\pi(p_t^i, z_t^i, n_t^i, \mathcal{S}_t) = \left(\frac{p_t^i}{P_t} - \frac{W_t}{z_t^i P_t}\right) \frac{Y_t}{n_t^i} \frac{1}{1+\psi} \left[\left(\frac{p_t^i}{\lambda_t n_t^i P_t} \right)^{\frac{\omega(1+\psi)}{1-\omega}} + \psi \right], \tag{3.14}$$

where $\mathcal{S}_t \equiv (P_t, W_t, Y_t, \lambda_t)$ collects all aggregate objects the firms need to know, and we assume that the firms know the law of motion for $\mathcal{S}_t$.

At the beginning of the period, each intermediate-good producer inherits their price from the previous period p_{t-1}^i. At that point, they choose whether or not to change their prices by solving the problem

$$V\left(p_{t-1}^i, z_t^i, n_t^i, \mathcal{S}_t\right) = \max \left[V_N\left(p_{t-1}^i, z_t^i, n_t^i, \mathcal{S}_t\right), V_A\left(z_t^i, n_t^i, \mathcal{S}_t\right) \right], \tag{3.15}$$

where $V_N(.)$ and $V_A(.)$ are the values for the firm if they do not change and change their prices, respectively.

The value of not adjusting is

$$V_N\left(p_{t-1}^i, z_t^i, n_t^i, \mathcal{S}_t\right) = \pi\left(p_{t-1}^i, z_t^i, n_t^i, \mathcal{S}_t\right) + \mathbb{E}_t\left[\Xi_{t,t+1} V\left(p_{t-1}^i, n_{t+1}^i, z_{t+1}^i, \mathcal{S}_{t+1}\right)\right], \tag{3.16}$$

which is equal to the flow profit evaluated at last period's price plus a continuation value. If the firm chooses to adjust its price, it pays the fixed price adjustment cost and chooses p_t^i to maximize the sum of current flow profit and the present discounted value of future profit

$$V_A\left(z_t^i, n_t^i, \mathcal{S}_t\right) = -f\frac{W_t}{P_t} + \max_{p_t^i} \left\{ \pi(p_t^i, z_t^i, n_t^i, \mathcal{S}_t) + \mathbb{E}_t\left[\Xi_{t,t+1} V\left(p_t^i, n_{t+1}^i, z_{t+1}^i, \mathcal{S}_{t+1}\right)\right] \right\}. \tag{3.17}$$

The intermediate-good producers solve this problem taking as given the laws of motion for the idiosyncratic state variables as in (3.13) and those for the aggregate variables in $\mathcal{S}_t$.

3.2.3 Kimball Aggregation and Pricing Decisions

The defining feature of the Kimball demand system is variable price elasticity. In particular, when $\psi < 0$, the price elasticity of demand becomes an increasing function of the relative price of the variety p/P, and a decreasing function of idiosyncratic demand, n_t^i, and the effective market share of the variety, ny/Y.[4] The non-constant price elasticity implied by the Kimball demand system has two important consequences. First of all, unlike the CES case where desired markups are constant and independent of idiosyncratic demand, variable price elasticity leads to variable desired markups. Moreover, the desired markup depends on both the firm's idiosyncratic productivity as well as demand. For example, a firm with lower costs or higher demand would choose to have a higher markup relative to an average firm. Second, the Kimball demand system creates strategic complementarities in pricing among firms, because a deviation from the aggregate price index is costly.

To demonstrate how productivity and demand shocks affect pricing decisions of firms under a Kimball demand system, it is instructive to set $f = 0$ and focus on the problem of an individual

[4] The price elasticity of demand is

$$\epsilon \equiv \frac{dy}{dp}\frac{p}{y} = \frac{\omega}{1-\omega}\frac{(1+\psi)\frac{ny}{Y} - \psi}{\frac{ny}{Y}} = \frac{\omega(1+\psi)}{1-\omega}\frac{\left(\frac{p}{\lambda nP}\right)^{\frac{\omega(1+\psi)}{1-\omega}}}{\left(\frac{p}{\lambda nP}\right)^{\frac{\omega(1+\psi)}{1-\omega}} + \psi}.$$

with $\frac{\partial\epsilon}{\partial(ny/Y)} = -\frac{\psi\omega}{\left(\frac{ny}{Y}\right)^2(\omega-1)}$ and $\frac{\partial\epsilon}{\partial n} = -\frac{\psi\omega Y}{n^2 y(\omega-1)}$ We can also compute the super-elasticity, defined as the elasticity of the demand elasticity with respect to price as

$$\gamma \equiv \frac{d\epsilon}{dp}\frac{p}{\epsilon} = \frac{\omega}{1-\omega}\cdot\frac{\psi(1+\psi)}{\left[\left(\frac{p}{\lambda nP}\right)^{\frac{\omega}{1-\omega}} + \psi\right]} = \frac{\omega}{1-\omega}\cdot\frac{\psi}{\frac{ny}{Y}}.$$

intermediate-good producer. The first-order condition to the static profit-maximization problem of an intermediate-good producer is

$$\left(\frac{p_i^*}{\lambda n_i P}\right)^{\frac{\omega(1+\psi)}{1-\omega}} \left[1 - \left(\frac{W}{z_i} - p_i^*\right)\left(\frac{\omega\left(1+\psi\right)}{1-\omega}\frac{1}{p_i^*}\right)\right] = -\psi \tag{3.18}$$

where p_i^* denotes the optimal price the firm chooses.

Log-linearizing (3.18) around a symmetric steady state and letting hatted variables denote log-deviations from the steady state yield the following expression for the optimal price

$$\hat{p}_i^* = \frac{\omega\psi}{\omega\psi - 1}\left(\hat{\lambda} + \hat{P} + \hat{n}_i\right) + \frac{1}{\omega\psi - 1}\hat{z}_i \tag{3.19}$$

Letting $\widehat{mc} \equiv 1/\hat{z}$ denote log-deviation in marginal cost, which is inversely proportional to productivity, the price elasticities with respect to cost and demand shocks, respectively, are given by

$$\frac{\partial \hat{p}_i^*}{\partial \widehat{mc}} = -\frac{1}{\omega\psi - 1} \tag{3.20}$$

$$\frac{\partial \hat{p}_i^*}{\partial \hat{n}_i} = \frac{\omega\psi}{\omega\psi - 1}, \tag{3.21}$$

In what follows we refer to these as cost and demand pass-through.[5]

When $\psi = 0$, cost pass-through is complete: price falls one to one with a positive z shock (negative $\widehat{mc}$ shock). However, when $\psi < 0$, the pass-through of a cost shock is incomplete. Variable demand elasticity is key to understanding this incomplete pass-through of cost. Consider a firm experiencing a positive productivity shock. As its marginal cost decreases, the firm finds

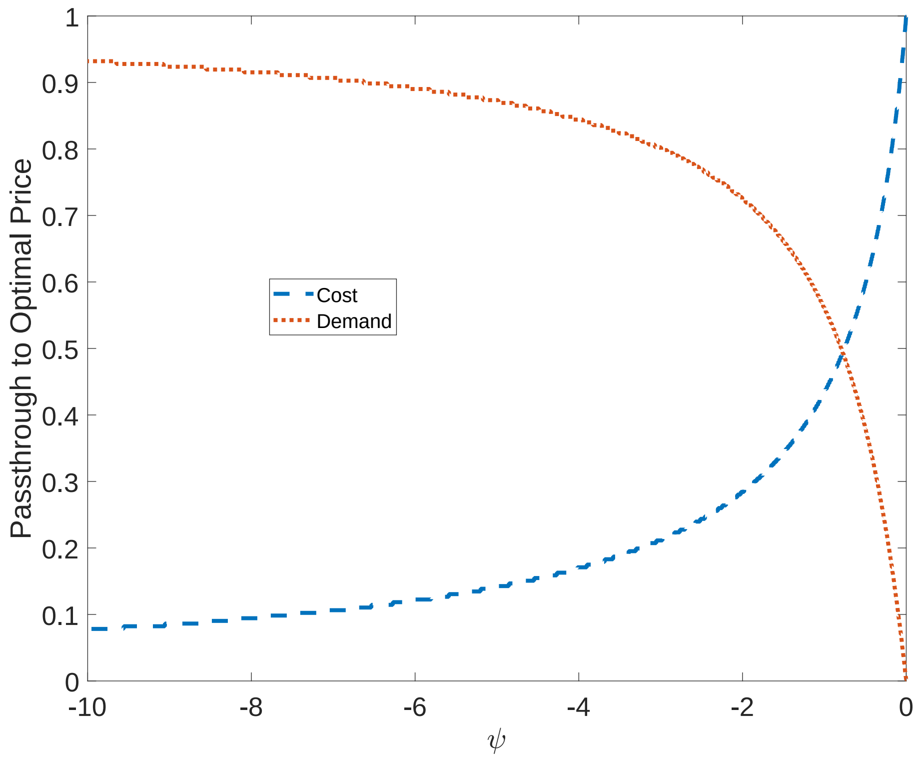
[5] See appendix A.1 for detailed derivations.

it optimal to reduce its price, which yields greater sales. However, as it moves along the demand curve, with a larger effective market share, its price elasticity decreases, which dampens the increase in revenue from cutting prices. As such, the optimal price cut is smaller than in the CES case. Furthermore, this line of reasoning implies that in the nonlinear solution, the size of the cost pass-through is smaller (larger) for a larger reduction (increase) in cost. Because variable price elasticities attenuate gains from price deviations, the Kimball demand system (with $\psi < 0$) generates strategic complementarities in that a firm wants to avoid moving its price too far away from its competitors. This is in line with the early literature such as Ball and Romer (1990) and Caplin and Leahy (1997), which emphasize the importance of strategic complementarities as a real rigidity.

Similarly, and obviously, when $\psi = 0$, demand pass-through is zero, which is the standard result under CES. When $\psi < 0$, demand pass-through becomes positive: a firm receiving a demand shock chooses to increase its price or, equivalently, chooses a higher markup over its marginal cost. Under a Kimball demand system, the elasticity of demand decreases in the effective market share. Firms with stronger demand for their product can raise prices without losing as much sales, resulting in higher markups.

The relative importance of demand and cost shocks hinges on the degree of strategic complementarity. Figure 3.1 plots the pass-through of cost and demand to the optimal frictionless price as a function of ψ, holding ω constant at our calibrated value. As ψ decreases, the pass-through of cost (productivity) shocks decreases and the pass-through of demand shocks increases. This figure makes it clear that the value of ψ will be critical for determining the pass-through of idiosyncratic shocks. There is overwhelming evidence in the international-trade, international-finance and firm-dynamics literatures that the pass-through from cost shocks to prices is less than

Figure 3.1: Pass-through of Demand and Cost Shocks to Price



Note: This plots pass-through of a small (1%) change in demand and productivity to the optimal frictionless price around a symmetric equilibrium with $\omega = 1.29$

complete, some of which we turn to in Section 3.3.1.3. This suggests that a constant elasticity of substitution specification ($\psi = 0$) cannot produce a cost pass-through that matches empirical estimates and strongly points toward a role for strategic complementarities ($\psi < 0$).

3.2.4 Equilibrium

Money supply, S_t, which must be equal to nominal aggregate expenditures, P_tC_t, in equilibrium, follows the stochastic process

$$\log(S_t) = \mu + \log(S_{t-1}) + \sigma_S \epsilon_t \text{ where } \epsilon_t \sim N(0,1), \tag{3.22}$$

where money supply grows at rate of μ every period with stationary fluctuations around it given by ϵ_t. As standard in the literature, because a one-time change in ϵ_t creates a permanent change in money balances, we interpret it as a monetary policy shock. This shock is the only source of aggregate uncertainty in our model. In calibrating our model, we set $\sigma_S = 0$ as it has minimal influence on the model-implied moments used for calibration. We explain our computational strategy in more detail in Appendix A.2.

3.3 Calibration

Most quantitative papers in the literature that uses menu cost models use a single idiosyncratic shock in their setup. Sometimes this shock directly moves the desired price around (e.g. Caplin and Spulber (1987)), where the authors are agnostic about the fundamental source of this shock. In other instances it is a productivity shock (e.g. Vavra (2014)), but the authors do not use firm-level evidence on productivity to calibrate it. In both cases, the process that drives either the desired price or productivity is typically calibrated to match various moments related to the distribution of firm-level price changes.

In this study, we aim to have a model that respects a broader set of micro-level evidence while also delivering significant monetary non-neutralities. To do so, we introduce two firm-level shocks in our model – productivity and demand. Furthermore, we calibrate the processes for these shocks to be consistent with direct firm-level evidence. This is a significant deviation from the common practice explained above. In order to demonstrate how different pieces of the calibration works and to emphasize that including demand shocks *and* deviating from CES are key to the success of the calibration, we consider three calibrated models with CES demand in addition to our baseline model with a Kimball demand system. Before turning to the details of these four models, we first review the moments we use from the data, either as calibration moments or as untargeted moments.

