ESSAYS IN LABOR ECONOMICS

Weilong Zhang

A DISSERTATION

in

Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2018

Supervisor of Dissertation

Petra E. Todd, Edmund J. and Louise W. Kahn Term Professor of Economics

Graduate Group Chairperson

Jesús Fernández-Villaverde, Professor of Economics

Dissertation Committee

Christopher Flinn, Professor of Economics

Sarah Moshary, Assistant Professor of Economics

This thesis is dedicated to my parents Guohuai Zhang and Keping Zhang, for granting me the wisdom, health and strength to undertake and complete this research project.

ACKNOWLEDGEMENT

The path toward this dissertation has been circuitous. Its completion is impossible without the continuous challenging, support, and guidance from my dissertation committee members along the way. I owe a debt of gratitude to Professor Petra Todd for her untiring support and encouragement. In her own way, she motivates and nurtures me from a naive and inexperienced Ph.D. student into a mature researcher I am today. I feel extremely lucky to have her as my advisor who cares so much about my work, and who responds to my questions and queries so promptly. To Professor Chris Flinn, I am extremely grateful for the time and efforts he has invested into my work and will forever cherish the time we have spent together, discussing ideas, brainstorming, and coming up with directions in my work. I will forever be indebted to him for pushing me to my limits. To Dr. Sarah Moshary, thank you for your time and careful attention to details despite your overwhelming schedule.

To my friends and roommates, thank you for the debates, dinners, and game nights as well as editing advice, supports during the job search, and general help and friendship were all greatly appreciated. Special mention goes to Zhesheng Qiu, for encouraging me to embark the journey of economic research. And to Hongxun Ruan, with whom I have shared moments of deep anxiety but also of big excitement.

Last but not the least, I would like to thank my fiancée, Yu Zheng for her unconditional love, support, and understanding. Without her, I would not be able to make such an accomplishment.

ABSTRACT

ESSAYS IN LABOR ECONOMICS

Weilong Zhang

Petra E. Todd

This thesis consists of three chapters. They explore develop and estimate economic models to analyze questions of interests to public policies.

Chapter 1 develops and estimates a spatial general equilibrium job search model to study the effects of local and universal (federal) minimum wage policies. In the model, firms post vacancies in multiple locations. Workers, who are heterogeneous in terms of location and education types, engage in random search and can migrate or commute in response to job offers. The model is estimated by combining multiple databases including the American Community Survey (ACS) and Quarterly Workforce Indicators (QWI). The estimated model is used to analyze how minimum wage policies affect employment, wages, job postings, vacancies, migration/commuting, and welfare. Empirical results show that minimum wage increases in local county lead to an exit of low type (education<12 years) workers and an influx of high type workers (education>12 years), which generates negative externalities for workers in neighboring areas. The model is used to simulate the effects of a range of minimum wages. Minimum wage increases up to \$14/hour increase the welfare of high type workers but lower welfare of low type workers, expanding inequality. Increases in excess of \$14/hour decrease welfare for all workers. Two counterfactual policies are further evaluated under this framework: restricting labor mobility and preempting local minimum wage laws. For a certain range of minimum wages, both policies have negative impacts on the welfare of high type workers, but beneficial effects for low type workers.

Chapter 2 poses a dynamic discrete choice model of schooling and occupational choices that incorporates time-varying personality traits, as measured by the so-called "Big Five" traits.

The model is estimated using the Household Income and Labor Dynamics in Australia (HILDA) longitudinal dataset from Australia. Personality traits are found to play a critical role in explaining education and occupational choices over the lifecycle. The traits evolve during young adult years but stabilize in the mid-30s. Results show that individuals with a comparative advantage in schooling and white-collar work have, on average, higher cognitive skills and higher personality traits, in all five dimensions. The estimated model is used to evaluate two education policies: compulsory senior secondary school and a 50% college subsidy. Both policies are found to be effective in increasing educational attainment, but the compulsory schooling policy provides greater benefits to lower socioeconomic groups. Allowing personality traits to evolve with age and with years of schooling proves to be important in capturing policy response heterogeneity.

Chapter 3 develops and estimates a model of how personality traits affect household time and resource allocation decisions and wages. In the model, households choose between two behavioral modes: cooperative or noncooperative. Spouses receive wage offers and allocate time to supply labor market hours and to produce a public good. Personality traits, measured by the so-called "Big Five" traits, can affect household bargaining weights and wage offers. Model parameters are estimated by Simulated Method of Moments using the Household Income and Labor Dynamics in Australia (HILDA) data. Personality traits are found to be important determinants of household bargaining weights and of wage offers and to have substantial implications for understanding the sources of gender wage disparities.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1 : Distributional Effects of Local Minimum Wage Hikes: A Spatial Job Search Approach

Weilong Zhang

1.1. Introduction

Traditional minimum wage studies estimate local labor market employment and wage effects by comparing a group that experienced the minimum wage change to a similar group in nearby region that did not experience a change.[1] This approach can be problematic when local minimum wage changes are large, because substantial local minimum wage increases likely induce labor mobility and have spillover effects on neighboring areas.[2] A full accounting of minimum wage effects must take into account workers from all affected areas.[3] Furthermore, when faced with higher labor costs, firms may substitute lower productivity workers with higher productivity ones to keep profitable (Horton, 2017). Therefore, some workers may benefit from minimum wage increases, while others are adversely affected. This paper studies the distributional and welfare effects of local and universal (federal) minimum wage policies taking into account worker heterogeneity, spatial mobility, and minimum wages of varying magnitudes.