3.3.1 Calibration Targets and Non-Targeted Moments

In this section we report all of the data moments used throughout the paper. It is important to note that a given moment may be a calibration target for some of the models described above while being an untargeted moment in other model specifications.

3.3.1.1 Firm-Level Productivity and Demand Processes

Using the Census of Manufactures, the seminal work of Foster et al. (2008) estimate firm-level productivity and demand for eleven product markets with minimal vertical differentiation.[6] Using data on sales, quantity sold, and input usage, they estimate the production function of firms assuming Cobb-Douglas technology and recover firm-level physical TFP (TFPQ) as the residual of the following estimation:

$$TFPQ_{it} = \ln q_{it} - \alpha_l \ln l_{it} - \alpha_k \ln k_{it} - \alpha_m \ln m_{it} - \alpha_e \ln e_{it}, \tag{3.23}$$

where $TFPQ_{it}$ is the firm-level physical TFP of firm i at time t, q_{it} is the actual quantity produced by the firm, l_{it} is the labor input, k_{it} is the capital input, m_{it} are the intermediate inputs used in production, and e_{it} is the energy used by the firm. Foster et al. (2008) also estimate revenue-based TFP, which can be obtained using the same method but replacing the dependent variable with the

[6] Examples include bread, block ice, and ready-mix concrete.

revenue of the firm,

$$TFPR_{it} = \ln p_{it}q_{it} - \alpha_l \ln l_{it} - \alpha_k \ln k_{it} - \alpha_m \ln m_{it} - \alpha_e \ln e_{it}. \tag{3.24}$$

To obtain firm-level estimates of idiosyncratic demand, Foster et al. (2008) estimate the demand function

$$\ln q_{it} = \alpha_0 + \alpha_1 \widehat{\ln p_{it}} + \sum_t \alpha_t \text{YEAR}_t + \alpha_2 \ln(\text{INCOME})_{mt} + n_t^i, \tag{3.25}$$

using an instrumental variable regression, where the log-price, $\ln p_{it}$, is instrumented by the estimate of TFPQ from (3.23), which serves as a demand shifter. The regression includes time fixed effects and the average income in a plant's local market, m, is defined using the Bureau of Economic Analysis' Economic Areas. The residual of this equation is interpreted as a pure demand shifter for that firm.

For the eleven products in the analysis, Foster et al. (2008) report average five-yearly autocorrelations of 0.32 and 0.62 for idiosyncratic TFPQ and demand, respectively. The cross-sectional dispersion of TFPQ and demand are 0.26 and 1.16 respectively. This means that demand shocks are more persistent and more dispersed across firms. Furthermore, they report a correlation of –0.54 between firm-level prices and TFPQ and a correlation of 0.75 between firm-level TFPQ and TFPR.

3.3.1.2 Pricing Moments

For moments related to micro-level pricing behavior, we reference Vavra (2014) who reports pricing moments using CPI micro-data from the Bureau of Labor Statistics spanning the period from 1988 through 2012.[7] Price data are at the product-outlet level and temporary sales are discarded from the analysis. In his sample, Vavra (2014) reports a monthly frequency of a regular price change to be 11%, of which 65% are upward adjustments. The average size of a price change excluding non-adjustments is 7.7%, and the standard deviation of price changes is 0.075.

3.3.1.3 Markup and Pass-Through of Cost Shocks to Prices

Following the methods of De Loecker et al. (2020), we estimate the markup distribution of U.S. public firms using Standard and Poor's Compustat data. To be in line with the sample in Foster et al. (2008), we restrict the analysis to data between 1980 and 2000. In particular, we follow the production approach and compute firm-level markup as the ratio of sales to cost of goods sold, multiplied by the output elasticity of variable inputs estimated at the two-digit NAICS level.[8] In our sample, the average markup is 56% and the median markup is 33%.

A major theoretical implication of a Kimball demand system is the incompleteness of cost pass-through to prices. One of the ways of capturing empirically the magnitude of cost pass-through can be found in the international finance literature. This literature looks at the pass-through of exchange rate shocks to importer prices, with the understanding that the exchange rate movements are exogenous from the viewpoint of importers. The empirical evidence is overwhelmingly in support of an incomplete pass-through of costs even in the medium and long-run: Campa and Goldberg (2005) estimate the long-run pass-through in the US to be 42%, Gopinath and Itskhoki (2010) find it to be between 20% to 40%, and Gopinath et al. (2010)

[7] The same dataset is widely used in the literature, see Bils and Klenow (2004); Nakamura and Steinsson (2008).

[8] Following the literature, we exclude the following two-digit industries: utilities, finance and insurance, real estate and rental and leasing, as well as public administration.

Table 3.1: Externally Calibrated Parameters

Parameter	Description	Value	Source
β	Discount Factor	0.9966	Annual discount rate of 4%
χ	Labor disutility	1	Normalization
μ	Growth rate of S	0.002	Annual inflation rate of 2.4%
σ_S	SD of shocks to nom. expenditure	0.0037	Vavra (2014)
σ_{zn}	Corr. b/w productivity and demand innov.	0	Foster et al. (2008)

Note: This table displays the externally calibrated parameters in the model.

find an aggregate pass-through of 30%. On the other hand, estimation of cost pass-through is more challenging in a purely domestic setting, due to the scarcity of appropriate data and well-identified shocks. To this end, recent studies using merged data on both costs and prices recover cost pass-through estimates that are similar to the international macro evidence. Using Chilean supermarket-supplier merged data, Aruoba et al. (2022) find that 29% of a supplier price change is passed onto the retail price conditional on a price change at the supermarket level. Carlsson et al. (2022) estimate that between 21% to 33% of innovations to firm productivity are passed through to prices using data on Swedish manufacturing firms. Overall, the evidence from both the open- and closed-economy literature points to incomplete cost pass-through to prices in the range of 20% to 40%.

3.3.2 Externally Calibrated Parameters

In this section we explain how we fix a subset of parameters. These parameters are kept the same across the four different versions of the model we consider. A period is a month, and we set the monthly discount rate β to 0.9966 such that the annual discount rate is 4%. Consistent with the usual choice in the literature (Golosov and Lucas, 2007), the disutility of labor χ is normalized to 1, so that the nominal wage, W_t, is equal to the money supply, S_t. The monthly growth rate of the money supply, μ, is 0.2%, which implies an annual inflation rate of approximately 2.4%. Following Foster et al. (2008), we assume that idiosyncratic demand and productivity innovations are uncorrelated.[9] Finally, when we run monetary policy experiments, we set $\sigma_S = 0.0037$ following Vavra (2014) to match the volatility of nominal output growth in the U.S. Table 3.1 summarizes the five externally calibrated parameters.

3.3.3 Internally Calibrated Parameters

The former analysis leaves 7 remaining parameters to be internally calibrated. Four of these parameters govern the AR(1) processes for idiosyncratic productivity (ρ_z, σ_z) and idiosyncratic demand (ρ_n, σ_n), two additional parameters govern the Kimball demand system (ω, ψ), and the final parameter is the fixed menu cost (f). In our baseline model, we jointly calibrate these parameters to match a set of empirical moments. Here are more details of how the calibration works for each of the four versions we consider:

- **CES I**: Here we use the standard, or agnostic, approach in the literature and calibrate (f, ρ_z, σ_z) to match three pricing moments: frequency of price changes, fraction of positive changes and the average size of price changes. In this version there are no demand shocks. We set $\omega = 1.33$ to obtain a desired markup of 33%, which is the median markup in our data, and, naturally $\psi = 0$ so that CES aggregation is obtained.

- **CES II**: In this version we take the first step in trying to be consistent with the firm-dynamics facts from Foster et al. (2008) and calibrate (ρ_z, σ_z) to match two moments:

[9] The estimation strategy of Foster et al. (2008) requires demand and productivity to be orthogonal. Using Colombian data and an alternative strategy that relaxes the assumptions about the covariance between demand and supply shocks, Eslava et al. (2022) report a minimal correlation between demand and productivity.

the five-yearly autocorrelation and the variance of TFP . We still calibrate f to match the frequency of price changes, turn off demand, and set $\psi = 0$ and $\omega = 1.33$.

- **CES III**: This version adds the firm-level demand shocks to the model and (ρ_n, σ_n) to match two moments: the five-yearly autocorrelation and the variance of demand from Foster et al. (2008). The rest of the strategy is identical to CES II.

- **Baseline**: The baseline model jointly calibrates all seven parameters to match seven moments: the four firm dynamics moments CES III uses, two additional moments from Foster et al. (2008): Corr(TFPQ,P) and Corr(TFPQ,TFPR), as well as the frequency of price changes as the only pricing moment.

The model-based moments we need for calibration can only be computed via simulation. To that end, we simulate 20,000 firms for 700 periods and drop the first 100 periods before computing any statistics. Computing moments that are monthly is straightforward. In order to compute moments that have their data counterpart in Foster et al. (2008), we aggregate the simulated data to the corresponding frequency and replicate their methodology. In particular, we aggregate the simulated monthly data into annual frequency by taking simple sums of revenue, sales, and employment. We then construct a panel dataset with the same time structure as Foster et al. (2008), namely five waves of annual observations that are five years apart. Because labor is the only input in the model, we recover firm-level TFPQ and TFPR as,

$$TFPQ_{it} = \ln q_{it} - \ln l_{it}, \tag{3.26}$$

$$TFPR_{it} = \ln (p_{it} q_{it}) - \ln l_{it}. \tag{3.27}$$

This is equivalent to mapping our unique inputs to their basket of inputs. We estimate the demand function using the same IV specification as Foster et al. (2008),

$$\ln(q_{it}) = \beta \ln(p_{it}) + \text{Time FE} + \eta_{it}, \tag{3.28}$$

where $\ln(p_{it})$ is instrumented by $TFPQ_{it}$ and recover firm-level demand shifters as the residuals, η_{it}. At the end of this process, we obtain five-yearly measures that are direct counterparts of those computed by Foster et al. (2008).

Before turning to the results, a discussion on the identification of parameters is in order. While all parameters influence the model's ability to match all calibration targets, some parameters are more responsible for matching specific target moments. Some of these are quite intuitive. The fixed cost, f, has a significant role in the model-implied frequency of price changes, and the shock-process parameters $(\rho_z, \sigma_z, \rho_n, \sigma_n)$ are mostly related to the corresponding five-yearly moments from Foster et al. (2008).

What may be less obvious is how ψ and ω are linked to the correlations of TFPQ with prices and TFPQ with TFPR. To understand this, it is instructive to start from a CES demand system (with $\psi = 0$). The profit-maximizing rule in that framework delivers a pricing strategy that sets price as a constant markup over marginal cost. As a result, a firm's optimal price is inversely proportional to its productivity, that is Corr$(P, TPFQ) = -1$. Moreover in a CES demand system TFPR is equalized across firms and as such Corr$(TFPR, TPFQ) = 0$. This is because optimizing firms will operate at the point where the marginal product of labor $(p_{i,t}z_{i,t})$ is equal to the nominal market wage. Under a Kimball demand system, both productivity and demand factors affect the optimal price of a firm. In particular, deviations from CES, controlled

Table 3.2: Internal Calibration

Moment	Data	CES I	CES II	CES III	Baseline
Frequency of price changes	0.11	**0.11**	**0.11**	**0.11**	**0.12**
Fraction of price increases	0.65	**0.64**	0.61	0.58	0.58
Size of price changes	0.08	**0.08**	0.15	0.14	0.07
5-yearly autocorr of z_t^i	0.32	0.00	**0.32**	**0.32**	**0.32**
Cross-sectional SD of z_t^i	0.26	0.03	**0.26**	**0.26**	**0.25**
5-yearly autocorr of n_t^i	0.62	0.01	0.00	**0.62**	**0.62**
Cross-sectional SD of n_t^i	1.16	0.01	0.04	**1.18**	**1.05**
Corr b/w TFPR and TFPQ	0.75	0.00	0.00	0.00	**0.74**
Corr b/w price and TFPQ	–0.54	–1.00	–1.00	–1.00	**–0.57**
Parameter	**Description**				
ψ	Super-elasticity	0	0	0	**–1.27**
ω	Elasticity of Substitution	1.33	1.33	1.33	**1.29**
ρ_z	Persistence of z_t^i	**0.66**	**0.98**	**0.98**	**0.98**
σ_z	Standard deviation of z_t^i	**0.04**	**0.05**	**0.05**	**0.06**
ρ_n	Persistence of n_t^i	–	–	**0.992**	**0.997**
σ_n	Standard deviation of n_t^i	–	–	**0.05**	**0.02**
f	Menu cost	**0.01**	**0.06**	**0.03**	**0.03**

Note: The top panel of this table compares the targeted moments and model-implied moments for the four model specifications, where the bolded numbers highlight moments that are targeted in the calibration. The bottom panel shows the parameter values for each calibration.

by the parameter ψ, diminish the role of productivity in pricing relatively to demand. This, in turn, would reduce the perfect negative correlation between price and TFPQ since some price changes will be due to demand shocks. The parameter ω, on the other hand, governs the elasticity of substitution between varieties, with a higher value of ω indicating less substitutability across varieties. In a Kimball demand system, more productive firms can charge a higher price if the elasticity of substitution is lower (ω is higher) leading to a higher correlation coefficient between TFPQ and TFPR.

The results of the internal calibration of the four versions of the model (three CES versions and the baseline Kimball version) are reported in Table 3.2. The top two panels show the moments considered, where targeted moments for each version is shown in bold. The first panel shows pricing moments described in Section 3.3.1.2 while the second panel shows the firm-dynamics moments from Foster et al. (2008), described in Section 3.3.1.1. The third panel shows the seven parameters where boldface indicate that the parameters calibrated jointly, while others are fixed as explained earlier.

Starting with CES I, which uses (ρ_z, σ_z, f) to match the first three pricing moments in the first panel, we see that it matches those moments very well. However, it completely misses the firm-dynamics moments. The size and persistence of the z process, which is used to match pricing moments, generates very small and nearly transitory movements in idiosyncratic productivity, as can be seen in the first two rows of the firm-dynamics moments. This version is, by definition, unable to say anything about the remaining firm-dynamics moments due to the absence of demand shocks.

Turning to CES II, where we now give up on matching pricing moments, except for the frequency of price changes, and use (ρ_z, σ_z) to match the first two firm-dynamics moments, the calibration is successful in the sense that the targeted moments are matched, including the dynamics of idiosyncratic productivity. However, this version misses the two untargeted pricing moments, especially the last one completely. In an effort to match the more volatile and persistent idiosyncratic productivity process, the model generates price changes that are about twice as large on average than the data. Moreover, due to the absence of demand shocks, the four remaining firm-dynamics moments are also not matched.

CES III attempts to match additional firm dynamics moments by including a idiosyncratic demand shock in addition to a idiosyncratic productivity shock, once again leaving all pricing moments except for the frequency of price changes as untargeted. This version is able to match the four firm-dynamics moments by picking appropriate parameters for the aforementioned shocks.

Due to the CES structure, as explained above, the last two moments are still elusive for this version. And just like CES II, it fails to deliver on the untargeted pricing moments.

Finally turning to the baseline model featuring a Kimball demand system, we use the only pricing moment common to all models (frequency of price changes) in addition to all of the six firm-dynamics moments as targets and calibrate all seven parameters. The second panel shows that the model matches all of the firm-dynamics moments very closely. The calibrated process for idiosyncratic demand is highly persistent with a monthly autocorrelation of 0.997, whereas the idiosyncratic productivity process exhibits less persistence with a monthly autocorrelation of 0.98. The standard deviation of the innovation to idiosyncratic productivity (0.06) is higher than that of idiosyncratic demand (0.02). Preferences appear to be very persistent and subject to relatively small shocks when compared to technology. However, the stationary distribution of idiosyncratic demand is highly dispersed due to the persistent nature of the process. The calibrated values of ω and ψ are 1.29 and –1.27, respectively. These parameters are mostly informed by the last two firm dynamics moments. The value of ω is not too far from the value we fix in the CES versions based on a 33% markup, though with a Kimball demand system it is no longer interpreted as the desired markup.

To sum up, the baseline model is successfully calibrated in that all targeted moments are matched reasonably well. We next turn to the discussion of how the baseline model performs in matching untargeted moments, including the pricing moments shown in Table 3.2 in Section 3.4.1.

3.3.4 Identification of Model Parameters

We alluded to certain calibration targets as being the main source of information in the calibration of certain parameters. In this section, we make this point explicit. In doing so we also demonstrate that the calibration targets are indeed very informative for the respective parameters by borrowing an exercise from Daruich (2022).

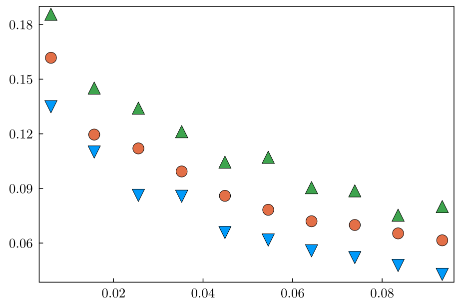
The main idea is to generate variation in the parameter space and investigate how the implied calibration targets are impacted – essentially taking a partial derivative. To do so, we first draw 500 parameter vectors from uniform Sobol points given a hypercube of the parameter space, which generates a quasi-random set of candidate parameter vectors.[10] Then, for each parameter vector, we solve and simulate the model to compute the relevant model-implied moments. This allows us to see how each of the seven parameters influences each of the seven calibration targets.

Figure 3.2 plots the values of each model-implied target moment against the values of the parameter it is assigned to. In particular, we group the values of each parameter in deciles, which we plot on the horizontal axis. Then, for each decile, we show the median value of the associated moment in red circled dots and the 25^{th} and 75^{th} percentiles in blue down-pointing triangles and green up-pointing triangles, respectively. The slope of the scatter plot is informative about the importance of that parameter, whereas the vertical dispersion reveals the influences of all other parameters on a particular moment.

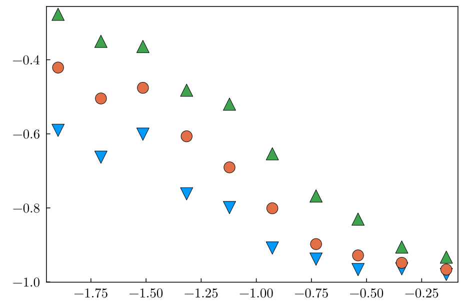
The frequency of price adjustment exhibits a strong negative correlation with the menu cost f. Meanwhile, other parameters also play a role as is evident in the vertical dispersion.

[10] A uniform Sobol sequence is a sequence of points that spans the n-dimensional hypercube in an even and quasi-random manner. For the purpose of the exercise, using quasi-random Sobol numbers are more efficient than drawing random numbers because Sobol numbers are designed to sample the space of possibilities evenly given the total number of draws, whereas a truly random sample is subject to sampling noise.

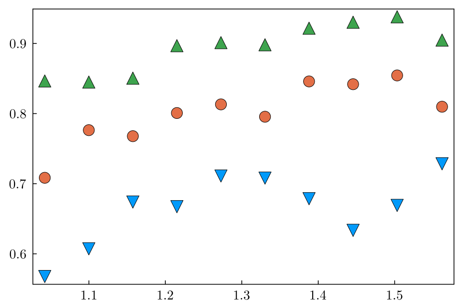
Figure 3.2: Identification of Internally-Calibrated Parameters



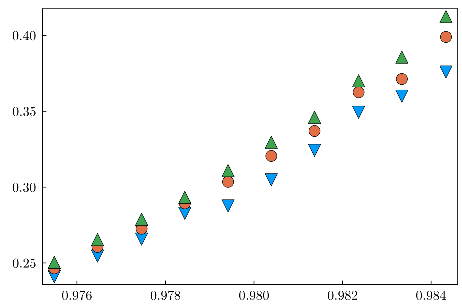
(a) Price Adjustment Frequency vs. f



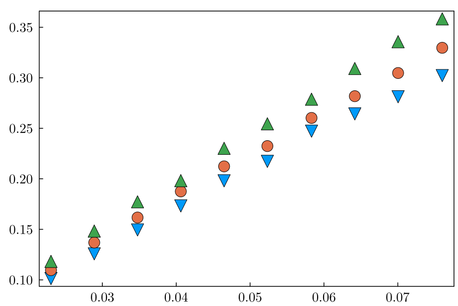
(b) Corr$(TFPQ, P)$ vs. ψ



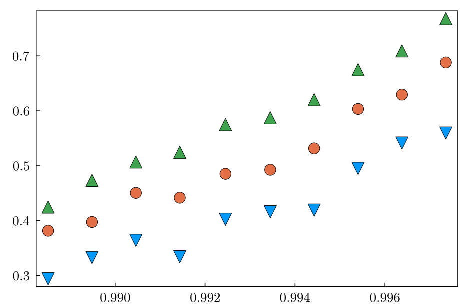
(c) Corr$(TFPQ, TFPR)$ vs. ω



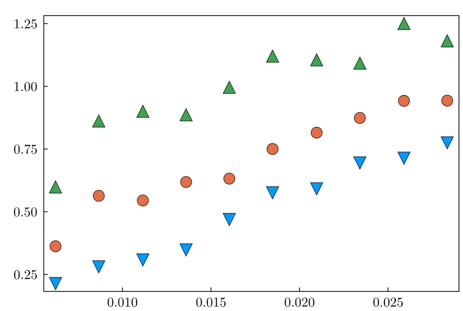
(d) Five-yearly AR of TFPQ vs. ρ_z



(e) Cross-sectional SD of TFPQ vs. σ_z



(f) Five-yearly AR of demand vs. ρ_n



(g) Cross-sectional SD of demand vs. σ_n

Note: For each decile of a given parameter plotted on the horizontal axis, the red dot shows the median of the moment that is assigned to the parameter. The blue down-pointing triangles and green up-pointing triangles show the 25^{th} and 75^{th} percentiles respectively.

For example, for a fixed value of f, larger idiosyncratic shocks generate more frequent price changes. The parameters (ρ_z, σ_z) are strongly correlated with the five-yearly autocorrelation and cross-sectional distribution of firm productivity, whereas other parameters play a minimal role as can be seen in the tight vertical variation in the scatter plots. For (ρ_n, σ_n), we observe a similar relationship, but there is noticeably more noise in the cross-sectional standard deviation of demand. This is mainly because at a given decile of σ_n, the remaining parameters, including ρ_n are randomly drawn. Since the value of ρ_n is generally very close to one, the resulting cross-sectional dispersion of demand is very sensitive to the value of ρ_n in addition to σ_n. Lastly, consistent with our reasoning, we recover a strong negative relationship between Corr(TFPQ,P) and ψ. We also observe a weak but visibly positive relationship between Corr(TFPQ,TFPR) and ω. The large variation in this correlation given a value of ω reveals that it is sensitive to the values of other parameters in addition to ω. In particular, we find that σ_n and σ_z, which determine the stationary distribution of idiosyncratic productivity and demand, have sizable effects on the level of this correlation. Given that all the parameters except for ω exhibit tight links with their associated targets, we argue that ω can be credibly identified by Corr(TFPQ,TFPR) when all other parameters are fixed and matched to their respective targets.

They key takeaway from this exercise is that for the most part, the links between the parameters and moments are quite tight. While it is too computationally intensive, if one were to consider a formal generalized method of moments approach to estimating the parameters of interest, this figure seems to suggest that one would obtain fairly tight standard errors for the estimates.

3.3.5 External Validity

The work of Foster et al. (2008) is the only study for the U.S. with a systematic estimation of productivity and demand shocks at the firm level for different industries. This estimation requires price and quantity data at the product level along with other information such as inputs. By using a carefully selected set of firms in the Census of Manufactures that produce uniform products, they are able to separately estimate shock processes for productivity and demand. However, the external validity of the estimation of Foster et al. (2008) to the broader economy may be a concern.

In this regard, we first note that we calibrate the strength of strategy complementarity in the baseline model using the correlations between TFPQ, prices, and TFPR from Foster et al. (2008). The calibration yields a cost pass-through of 38%, which is in line with evidence on the exchange-rate pass-through from the international finance literature as well as recent estimates based on domestic cost-price linked data, both of which report estimates in the range of 20 to 40 percent.

Furthermore, several other countries have similar data for a wider set of firms, and researchers have estimated some of the moments that we use for our identification strategy. For instance, Eslava et al. (2013) use Colombian firm-level data covering the entire manufacturing industry to separately identify productivity and demand processes at the firm level. They report similar values for Corr($TFPQ, TFPR$) and Corr($TFPQ, P$), which are crucial for pinning down the Kimball demand system parameters. Specifically, they report Corr($TFPQ, TFPR$) = 0.69 and Corr($TFPQ, P$) = –0.65 versus the Foster et al. (2008) values of 0.75 and –0.54, respectively.[11]

[11]Moreover, they estimate a correlation between productivity and demand innovations in the neighborhood of zero, consistent with Foster et al. (2008) and our assumption.

Given that idiosyncratic demand and productivity processes determine the ergodic properties of firm growth, we can also compare the cross-sectional dispersion of output growth rates computed from model-simulated data with external evidence to gauge if the estimates from Foster et al. (2008) can be generalized beyond the eleven industries. We benchmark our estimates to Davis et al. (2006), a study using the Longitudinal Business Database and thus a good measure of the U.S. business dynamics. They estimate the cross-sectional standard deviation of firm revenue growth rate to be 0.39 over the period 1982-1997, while in our simulated data this untargeted moment is 0.41. Therefore, we conclude that the estimation of Foster et al. (2008) can be used to shed some light about the general behavior firms.

3.4 Results

We calibrated our model using firm-dynamics moments from Foster et al. (2008) and one pricing moment, the average frequency of price adjustments. In this section, we turn to investigating the implications of the model in three dimensions: other pricing moments, markup distribution, and monetary non-neutrality, none of which was targeted in our calibration.

3.4.1 Untargeted Pricing Moments

Table 3.3 shows three important pricing moments we do not target in the calibration of the baseline model, along with their data counterparts and the results from two CES versions where these pricing moments are not targeted. These are the average size of a price change conditional on a change, the fraction of adjustments that are positive, and the dispersion of non-zero price changes. In CES II and CES III we calibrate the properties of the stochastic processes for demand

Table 3.3: Untargeted Pricing Moments

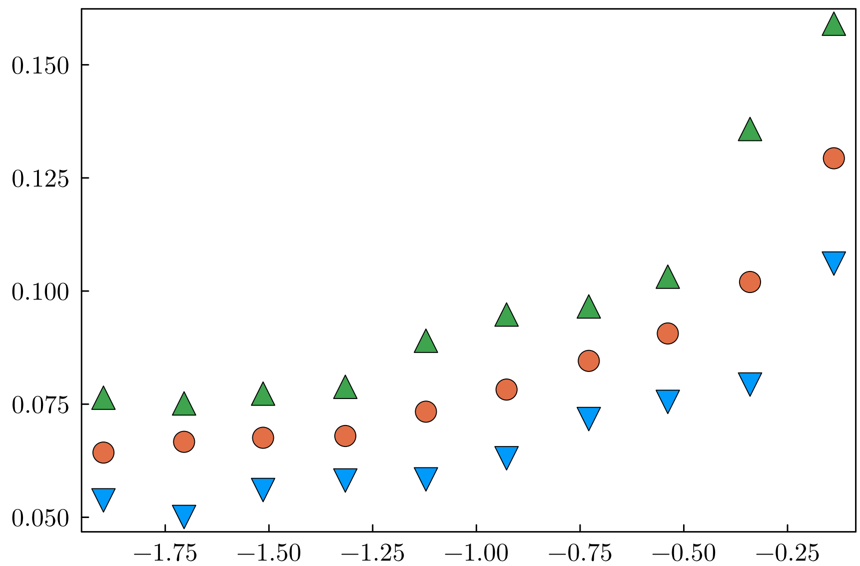
Moments	Data	CES II	CES III	Baseline
Average Size	0.08	0.15	0.14	0.07
Fraction Up	0.65	0.61	0.58	0.58
SD(Δp)	0.08	0.16	0.17	0.07

Note: This table shows the three untargeted moments: average size of adjustment conditional on a price change, the fraction of adjustments that are positive, and the standard deviation of price changes excluding zeros from the empirical data and from model simulated data.

and productivity to match the firm dynamics moments. The productivity process, in turn, ends up producing price changes that are about twice as large as what is in the data. The table shows that, unlike the CES versions, the baseline version is actually able to match the three untargeted pricing moments quite well – most importantly the size of price changes is tempered relative to the CES versions. The reason for this is, unlike CES where the pass-through of cost shocks are 100%, with a Kimball demand system the pass-through is much less and therefore for the same level of productivity shocks, price changes are smaller. A back-of-the-envelope calculation using the 38% cost pass-through reported above would suggest that for the same size productivity shocks, the size of price changes will be about 0.053. Because firms also change prices in response to demand fluctuations, the introduction of demand shocks increases this number to around 0.07, which is very close to the moment in the data at 0.08.

In comparison, Klenow and Willis (2016) use a model with a Kimball demand system, but they do not include an idiosyncratic demand shock. They follow the approach in CES I and calibrate their model to match pricing moments. Their conclusion, similar to our findings for CES I, is that the properties of the firm-level productivity process needed for matching the pricing moments is inconsistent with those found in the firm-dynamics literature. The baseline calibration in Klenow and Willis (2016) ($\theta = 5, \epsilon = 10$ using their specification) translates

Figure 3.3: Value of ψ and Average Size of Non-zero Price Changes



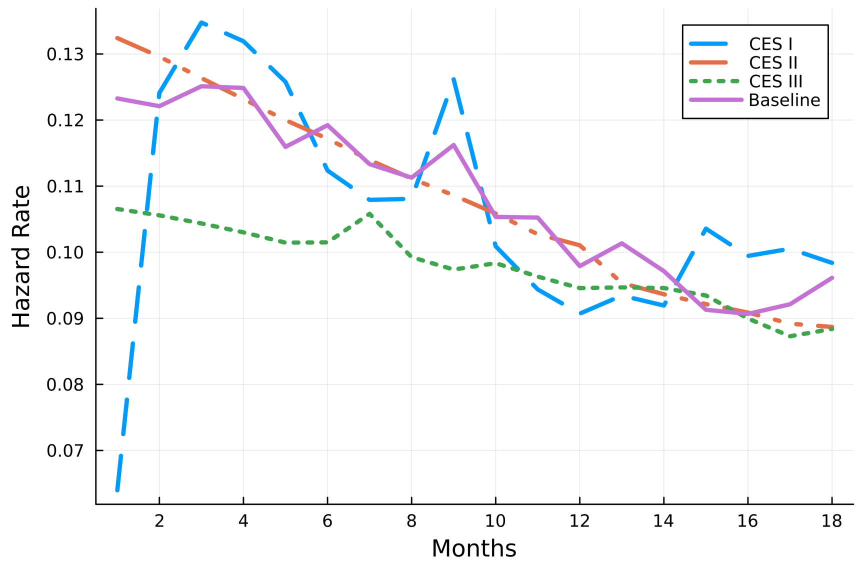
Note: We group values of ψ into deciles. For each decile of ψ values plotted on the horizontal axis, the red dots, blue down-pointing triangles, and green up-pointing triangles show the median, 25^{th}, and 75^{th} percentiles of a given untargeted pricing moment respectively. The underlying data is the same as that used for Figure 3.2. Specifically, they are random draws from a hypercube of parameter space.

roughly to parameter values $\omega = 1.25$ and $\psi = -2$ under our specification of the Kimball aggregator and a cost pass-through of 28%. This is not too far from the pass-through that we obtain in our calibration. Since the cost pass-through is a third of CES, in order to match the same pricing moments, one would need even larger shocks than what we obtained in CES I. This is why Klenow and Willis (2016) conclude that their calibrated model was not consistent with the firm dynamics literature. Our model shows that the inclusion of idiosyncratic demand shocks compensates for the weaker role of productivity shocks. It turns out that idiosyncratic demand and productivity processes that are consistent with firm dynamics estimates can generate realistic price-adjustment facts at a degree of strategic complementarity that is also consistent with firm-level evidence.

One may wonder if the success of the model has much to do with the precise calibration of ψ or whether any deviation from CES by reducing ψ below zero would have done the trick. In Figure 3.3 we plot the average size of non-zero price changes versus different values of ψ, similar to Figure 3.2, where the vertical variation for a given level of ψ is due to the differences in other parameters. This figure shows that there is a tight relationship between the average size of adjustments and ψ. As ψ falls, the increasing real rigidities (strategic complementarities) make the firms less and less willing to deviate from their competitors, avoiding large price changes. This makes the average size of price changes to fall. Thus, it is even more remarkable that our model matches these moments as good as it does, given that a different value of ψ from the one calibrated to firm-dynamics moments may have led to a worse fit.

We also examine the hazard function of price change from the model. The hazard of a price change is the probability that a price will change t periods after the last adjustment, conditional on the price spell lasting t periods. Empirically, the hazard function is found to be either downward-sloping (Nakamura and Steinsson (2008)) or flat (Klenow and Kryvtsov (2008)). As pointed out by Nakamura and Steinsson (2008), plain vanilla menu cost models are typically not able to generate hazard functions that are consistent with the empirical evidence. This largely hinges on the calibration of the idiosyncratic processes. In a model with trend inflation, the hazard function is upward-sloping when idiosyncratic shocks are small. Larger idiosyncratic shocks and more persistent idiosyncratic processes flatten the hazard function as they lead to temporary price changes that are often reversed quickly. When idiosyncratic shocks are sufficiently large, a plain vanilla menu cost is able to generate a downward-sloping hazard. Nakamura and Steinsson (2008) argue that such calibrations are unrealistic, due to the fact that they are inconsistent with micro pricing facts. This conclusion has spurred alternative pricing models that seek to rationalize a downward-sloping hazard, such as Baley and Blanco (2019) who introduce firm-level uncertainty

Figure 3.4: Hazard Function of Price Change

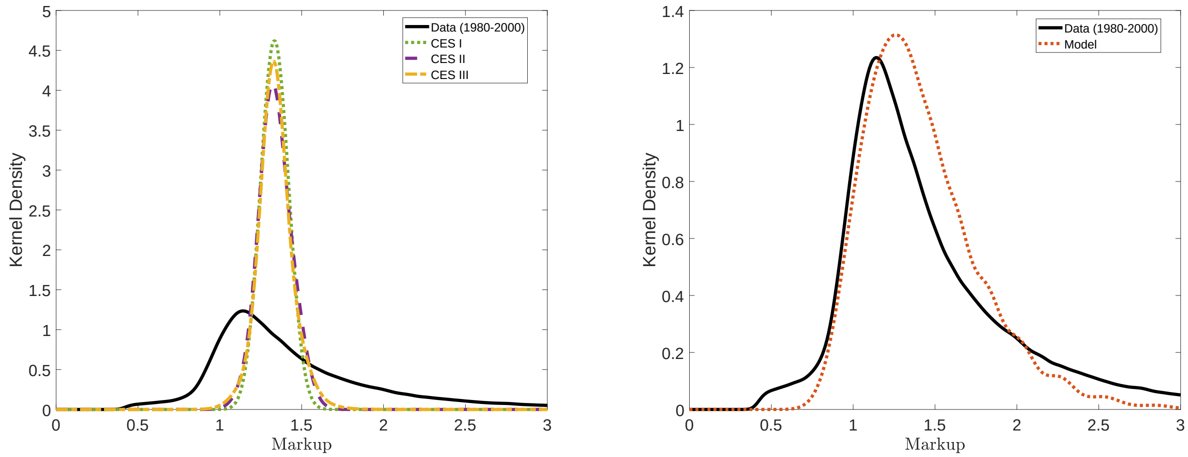


Note: This figure plots the pricing hazard for the four model calibrations over the first 18 months.

and learning to generate frequent price changes shortly following an adjustment.

Figure 3.4 plots the hazard function generated by our model across the four specifications. In CES I, which is the approach commonly taken by the literature, the hazard function is increasing over the first few months and flattens out afterwards. This is consistent with the baseline calibration of Nakamura and Steinsson (2008). The other three model specifications, where idiosyncratic processes are calibrated to firm dynamics evidence, instead exhibit a downward-sloping hazard. This shows that calibrations that imply downward-sloping hazards are not necessarily unrealistic. In fact, a simple menu cost model with Kimball aggregation, calibrated to firm dynamics estimates generates a price change hazard that is much more consistent with the empirical counterpart, while remaining consistent with micro pricing facts.

Figure 3.5: Gross Markup Distribution: Model vs. Data



(a) Cross-sectional Distribution of Markup

(b) Cross-sectional Distribution of Markup

These figures plot the kernel density of the empirical markup distribution from publicly traded firms in the U.S. as well as the kernel density of the markup distribution in the ergodic distribution of the four versions of the model. Both kernel densities are computed using the optimal bandwidth for normal densities.

3.4.2 Cost Pass-through and Markups

In a flexible-price model with CES and symmetric firms, all prices are set to a fixed markup above marginal cost and thus all firms have the same markup, leading to a degenerate markup distribution. With pricing rigidities, firms that cannot change their price in a period could potentially deviate from the desired markup. As a firm's marginal cost changes via changes in productivity, so would its price, reflecting the cost change one to one. Since demand shocks do not change prices, they have no impact on markups. Panel (a) of Figure 3.5 plots the kernel density of markup distribution for our CES versions along with its data counterpart. For all CES versions the markup varies very little and does so symmetrically around $\omega = 1.33$, or a net markup of 33%. This is completely at odds with the distribution we obtain from the data, which has a mode just above 0% with a very wide right tail reaching a level of 200%, though markups of as low as –50% are also observed.

The desired markup of a firm facing a Kimball demand system depends on both its idiosyncratic productivity and demand. Because the pass-through of marginal cost to price is incomplete, more productive firms do not pass on their cost advantage to price one-for-one, resulting in higher markups. Also, firms with larger idiosyncratic demand optimally choose higher prices and hence higher markups. The cross-sectional distribution of productivity and demand, alongside pricing frictions, result in a non-degenerate markup distribution in the model. Panel (b) of Figure (3.5) plots the kernel density of the cross-sectional distribution of gross markup both from the data and the model. The markup distribution from the model mimics the non-targeted empirical distribution reasonably well. The median gross markup in the model is 1.35, compared to 1.33 in the data. The model-implied markup distribution, however, exhibits lower variance relative to the empirical distribution, much of which stems from the tails. Our model does not generate as many firms that have markups larger than 200%, nor does it deliver markups deep into the negative territory. Matching these extreme tails of the distribution would require additional features such as non-Gaussian shocks to demand or monopolies (large positive markups) and customer capital (negative markups).

The success of our model in replicating the empirical markup distribution is perhaps not very surprising given the work of Arkolakis et al. (2019) and Edmond et al. (2018), among others, who use Kimball demand systems for modeling firm markups. What is remarkable is that our model, calibrated to results from Foster et al. (2008) that relies on selected manufacturing industries, provide a good fit to the distribution of markups estimated for a broader set of industries.

3.4.3 Monetary Policy and Non-Neutrality

Nominal rigidities alone are typically insufficient at generating sizable real effects to nominal shocks (Caplin and Spulber, 1987; Golosov and Lucas, 2007). A solution put forth by Ball and Romer (1990) is to introduce real rigidities in conjunction with nominal pricing frictions. Since then, the macroeconomics literature has explored the plausibility of various potential sources of real rigidities. Strategic complementarities in pricing induced by a Kimball demand system is one such mechanism. When a monetary policy shock hits the economy, some firms choose not to respond due to the presence of *nominal* rigidities, which makes monetary policy effective in stimulating real activity, or non-neutral. With strategic complementarities, the firms adjusting their prices choose to adjust less than under CES demand in order to remain closer to their competitors who choose not to adjust their prices, and this *real* rigidity adds to the degree of non-neutrality. This intuition is formalized in Alvarez et al. (2022), who derive analytic results in a menu cost model featuring strategic complementarity casted as a Mean Field Game and show that complementarity makes the impulse response of output to a nominal shock larger at each horizon.

In spite of its theoretical soundness, the literature has largely dismissed strategic pricing complementarity as an *empirically relevant* source of real rigidity due to the findings of Klenow and Willis (2016), who conclude that a menu-cost model featuring a Kimball demand system that is consistent with micro pricing facts and can generate sufficient monetary non-neutrality was inconsistent with the evidence from the firm dynamics literature. In the previous section, we showed that such a model can in fact be simultaneously consistent with both the micro pricing facts and the firm-dynamics literature. With this in mind, we turn to assessing whether our model is able to generate sizable real responses to nominal shocks.

We consider four measures of non-neutrality. The first measure is the unconditional standard deviation of consumption in a long simulation. Our model features no aggregate shocks other than the nominal expenditure shock ϵ_t in (3.22), which can also be interpreted as a monetary policy shock. As such, if monetary policy was perfectly neutral, then aggregate consumption would be constant. Thus, the standard deviation of aggregate consumption serves as a measure of deviation from neutrality. The other three measures are obtained from a response of the economy to a one-time change in the nominal expenditure shock, ϵ_t.[12] In particular, we compute the response of real output to a positive nominal expenditure shock of size 0.2%, which doubles the monthly growth rate of aggregate expenditure. The first object we compute from this response is the peak response of output as a fraction of the size of the shock, which in a model like ours without any internal propagation, will happen on impact. The second object we can compute from this response is the half-life of the shock, which tells us how persistent the effect of the shock is. Finally, these two measures can be summarized by the cumulative impulse response (CIR) which adds up the response of output as a fraction of the shock over the period it is non-zero and divides by the number of periods in a year – 12 in our case.

Before we turn to the results, it is worth mentioning that a CES model with Calvo pricing, where firms receive an i.i.d. shock that determines when they can adjust their prices, would yield an output response as a fraction of the shock that is exactly $(1-\alpha)^h$ where α is the probability that firms can change their price and h is the horizon where $h = 1$ is the period of the shock.[13] Based on the evidence we presented in Section 3.3.1.2, α would be set to 0.11 in such a model, indicating

[12] If the aggregate price does not respond at all, then the entirety of the increase in nominal expenditure is reflected in an increase in real output, leading to complete transmission to real output. On the other extreme, if the aggregate price is perfectly flexible, the real effect of the nominal shock would be zero.

[13] See Appendix A.1.2 for the derivations.

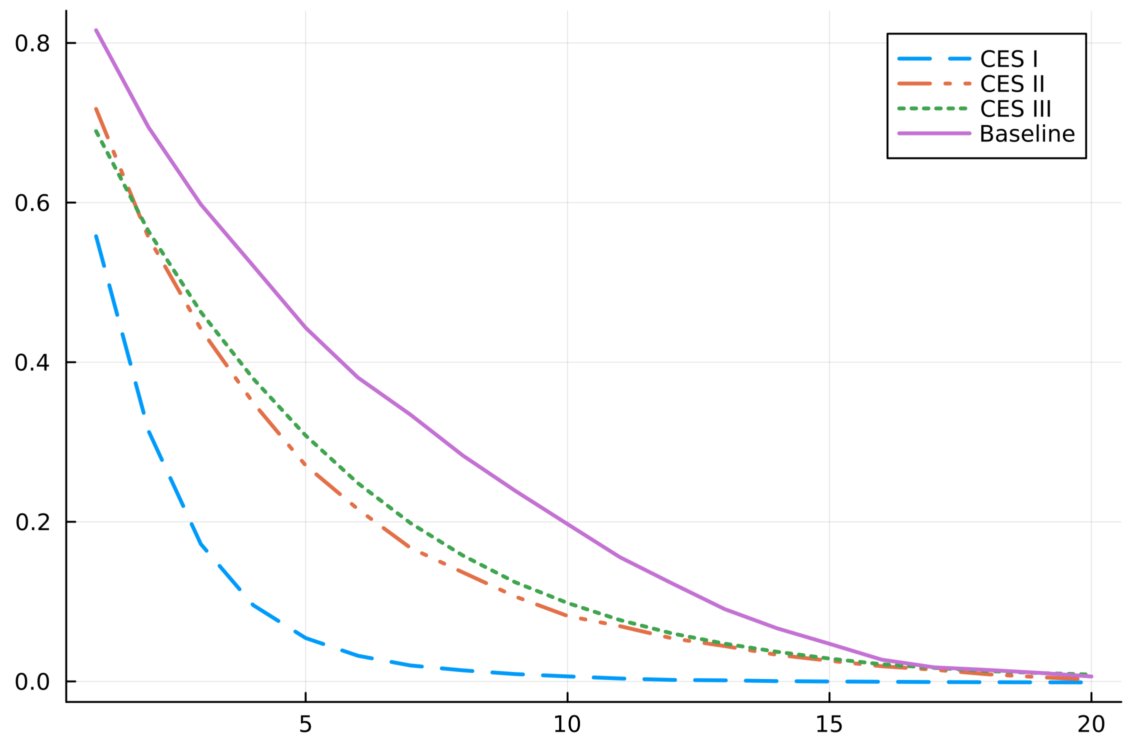


Figure 3.6: Impulse Response of Real Output to a Nominal Expenditure Shock

Note: This figure plots the impulse response of real output expressed as a fraction of the nominal expenditure shock on the vertical axis and periods elapsed since the shock on the horizontal axis.

that the initial (and peak) response of output would be 89% of the shock. The cumulative response according to the Calvo model would be 0.76.[14] This coincides with the result of Alvarez et al. (2016) who show that the CIR can be expressed in terms of the kurtosis and frequency of price changes. According to their formula, the Calvo model has a CIR of 0.76 whereas a menu cost model à la Golosov and Lucas (2007) has CIR of 0.13.

Figure 3.6 plots the impulse response of real output expressed as a fraction of the size of the shock for the baseline model as well as the three CES versions we introduced earlier. Table 3.4 contains the four statistics that summarize the degree of monetary non-neutrality we discussed above. CES I (blue dashed line), which was calibrated to pricing moments, deliver a

[14]The CIR in the Calvo model is given by $\sum_{t=0}^{\infty} \frac{0.89^t}{12} = 0.76$

Table 3.4: Measures of Monetary Non-neutrality

Moment	CES I	CES II	CES III	Baseline
SD(C)	0.22%	0.39%	0.44%	0.52%
Impact	0.56	0.69	0.69	0.82
Half-life	1.25	2.90	3.46	4.50
CIR	0.11	0.28	0.30	0.42

Note: This table displays four measures of monetary non-neutrality for the four model calibrations.

peak response of 0.56, which is substantially lower than the Calvo response. The response is also fairly short-lived with a half-life of 1.25 months. The resulting cumulative impulse response is 0.11, which is close to the number reported in Alvarez et al. (2016) for a Golosov-Lucas type menu cost model. CES II (orange dashed dotted line) which calibrates the model to the productivity process of Foster et al. (2008) produces more non-neutrality with a peak response of 0.69. It is also somewhat more long-lived with a half-life of 2.9 months and has a larger CIR of 0.28 compared to CES I. Neither of these versions delivers a reasonable level of non-neutrality, which is the key result of Golosov and Lucas (2007).

The results change once a demand shock is added to the model in the CES III (green dotted line) and the baseline (purple solid lines) versions. CES III shows an impact response of 0.69, a half-life of 3.5 months, and a CIR of 0.30. Monetary policy in this version exhibits slightly greater non-neutrality due to the fact that idiosyncratic demand shocks act as random menu costs.[15] In the presence of random menu costs, firms are responding not only to the aggregate shock but also to idiosyncratic realizations of the adjustment cost, thereby weakening the selection effect of responding to idiosyncratic productivity shocks and thereby raising monetary non-neutrality. However, the larger non-neutrality in CES III relative to other CES versions come at the expense

[15] Random menu costs were first introduced by Dotsey et al. (1999). Recent works by Nakamura and Steinsson (2010), Midrigan (2011), and Alvarez et al. (2016) explicitly explore the implications of random menu costs on monetary non-neutrality.

of matching the pricing moments – recall that the average size of price changes were about twice the level in the data. Moreover, CES III, as with other CES versions, is unable to deliver a realistic markup distribution.

Finally turning to our baseline version, on impact, approximately 82% of the increase in nominal expenditure is reflected in an increase in the real output, which is 40% larger than the impact response in CES I and 20% larger than the other two CES calibrations. The real effects of the shock dampen as time goes by, though much more slowly than any of the CES versions, and eventually die out about 20 months after the shock. The half-life of the shock is 4.5 months. The larger response on impact coupled with slower decay is reflected in the relatively large CIR of 0.42.

The degree of monetary non-neutrality that the model generates is larger than the CES versions. In particular, it generates a cumulative response that is four times as large as CES I which has the archetypal setup. The real effect is also close to the Calvo version which generates a cumulative response that is six times as large as the Golosov-Lucas model (Alvarez et al., 2016). Lastly, it is fair to say that a response of this magnitude is considered appropriately large in the literature as summarized in Mongey (2021).[16]

The model exhibits substantial monetary non-neutrality due to two key features. First, as discussed, a Kimball demand system introduces strategic complementarity in pricing among

[16]As Table A.1 of Mongey (2021) summarizes, menu-cost models without real rigidities typically generate a peak output response in the range of 0.35 to 0.50 (Golosov and Lucas, 2007). Some of the other studies exploring alternative sources of real rigidities find higher numbers. Nakamura and Steinsson (2010) consider a multi-sector model with a round-about production structure and obtain a peak output response of 0.80. Gertler and Leahy (2008) model segmented labor markets and report a real response of 0.75. Burstein and Hellwig (2007) incorporate decreasing returns to scale and wage rigidity and obtain 0.70, whereas Blanco et al. (2022) builds a similar single-product model with DRS and find a peak output response of 0.80. Lastly, Mongey (2021) studies a pricing model with duopoly competition and finds a peak response of 0.74. Relative to the standard setup with monopolistic competition, the duopoly model generates a CIR that is 2.3 times as large.

firms as deviation from the prices of competitors are costly. Therefore, firms are hesitant to respond aggressively, if at all, to the nominal shock if a sizeable share of other firms are not responding. Here, we show that a degree of strategic complementarity, as controlled by ψ and summarized by the cost pass-through to prices, that is consistent with empirical evidence contributes to significant non-neutrality. Second, the presence of both productivity and demand shocks plays a role in weakening the selection effect in price adjustments to an aggregate shock which further contributes to a larger output response. As shown in Golosov and Lucas (2007), the real response to nominal shocks hinges not on how many prices adjust but which prices adjust. In the absence of idiosyncratic shocks, prices respond only to aggregate shocks and the only prices that adjust are those that are most out of line with the aggregate shock. Adding idiosyncratic shocks weakens the selection effect as firms respond not only to aggregate shocks but also to disturbances to their idiosyncratic states. Having two idiosyncratic shocks further weakens this self-selection as firms in our model response to two orthogonal shocks in productivity and demand on top of shocks to aggregate nominal expenditure.

3.5 Conclusion

The advent of rich micro datasets has spurred the advancement of models of firm dynamics over the past twenty years to study a range of macroeconomic and international topics, including monetary non-neutrality, exchange-rate pass-through, and firm markups. From each of these literatures, key modeling ingredients have emerged as important for explaining aggregate and firm-level dynamics. From the exchange-rate pass-through and markup-dynamics literatures, models of strategic complementarity have been shown to play an important role. From the price-setting literature, nominal rigidities play a central role, but additional features in the form of real rigidities have also been necessary to approximate the degree of monetary non-neutrality from quantitative estimates.

However, each of these literatures has faced limitations that have prevented the emergence of a single modeling framework able to deliver the main results across all of these areas of study. From the price-setting literature, Klenow and Willis (2016) show that in order to generate realistic price-adjustment moments, a model with strategic complementarities requires a large magnitude of idiosyncratic productivity shocks that is inconsistent with micro evidence. A related limitation across these literatures has been the absence of micro data sets containing both prices and quantities at the firm level. Thus, most of these studies estimate or calibrate the parameters of idiosyncratic productivity processes to match pricing moments, because these pricing datasets lack data on quantities.

We propose a parsimonious framework that brings together the modeling elements from these literatures and also resolves prior limitations on the selection of modeling ingredients. First, we calibrate separate idiosyncratic shock processes for demand and productivity using empirical estimates from the firm-dynamics literature that employ both prices and quantities, along with revenue and inputs. This eliminates the need to calibrate these shock process parameters to produce observed pricing moments without any connection to quantities. Second, this approach allows us to re-investigate the role for strategic complementarities in a richer structure than was used in prior studies.

Our calibrated model is able to generate real and nominal dynamics that match key features of interest across these literatures. The combination of menu costs, a Kimball demand system, and idiosyncratic shocks for demand and productivity in the model produces cost pass-through in line with previous studies along with a distribution of markups that is similar to that estimated from the data. The model also produces simulated moments consistent with non-targeted moments on the size, direction, and dispersion of price changes. And when the model is extended to feature nominal expenditure shocks, it produces real effects from a nominal shock that are within the range of results from other studies of monetary non-neutrality. The two features generating this non-neutrality are the Kimball demand system and the presence of both idiosyncratic productivity and demand shocks, the latter of which dampens the selection effect in price adjustments to an aggregate shock and results in a larger real output response. By calibrating the model to directly match firm-level empirical estimates of the underlying shock processes for productivity and demand, we are able to avoid the critique of Klenow and Willis (2016), while also highlighting the importance of the joint inclusion of idiosyncratic productivity and demand shock processes in the study of models of strategic complementarities. Thus, we view our work as opening the door to future research that jointly models real (e.g. investment, employment, entry / exit), nominal (price setting) and other (e.g. markup, pass-through) decisions of firms using one unified framework.

Appendix A: Chapter 3

A.1 Derivations

A.1.1 Pass-through

Consider the static optimization problem of an intermediate firm without any pricing frictions. The nominal profit of an intermediate firm is

$$\pi_i = \left(\frac{p_i}{P} - \frac{W}{Pz_i}\right)\frac{Y}{n_i}\frac{1}{1+\psi}\left[\left(\frac{p_i}{\lambda n_i P}\right)^{\frac{\omega(1+\psi)}{1-\omega}} + \psi\right] \tag{A.1}$$

The first-order condition with respect to p_i is given by

$$\frac{1}{p_i - \frac{W}{z_i}} + \frac{\frac{\omega(1+\psi)}{1-\omega}\left(\frac{p_i}{\lambda n_i P}\right)^{\frac{\omega(1+\psi)}{1-\omega}-1}\frac{1}{\lambda n_i P}}{\left(\frac{p_i}{\lambda n_i P}\right)^{\frac{\omega(1+\psi)}{1-\omega}} + \psi} = 0 \tag{A.2}$$

$$\left(\frac{p_i}{\lambda n_i P}\right)^{\frac{\omega(1+\psi)}{1-\omega}}\left[1 - \left(\frac{W}{z_i} - p_i\right)\left(\frac{\omega\left(1+\psi\right)}{1-\omega}\frac{1}{p_i}\right)\right] = -\psi \tag{A.3}$$

Log-linearizing the first-order condition around a symmetric steady state yield

$$\left(\hat{p}_i - \hat{\lambda} - \hat{P} - \hat{n}_i\right) + \frac{1 + \left(\frac{W}{\bar{z}_i} - \bar{p}_i\right)\left(\frac{1}{\bar{p}_i}\right)}{1 - \left(\frac{W}{z_i} - p_i\right)\left(\frac{\omega(1+\psi)}{1-\omega}\frac{1}{p_i}\right)}\hat{p}_i + \frac{\frac{W}{\bar{z}_i}\left(\frac{1}{\bar{p}_i}\right)}{1 - \left(\frac{W}{z_i} - p_i\right)\left(\frac{\omega(1+\psi)}{1-\omega}\frac{1}{p_i}\right)}\hat{z}_i = 0 \tag{A.4}$$

where hatted variables denote log-deviations from the steady state.

Note that in a symmetric steady state, all firms are identical, have the same market share, and set the same price. Specifically, the optimal price is a fixed markup over cost $\bar{p}_i = \omega \frac{W}{\bar{z}_i}$ as demand does not matter in the steady state where all firms have the same demand. Substituting this into the log-linearized first-order condition gives

$$\hat{p}_i = \frac{\omega\psi}{\omega\psi - 1}\left(\hat{\lambda} + \hat{P} + \hat{n}_i\right) + \frac{1}{\omega\psi - 1}\hat{z}_i \tag{A.5}$$

A.1.2 Transmission of Nominal Shock under Calvo Pricing and CES

Consider a simple model where CES demand and Calvo pricing. Each period, a random fraction α of firms can adjust their prices freely while the other fraction $1-\alpha$ cannot change their prices. For simplicity, assume that there are no aggregate risks and the economy is initially in a symmetric equilibrium where all firms set nominal prices to $\bar{p}$ and nominal expenditure $S = PC$ is equal to $\bar{S}$.

Under the Calvo setup, the aggregate price index P_t can be written as a combination of the optimal reset price X_t, which is the price chosen by firms that are adjusting, and the lagged price index, which summarizes the prices of non-adjusters, as follows

$$P_t = \left[\alpha X_t^{1-\theta} + (1-\alpha) P_{t-1}^{1-\theta}\right]^{\frac{1}{1-\theta}} \tag{A.6}$$

Log-linearizing around the initial steady state where $P_t = X_t = \bar{p}$ and using hatted variables to denote log-deviations from the steady state yields

$$\hat{P}_t^{1-\theta} = \alpha \hat{X}_t^{1-\theta} + (1-\alpha)\,\hat{P}_{t-1}^{1-\theta} \tag{A.7}$$

Suppose that in period $t = 1$, there is an unanticipated permanent shock $\mu > 0$ to nominal expenditure shock S, such that $\hat{S}_1 = \mu$. Because the nominal wage is proportional to nominal expenditure, firms with the opportunity to adjust will respond to the shock by increasing their prices by μ as the optimal markup is a constant $\frac{\theta}{\theta-1}$ over the marginal cost. In other words, the optimal reset price is $\hat{X} = \mu$.

As such, the aggregate price in period one is given by

$$\hat{P}_1 = \alpha\mu \tag{A.8}$$

Iterating forward, the aggregate price in period h can be written as

$$\hat{P}_h = \left(\sum_{i=1}^{h} (1-\alpha)^{i-1}\right) \alpha\mu \tag{A.9}$$

The response of real output is therefore given by

$$\begin{aligned} \hat{C}_h &= \hat{S} - \hat{P}_h && \text{(A.10)} \\ &= \mu - \left(\sum_{i=1}^{h} (1-\alpha)^{i-1}\right) \alpha\mu && \text{(A.11)} \\ &= (1-\alpha)^h \mu && \text{(A.12)} \end{aligned}$$

which implies that the output response as a fraction of the shock is $(1-\alpha)^h$ at period h.

A.2 Model Solution

A.2.1 Rewriting the Problem

Note that firms need to observe $\mathcal{S}$ and know its law of motion in order to solve their problem. These relevant aggregate variables in $\mathcal{S}$ can be summarized by a single aggregate state variable P_{t-1}/S_t.

Because money supply $S_t = P_t C_t$ exhibits positive growth on average, nominal prices are also ever-increasing. We normalize all nominal variables by S_t to ensure that the state variables are stationary. As such, we can rewrite the firm's profit function

$$\pi\left(\frac{p_t^i}{S_t}; n_t^i, z_t^i, \frac{P_t}{S_t}, \lambda_t\right) = \left(\frac{p_t^i/S_t}{P_t/S_t} - \frac{1}{z_t^i P_t/S_t}\right) \frac{(P_t/S_t)^{-1}}{n_t^i\,(1+\psi)} \left[\left(\frac{p_t^i/S_t}{\lambda_t n_t^i\,(P_t/S_t)}\right)^{\frac{\omega(1+\psi)}{1-\omega}} + \psi\right] \tag{A.13}$$

where we also use $Y_t = C_t = \frac{P_t}{S_t}$ from goods market clearing and $W_t = S_t$ from the household's intratemporal optimality condition.

P_t/S_t and λ_t are the collective results of the pricing decisions of all firms. To know these, firms must know the entire firm distribution over the idiosyncratic states which is an infinite-dimensional object. Following the application of the Krusell and Smith (1998) algorithm in menu-cost models (Midrigan, 2011; Nakamura and Steinsson, 2010; Vavra, 2014), we conjecture the following forecasting rules for P_t/S_t and λ_t

$$\log\left(\frac{P_t}{S_t}\right) = F\left(\frac{P_{t-1}}{S_t}\right) = \alpha_0 + \alpha_1 \log\left(\frac{P_{t-1}}{S_t}\right) \tag{A.14}$$

$$\log(\lambda_t) = G\left(\frac{P_{t-1}}{S_t}\right) = \beta_0 + \beta_1 \log\left(\frac{P_{t-1}}{S_t}\right) \tag{A.15}$$

Using these, the law of motion of the aggregate variable P_{t-1}/S_t is also given by

$$\log\left(\frac{P_t}{S_{t+1}}\right) = \log\left(\frac{P_t}{S_t}\right) + \log\left(\frac{S_t}{S_{t+1}}\right) \tag{A.16}$$

$$= \alpha_0 + \alpha_1 \log\left(\frac{P_{t-1}}{S_t}\right) - (\mu + \sigma_S \epsilon_{t+1}) \tag{A.17}$$

Now, we rewrite the intermediate producers' problem using these state variables. At the beginning of a period, each intermediate producer starts off with a price p^i_{t-1}/S_t, idiosyncratic demand n^i_t, and idiosyncratic productivity z^i_t. They also observe P_{t-1}/S_t and forecast P_t/S_t and λ_t using the aforementioned laws of motion. The value of not adjusting is

$$V_N\left(\frac{p^i_{t-1}}{S_t}; n^i_t, z^i_t, \frac{P_{t-1}}{S_t}\right) = \pi\left(\frac{p^i_{t-1}}{S_t}; n^i_t, z^i_t, \frac{P_{t-1}}{S_t}\right) + \mathbb{E}_t\left[\Xi_{t,t+1} \cdot V\left(\frac{p^i_t}{S_{t+1}}, n^i_{t+1}, z^i_{t+1}, \frac{P_t}{S_{t+1}}\right)\right] \tag{A.18}$$

which is equal to the flow profit evaluated at last period's price adjusted for inflation plus a continuation value.

If the firm chooses to adjust its price, it pays the fixed price adjustment cost and chooses p^i_t to maximize the sum of the current flow profit and the present discounted value of future profit

$$V_A\left(n^i_t, z^i_t, \frac{P_{t-1}}{S_t}\right) = -f^i_t \frac{P_t}{S_t} + \max_{p^i_t}\left\{\pi\left(\frac{p^i_t}{S_t}; n^i_t, z^i_t, \frac{P_{t-1}}{S_t}\right) + \mathbb{E}_t\left[\Xi_{t,t+1} \cdot V\left(\frac{p^i_t}{S_{t+1}}, n^i_{t+1}, z^i_{t+1}, \frac{P_t}{S_{t+1}}\right)\right]\right\} \tag{A.19}$$

A firm chooses to adjust its price if and only if the value of doing so exceeds the value of inaction. Therefore, the value function of the firm is

$$V\left(\frac{p^i_{t-1}}{S_t}; n^i_t, z^i_t, \frac{P_{t-1}}{S_t}\right) = \max\left[V_N\left(\frac{p^i_{t-1}}{S_t}; n^i_t, z^i_t, \frac{P_{t-1}}{S_t}\right), V_A\left(n^i_t, z^i_t, \frac{P_{t-1}}{S_t}\right)\right] \tag{A.20}$$

A.2.2 Computational Strategy

A sketch of the computation algorithm is as follows. We first make guesses of the coefficients $(\alpha_0^0, \alpha_1^0, \beta_0^0, \beta_1^0)$ in the forecasting equations F and G. Given the guesses, use value function iteration to solve for the intermediate-good producers' value functions as well as the optimal pricing rules. Using pricing rules, simulate the model for a large number of periods and obtain simulated sequences of $\frac{P_t}{S_t}$, λ_t, and $\frac{P_{t-1}}{S_t}$. Estimate the regressions F and G with model simulated data and obtain estimated coefficients $(\alpha_0^1, \alpha_1^1, \beta_0^1, \beta_1^1)$ which is then used to update the initial guesses. Repeat this process until the coefficient guesses and sufficiently close to the estimated coefficients from the linear regressions. In doing so, we find that the conjectured law of motion approximates the true law of motion from the model simulation well, as the regression yields an R^2 larger than 0.99.

Bibliography

Abbott, Brant, Giovanni Gallipoli, Costas Meghir, and Giovanni Violante, "Education Policy and Intergenerational Transfers in Equilibrium," *Journal of Political Economy*, 2019.

Acemoglu, Daron, "Technical Change, Inequality, and the Labor Market," *Journal of Economic Literature*, 2002, *40* (1), 7–72.

Adamopoulos, Tasso, Loren Brandt, Jessica Leight, and Diego Restuccia, "Misallocation, selection and productivity: A quantitative analysis with panel data from China," Technical Report, National Bureau of Economic Research 2017.

Agostinelli, Francesco and Giuseppe Sorrenti, "Money vs. Time: Family Income, Maternal Labor Supply, and Child Development," *Working Paper*, 2018, (100018 165616).

Aiyagari, S Rao, "Uninsured idiosyncratic risk and aggregate saving," *The Quarterly Journal of Economics*, 1994, *109* (3), 659–684.

Akerlof, George A and Janet L Yellen, "Can Small Deviations from Rationality Make Significant Differences to Economic Equilibria?," *The American Economic Review*, 1985, *75* (4), 708–720.

Alesina, Alberto, Andrea Ichino, and Loukas Karabarbounis, "Gender-based Taxation and the Division of Family Chores," *American Economic Journal: Economic Policy*, 2011, *3* (2), 1–40.

Alvarez, Fernando E, Francesco Lippi, and Takis Souganidis, "Price setting with strategic complementarities as a mean field game," Technical Report, National Bureau of Economic Research 2022.

Alvarez, Fernando, Hervé Le Bihan, and Francesco Lippi, "The real effects of monetary shocks in sticky price models: a sufficient statistic approach," *American Economic Review*, 2016, *106* (10), 2817–2851.

Amiti, Mary, Oleg Itskhoki, and Jozef Konings, "International Shocks, Variable Markups, and Domestic Prices," *The Review of Economic Studies*, 2019, *86* (6), 2356–2402.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare, "The Elusive Pro-competitive Effects of Trade," *The Review of Economic Studies*, 2019, *86* (1), 46–80.

Aruoba, S Boragan, Andrés Fernández, Daniel Guzmán, Ernesto Pastén, and Felipe Saffie, "Pricing under Distress," 2022.

Autor, David, David Dorn, and Gordon Hanson, "When work disappears: Manufacturing decline and the falling marriage market value of young men," *American Economic Review: Insights*, 2019, *1* (2), 161–178.

Baley, Isaac and Andrés Blanco, "Firm uncertainty cycles and the propagation of nominal shocks," *American Economic Journal: Macroeconomics*, 2019, *11* (1), 276–337.

Ball, Laurence and David Romer, "Real Rigidities and the Mon-neutrality of Money," *The Review of Economic Studies*, 1990, *57* (2), 183–203.

Bardóczy, Bence, "Spousal Insurance and the Amplification of Business Cycles," Technical Report 2020.

Becker, Gary S and Nigel Tomes, "An equilibrium theory of the distribution of income and intergenerational mobility," *Journal of political Economy*, 1979, *87* (6), 1153–1189.

Berger, David and Joseph Vavra, "Shocks versus Responsiveness: What Drives Time-Varying Dispersion?," *Journal of Political Economy*, 2019, *127* (5), 2104–2142.

Bils, Mark and Peter J Klenow, "Some evidence on the importance of sticky prices," *Journal of political economy*, 2004, *112* (5), 947–985.

Birinci, Serdar, "Spousal Labor Supply Response to Job Displacement and Implications for Optimal Transfers," *FRB St. Louis Working Paper*, 2019, (2019-020).

Blanco, Andrés, "Optimal Inflation Target in an Economy with Menu Costs and a Zero Lower Bound," *American Economic Journal: Macroeconomics*, 2021, *13* (3), 108–41.

Blanco, Andres, Corina Boar, Callum Jones, and Virgiliu Midrigan, "Nonlinear Inflation Dynamics in Menu Cost Economies," 2022.

Blau, Francine D and Lawrence M Kahn, "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of economic literature*, 2017, *55* (3), 789–865.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten, "Consumption inequality and family labor supply," *American Economic Review*, 2016, *106* (2), 387–435.

Borella, Margherita, Mariacristina De Nardi, and Fang Yang, "The Effects of Marriage-related Taxes and Social Security Benefits," Technical Report, National Bureau of Economic Research 2017.

_ , _ , and _ , "Are Marriage-Related Taxes and Social Security Benefits Holding Back Female Labour Supply?," *The Review of Economic Studies*, 2023, *90* (1), 102–131.

__ , __ , **Michael Pak, Nicolo Russo, and Fang Yang**, "The Importance of Modeling Income Taxes Over time. US Reforms and Outcomes," Technical Report, National Bureau of Economic Research 2022.

Botezat, Alina and Friedhelm Pfeiffer, "The Impact of Parents Migration on the Well-Being of Children Left Behind Initial Evidence from Romania," *IZA Discussion Paper Series*, 2014, (8225).

Braxton, J Carter, Kyle F Herkenhoff, Jonathan L Rothbaum, and Lawrence Schmidt, "Changing Income Risk across the US Skill Distribution: Evidence from a Generalized Kalman Filter," Technical Report, National Bureau of Economic Research 2021.

Bronson, Mary Ann and Maurizio Mazzocco, "Taxation and household decisions: an intertemporal analysis," Technical Report 2021.

Bryan, Gharad and Melanie Morten, "The aggregate productivity effects of internal migration: Evidence from indonesia," *Journal of Political Economy*, 2019, *127* (5), 2229–2268.

Burstein, Ariel and Christian Hellwig, "Prices and Market Shares in a Menu Cost Model," 2007.

Campa, Jose Manuel and Linda S Goldberg, "Exchange rate pass-through into import prices," *Review of Economics and Statistics*, 2005, *87* (4), 679–690.

Caplin, Andrew and John Leahy, "Aggregation and optimization with state-dependent pricing," *Econometrica*, 1997, pp. 601–625.

Caplin, Andrew S and Daniel F Spulber, "Menu Costs and the Neutrality of Money," *The Quarterly Journal of Economics*, 1987, *102* (4), 703–725.

Carlsson, Mikael, Alex Clymo, and Knut-Eric Joslin, "Dispersion Over The Business Cycle: Passthrough, Productivity, And Demand," Technical Report 2022.

Carneiro, Pedro and James J. Heckman, "The Evidence on Credit Constraints in Post-secondary Schooling," *Economic Journal*, 2002, *112* (482), 705–734.

Caucutt, Elizabeth and Lance John Lochner, "Early and Late Human Capital Investments, Borrowing Constraints, and the Family," *Journal of Political Economy*, 2019.

__ **and** __ , "Early and Late Human Capital Investments, Borrowing Constraints, and the Family," *Journal of Political Economy*, 2019.

Chamon, Marcos, Kai Liu, and Eswar Prasad, "Income uncertainty and household savings in China," *Journal of Development Economics*, 2013, *105*, 164–177.

Chen, Joyce J, "Identifying non-cooperative behavior among spouses: Child outcomes in migrant-sending households," *Journal of Development Economics*, 2013, *100* (1), 1–18.

Chen, Zexing, Bing Li, and Tao Li, "Exports and left-behind children: Empirical evidence from the China Migrants Dynamic Survey," *Review of International Economics*, 2019, *27* (4), 1081–1107.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *The Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

Choo, Eugene and Aloysius Siow, "Who marries whom and why," *Journal of political Economy*, 2006, *114* (1), 175–201.

Christiano, Lawrence J, Martin Eichenbaum, and Charles L Evans, "Monetary Policy Shocks: What Have We Learned and to What End?," *Handbook of Macroeconomics*, 1999, *1*, 65–148.

Coibion, Olivier, Yuriy Gorodnichenko, and Johannes Wieland, "The Optimal Inflation Rate in New Keynesian Models: Should Central Banks Raise Their Inflation Targets in Light of the Zero Lower Bound?," *The Review of Economic Studies*, 03 2012, *79* (4), 1371–1406.

Cunha, Flavio and James Heckman, "The Technology of Skill Formation," *American Economic Review*, 2007, *97* (2), 31–47.

_, James J Heckman, and Susanne M Schennach, "Estimating the technology of cognitive and noncognitive skill formation," *Econometrica*, 2010, *78* (3), 883–931.

Daruich, Diego, "The Macroeconomic Consequences of Early Childhood Development Policies," 2019.

_ , "The Macroeconomic Consequences of Early Childhood Development Policies," 2022.

Davis, Steven J, John Haltiwanger, Ron Jarmin, Javier Miranda, Christopher Foote, and Eva Nagypal, "Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms," *NBER Macroeconomics Annual*, 2006, *21*, 107–179.

Diamond, Peter A, "Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates," *American Economic Review*, 1998, pp. 83–95.

Doepke, Matthias and Michèle Tertilt, "Families in Macroeconomics," in "Handbook of Macroeconomics," Vol. 2, Elsevier, 2016, pp. 1789–1891.

_, Giuseppe Sorrenti, and Fabrizio Zilibotti, "The Economics of Parenting," *Annual Review of Economics*, 2019, *11*, 55–84.

Dotsey, Michael and Robert G King, "Implications of State-dependent Pricing for Dynamic Macroeconomic Models," *Journal of Monetary Economics*, 2005, *52* (1), 213–242.

_, _, and Alexander L Wolman, "State-dependent pricing and the general equilibrium dynamics of money and output," *The Quarterly Journal of Economics*, 1999, *114* (2), 655–690.

Dustmann, Christian, "Return migration, uncertainty and precautionary savings," *Journal of Development Economics*, 1997, *52* (2), 295–316.

— **and Albrecht Glitz**, "Migration and education," in "Handbook of the Economics of Education," Vol. 4, Elsevier, 2011, pp. 327–439.

— **and Joseph-Simon Görlach**, "The economics of temporary migrations," *Journal of Economic Literature*, 2016, *54* (1), 98–136.

Edin, Kathryn and Joanna M Reed, "Why don't they just get married? Barriers to marriage among the disadvantaged," *The Future of Children*, 2005, pp. 117–137.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "How Costly are Markups?," Technical Report, National Bureau of Economic Research 2018.

Eichenbaum, Martin and Jonas DM Fisher, "Estimating the Frequency of Price Re-optimization in Calvo-style Models," *Journal of Monetary Economics*, 2007, *54* (7), 2032–2047.

Eslava, Marcela, John Haltiwanger, Adriana Kugler, and Maurice Kugler, "Trade and Market Selection: Evidence from Manufacturing Plants in Colombia," *Review of Economic Dynamics*, 2013, *16* (1), 135–158.

— **,** — **, and Nicolas Urdaneta**, "The Size and Life-Cycle Growth of Plants: The Role of Productivity, Demand, and Wedges," Technical Report 2022.

Fan, Jingting, "Internal Geography, Labor Mobility, and the Distributional Impacts of Trade," *American Economic Journal: Macroeconomics*, 2019, *11* (3), 252–288.

Fan, Yi, Junjian Yi, and Junsen Zhang, "Rising Intergenerational Income Persistence in China," *American Economic Journal: Economic Policy*, Forthcoming, pp. 1–48.

Fernández, Raquel and Richard Rogerson, "Sorting and Long-Run Inequality," *The Quarterly Journal of Economics*, 2001, *116* (4), 1305–1341.

Fernandez, Raquel, Nezih Guner, and John Knowles, "Love and money: A Theoretical and Empirical Analysis of Household Sorting and Inequality," *The Quarterly Journal of Economics*, 2005, *120* (1), 273–344.

Fogli, Alessandra and Veronica Guerrieri, "The End of the American Dream? Inequality and Segregation in US Cities," *NBER Working Paper*, 2019.

Foster, Lucia, John Haltiwanger, and Chad Syverson, "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *American Economic Review*, 2008, *98* (1), 394–425.

Garriga, Carlos, Aaron Hedlund, Yang Tang, and Ping Wang, "Rural-Urban Migration, Structural Transformation, and Housing Markets in China," Working Paper 23819, National Bureau of Economic Research September 2017.

Gayle, George-Levi and Andrew Shephard, "Optimal Taxation, Marriage, Home Production, and Family Labor Supply," *Econometrica*, 2019, *87* (1), 291–326.

Gertler, Mark and John Leahy, "A Phillips Curve with an Ss Foundation," *Journal of Political Economy*, 2008, *116* (3), 533–572.

Golosov, Mikhail and Robert E. Jr. Lucas, "Menu Costs and Phillips Curves," *Journal of Political Economy*, 2007, *115* (2), 171–199.

Gopinath, Gita and Oleg Itskhoki, "Frequency of Price Adjustment and Pass-through," *The Quarterly Journal of Economics*, 2010, *125* (2), 675–727.

_ , _ , and Roberto Rigobon, "Currency choice and exchange rate pass-through," *American Economic Review*, 2010, *100* (1), 304–336.

Gorbachev, Olga, "Has the Increased Attachment of Women to the Labor Market Changed a Family's Ability to Smooth Income Shocks?," *American Economic Review*, 2016, *106* (5), 247–251.

Gottschalk, Peter, Robert Moffitt, Lawrence F Katz, and William T Dickens, "The Growth of Earnings Instability in the US Labor Market," *Brookings Papers on Economic Activity*, 1994, *1994* (2), 217–272.

Gould, Eric D and M Daniele Paserman, "Waiting for Mr. Right: rising inequality and declining marriage rates," *Journal of Urban Economics*, 2003, *53* (2), 257–281.

Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov, and Cezar Santos, "Marry your like: Assortative mating and income inequality," *American Economic Review*, 2014, *104* (5), 348–53.

_ , _ , _ , and _ , "Technology and the changing family: A unified model of marriage, divorce, educational attainment, and married female labor-force participation," *American Economic Journal: Macroeconomics*, 2016, *8* (1), 1–41.

_ , Richard Rogerson, and Randall Wright, "Household Production in Real Business Cycle Theory," in "Frontiers of Business Cycle Research," Princeton University Press, 1995, pp. 157–174.

Guner, Nezih, Remzi Kaygusuz, and Gustavo Ventura, "Taxation and Household Labour Supply," *The Review of economic studies*, 2012, *79* (3), 1113–1149.

Guvenen, Fatih, Alisdair McKay, and Conor Ryan, "A Tractable Income Process for Business Cycle Analysis," Technical Report 2022.

_ , Greg Kaplan, Jae Song, and Justin Weidner, "Lifetime Incomes in the United States over Six Decades," Technical Report, National Bureau of Economic Research 2017.

_ , Serdar Ozkan, and Jae Song, "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, 2014, *122* (3), 621–660.

Haan, Peter and Victoria L Prowse, "Optimal Social Assistance and Unemployment Insurance in a Life-cycle Model of Family Labor Supply and Savings," Technical Report 2020.

Harding, Martín, Jesper Lindé, and Mathias Trabandt, "Resolving the Missing Deflation Puzzle," *Journal of Monetary Economics*, 2022, *126*, 15–34.

Harding, Martın, Jesper Lindé, and Mathias Trabandt, "Understanding Post-Covid Inflation Dynamics," *Manuscript*, 2022.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L Violante, "The macroeconomic implications of rising wage inequality in the United States," *Journal of political economy*, 2010, *118* (4), 681–722.

_ , _ , and _ , "Consumption and Labor Supply with Partial Insurance: An Analytical Framework," *American Economic Review*, 2014, *104* (7), 2075–2126.

_ , _ , and _ , "Optimal tax progressivity: An analytical framework," *The Quarterly Journal of Economics*, 2017, *132* (4), 1693–1754.

_ , _ , and _ , "How Should Tax Progressivity Respond to Rising Income Inequality?," *Journal of the European Economic Association*, 2020, *18* (6), 2715–2754.

Huggett, Mark, "Wealth distribution in life-cycle economies," *Journal of Monetary Economics*, 1996, *38* (3), 469–494.

Ilin, Elias, Laurence J Kotlikoff, and Melinda Pitts, "Is Our Fiscal System Discouraging Marriage? A New Look at the Marriage Tax," Technical Report, National Bureau of Economic Research 2022.

Karahan, Fatih and Serdar Ozkan, "On the Persistence of Income Shocks over the Life Cycle: Evidence, Theory, and Implications," *Review of Economic Dynamics*, 2013, *16* (3), 452–476.

Katz, Lawrence F and Kevin M Murphy, "Changes in relative wages, 1963–1987: supply and demand factors," *The Quarterly Journal of Economics*, 1992, *107* (1), 35–78.

Keane, Michael and Kenneth Wolpin, "The Career Decisions of Young Men," *Journal of Political Economy*, 1997, *105* (3), 473–522.

Keane, Michael P. and Kenneth I. Wolpin, "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment," *International Economic Review*, 2001, *42* (4).

Kimball, Miles S, "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit, and Banking*, 1995, *27* (4), 1241.

Klenow, Peter J and Jonathan L Willis, "Real Rigidities and Nominal Price Changes," *Economica*, 2016, *83* (331), 443–472.

_ and Oleksiy Kryvtsov, "State-dependent or time-dependent pricing: Does it matter for recent US inflation?," *The Quarterly Journal of Economics*, 2008, *123* (3), 863–904.

Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez, "The optimal income taxation of couples," *Econometrica*, 2009, *77* (2), 537–560.

Krueger, Alan, "The Rise and Consequences of Inequality in the United States," Jan 2012.

Krueger, Dirk and Alexander Ludwig, "On the Optimal Provision of Social Insurance: Progressive Taxation versus Education Subsidies in General Equilibrium," *Journal of Monetary Economics*, 2016, *77*, 72–98.

Krusell, Per and Anthony A Smith Jr, "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 1998, *106* (5), 867–896.

Lagakos, David, Ahmed Mushfiq Mobarak, and Michael E Waugh, "The welfare effects of encouraging rural-urban migration," Technical Report, National Bureau of Economic Research 2018.

Lee, Sang Yoon and Ananth Seshadri, "On the Intergenerational Transmission of Economic Status," *Journal of Political Economy*, 2019, *127* (2), 855–921.

Li, Haizheng and Steven Zahniser, "The determinants of temporary rural-to-urban migration in China," *Urban Studies*, 2002, *39* (12), 2219–2235.

Loecker, Jan De, Jan Eeckhout, and Gabriel Unger, "The Rise of Market Power and the Macroeconomic Implications," *The Quarterly Journal of Economics*, 2020, *135* (2), 561–644.

Mankiw, N Gregory, "Small Menu Costs and Large Business Cycles: A Macroeconomic Model of Monopoly," *The Quarterly Journal of Economics*, 1985, *100* (2), 529–537.

McGrattan, Ellen R, Richard Rogerson, and Randall Wright, "An Equilibrium Model of the Business Cycle with Household Production and Fiscal Policy," *International Economic Review*, 1997, pp. 267–290.

Meghir, Costas and Luigi Pistaferri, "Income Variance Dynamics and Heterogeneity," *Econometrica*, 2004, *72* (1), 1–32.

Meng, Xin, "Labor market outcomes and reforms in China," *Journal of Economic Perspectives*, 2012, *26* (4), 75–102.

_ **and Chikako Yamauchi**, "Children of Migrants: the Cumulative Impact of Parental Migration on Children's Education and Health Outcomes in China," *Demography*, 2017, *54* (5), 1677–1714.

Meyerhoefer, Chad D and CJ Chen, "The Effect of Parental Labor Migration on Children's Educational Progress in Rural China," *Review of Economics of the Household*, 2011, *9* (3), 379–396.

Midrigan, Virgiliu, "Menu Costs, Multiproduct Firms, and Aggregate Fluctuations," *Econometrica*, 2011, *79* (4), 1139–1180.

Mirrlees, James A, "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 1971, *38* (2), 175–208.

Moffitt, Robert A, "Reconciling Trends in US Male Earnings Volatility: Results from a Four Data Set Project," Technical Report, National Bureau of Economic Research 2020.

Mongey, Simon, "Market Structure and Monetary Non-neutrality," Technical Report, National Bureau of Economic Research 2021.

Mu, Ren and Alan De Brauw, "Migration and Young Child Nutrition: Evidence from Rural China," *Journal of Population Economics*, 2015, *28* (3), 631–657.

Nakamura, Emi and Jón Steinsson, "Five Facts about Prices: A Reevaluation of Menu Cost Models," *The Quarterly Journal of Economics*, 2008, *123* (4), 1415–1464.

_ **and Jon Steinsson**, "Monetary Non-neutrality in a Multisector Menu Cost Model," *The Quarterly Journal of Economics*, 2010, *125* (3), 961–1013.

Nardi, Mariacristina De, Giulio Fella, and Gonzalo Paz-Pardo, "Nonlinear Household Earnings Dynamics, Self-insurance, and Welfare," *Journal of the European Economic Association*, 2020, *18* (2), 890–926.

_ **,** _ **, and** _ , "Wage Risk and Government and Spousal Insurance," Technical Report, National Bureau of Economic Research 2021.

Navarro, Salvador and Jin Zhou, "Human Capital and Migration: a Cautionary Tale," Technical Report 2018.

Ngai, L Rachel, Christopher A Pissarides, and Jin Wang, "China's mobility barriers and employment allocations," *Journal of the European Economic Association*, 2018, *17* (5), 1617–1653.

Nickell, Stephen, Luca Nunziata, and Wolfgang Ochel, "Unemployment in the OECD since the 1960s. What do we know?," *The Economic Journal*, 2005, *115* (500), 1–27.

Ortigueira, Salvador and Nawid Siassi, "How Important is Intra-household Risk Sharing for Savings and Labor Supply?," *Journal of Monetary Economics*, 2013, *60* (6), 650–666.

Piketty, Thomas and Emmanuel Saez, "How progressive is the US federal tax system? A historical and international perspective," *Journal of Economic perspectives*, 2007, *21* (1), 3–24.

Ramakrishnan, Prasanthi, "Education, Marriage, and Social Security," Technical Report 2022.

Ramey, Valerie A, "Macroeconomic Shocks and their Propagation," *Handbook of Macroeconomics*, 2016, *2*, 71–162.

Regalia, Ferdinando, Jose-Victor Rios-Rull, and Jacob Short, "What Accounts for the Increase in the Number of Single Households?," 2019.

Restuccia, Diego and Carlos Urrutia, "Intergenerational Persistence of Earnings: the Role of Early and College Education," *American Economic Review*, 2004, *94* (5), 1354–1378.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek, "IPUMS USA: version 12.0 [dataset]," *Minneapolis, MN: IPUMS*, 2022, *10*, D010.

Saez, Emmanuel and Gabriel Zucman, "The Rise of Income and Wealth Inequality in America: Evidence from Distributional Macroeconomic accounts," *Journal of Economic Perspectives*, 2020, *34* (4), 3–26.

Santos, Cezar and David Weiss, ""Why not settle down already?" a quantitative analysis of the delay in marriage," *International Economic Review*, 2016, *57* (2), 425–452.

Shore, Stephen H and Todd Sinai, "Commitment, risk, and consumption: do birds of a feather have bigger nests?," *The Review of Economics and Statistics*, 2010, *92* (2), 408–424.

Sicular, Terry, Yue Ximing, Björn Gustafsson, and Li Shi, "The urban–rural income gap and inequality in China," *Review of Income and Wealth*, 2007, *53* (1), 93–126.

Sieg, Holger, Chamna Yoon, and Jipeng Zhang, "The impact of migration controls on urban fiscal policies and the intergenerational transmission of human capital in China," Technical Report, National Bureau of Economic Research 2020.

Smets, Frank and Rafael Wouters, "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 2007, *97* (3), 586–606.

Stevenson, Betsey and Justin Wolfers, "Marriage and Divorce: Changes and their Driving Forces," *Journal of Economic perspectives*, 2007, *21* (2), 27–52.

Tombe, Trevor and Xiaodong Zhu, "Trade, Migration, and Productivity: A Quantitative Analysis of China," *American Economic Review*, 2019, *109* (5), 1843–1872.

Vavra, Joseph, "Inflation Dynamics and Time-varying Volatility: New Evidence and an Ss Interpretation," *The Quarterly Journal of Economics*, 2014, *129* (1), 215–258.

Walker, Laquitta and Danielle Taylor, *Same-sex Couple Households, 2019*, US Department of Commerce, Economics and Statistics Administration, USA, 2021.

Wen, Yi, "Income and Living Standards across China," Jan 2018.

Wu, Chunzan, "More Unequal Income but Less Progressive Taxation," *Journal of Monetary Economics*, 2021, *117*, 949–968.

_ **and Dirk Krueger**, "Consumption Insurance Against Wage Risk: Family Labor Supply and Optimal Progressive Income Taxation," *American Economic Journal: Macroeconomics*, 2021, *13* (1), 79–113.

You, Wei and Wenbin Wu, "The Welfare Implications of Internal Migration Restrictions : Evidence from China," 2019.

Yu, Jihai and Guozhong Zhu, "How uncertain is household income in China," *Economics Letters*, 2013, *120* (1), 74–78.

Yum, Minchul, "Parental time investment and intergenerational mobility," Technical Report, Working Paper Series 2016.

Zhang, Hongliang, Jere R Behrman, C Simon Fan, Xiangdong Wei, and Junsen Zhang, "Does Parental Absence Reduce Cognitive Achievements? Evidence from Rural China," *Journal of Development Economics*, 2014, *111*, 181–195.

Zhao, Yaohui, "Leaving the countryside: rural-to-urban migration decisions in China," *American Economic Review*, 1999, *89* (2), 281–286.

_ , "Causes and consequences of return migration: recent evidence from China," *Journal of Comparative Economics*, 2002, *30* (2), 376–394.

Zheng, Angela and James Graham, "Public Education Inequality and Intergenerational Mobility," 2019, pp. 1–39.

Zhong, Ling, "Internal Migration and Extended Families in China," 2019.

Zhu, Nong, "The impacts of income gaps on migration decisions in China," *China Economic Review*, 2002, *13* (2-3), 213–230.