To this end, I develop a spatial general equilibrium model that extends Flinn (2006) to a spatial search context. The economy consists of two adjacent regions, similar to the cross-border contiguous county pairs in Dube et al. (2016). Workers are differentiated by their types and locations.[4] They receive job offers from local firms and from firms in a neighboring

[1]There is an ongoing debate concerning the effect of minimum wages on employment. See Card and Krueger (1994, 2000); Dube et al. (2007, 2010, 2016); Neumark (2001); Neumark et al. (2014a,b); Jardim et al. (2017).

[2]Recent studies have documented increased labor mobility induced by minimum wage changes, especially for low skilled workers (Monras, 2015; McKinnish, 2017).

[3]As of September 2017, 39 counties and cities have passed new minimum wage laws according to the UC Berkeley Labor Center. 23 out of 39 cities/counties have passed minimum wages of $15 or more, while the current federal minimum wage remains at $7.25. See http://laborcenter.berkeley.edu/minimum-wage-living-wage-resources/inventory-of-us-city-and-county-minimum-wage-ordinances/for more details.

[4]Ideally, type could be a summary statistic to rank workers expected productivity. I empirically use educational attainment as a proxy for worker types. Low type represents high school dropouts while high type represents high school graduates or more. According to 2015 the Current Population Survey (CPS),

county. Workers accept a local offer if its value exceeds the value of unemployment. When considering offers from neighboring regions, workers require extra compensation to offset migration/commuting costs. Firms decide in which counties to post vacancies, where the number of vacancies is determined by a free entry condition. Given the assumption of random search, heterogeneous workers in all locations are contacted by firms at identical rates. An individual's productivity when meeting a firm is determined by his/her type and an idiosyncratic random matching quality. The bargained wage is determined by a surplus division rule, subject to the minimum wage constraint, which left-truncates the original wage distribution (Flinn, 2006). The new wage structure is a continuous distribution with a mass point at the minimum wage level.

I estimate this spatial job search model using a Simulated Method of Moments estimator that combines county-level data moments from various sources. The migration and commuting flows are obtained from the American Community Survey (ACS). Local labor market conditions (hiring rates, separation rates and employment rates) are obtained from Quarterly Workforce Indicators (QWI) survey. The payroll share of firms' expenditures, and the ratio of job postings to workers come from the Economics Wide Key Statistics (EWKS) and the Conference Board Help Wanted Online (HWOL).

This model provides a framework to access the effects of minimum wage increases of a range of magnitude. Previous studies have focused on the most disadvantaged workers, without considering the welfare consequences for high type workers. To study the impacts of minimum wage increases for heterogeneous workers, my model incorporates four important effects. First, conditional on being employed, workers receive a higher wage from the same matches (the "*wage enhancement effect*"). Second, a minimum wage increase also causes a disemployment effect, because it dissolves marginally acceptable matches (the "*disemployment effect*"). Low type workers are more likely to be the marginally hired worker. Third, when firms are mandatory to pay workers more, they receive a smaller fraction of the sur-

5.8 percent workers are paid an hourly rate at or below federal minimum wage for the low type group, while this rate drops to 2.9 percent for the high type group.

plus from same matches (the "*share reduction effect*"). Fourth, the probability of filling the vacancy with a high productivity worker increases in the higher minimum wage county but decreases in its neighboring county (the "*worker relocation effect*"). The incentive for firms to post vacancies is reduced in both counties, but especially in the county that does not change its minimum wage, due to negative spillover effects.

My analysis yields three main results. First, local minimum wage hikes have contrasting impacts on differentiated workers, expanding the inequality between low type and high type workers. Low type workers are adversely affected by higher minimum wages, primarily due to greater *disemployment effect*. For high type workers, the *wage enhancement effect* dominates the *disemployment effect* when the minimum wage level is less than $14, above which the countervailing *disemployment effect* start to dominate. Therefore, the welfare of high type workers displays a hump shape with a peak at $14/hour. When simulating the welfare difference of a range of minimum wages, the inequality between high and low type workers grows as the local minimum wage increases and reaches its peak at $15.

Second, I use the estimated model to evaluate two policies: restricting labor mobility and preempting local minimum wage laws.[5] For a range of minimum wage values, I find that the welfare of high type workers is negatively impacted, but both policies have beneficial effects for low type workers. In the experiment of restricting spatial labor mobility, the low type workers in neighboring counties prefer two labor markets to be isolated when local minimum wage increases are large (above $10), because the cost of lost working opportunities is fully compensated by the benefit of eliminating spillover externalities. In the experiment of preempting local minimum wage laws, I compare local minimum wage changes with universal minimum wage changes. I find that low type workers prefer universal minimum wage hikes over local minimum wage hikes when the minimum wage change is moderate (below $14.5). The benefit of reducing spillover externalities outweighs the cost of a larger

[5]The minimum wage preemption laws prohibit cities from enacting their own minimum wage laws. As of July 6, 2017, 25 states have passed such laws. See http://www.nelp.org/publication/fighting-preemption-local-minimum-wage-laws/ for a more comprehensive policy review.

disemployment effect.

Third, I find the disemployment effect of a minimum wage increase is underestimated if one ignores labor mobility. On one hand, low educated workers tend to move away in response to a minimum wage increase and thus "disappear" from the "treated" county. On the other hand, they "reappear" in the neighboring area, contaminating the control group. I obtain with the model an estimate of the elasticity of employment with respect to the minimum wage equal to -0.073; ignoring labor mobility cuts this value in half to -0.034. The bias is most severe for counties with higher fractions of mobile workers.

My paper contributes to four broad strands of the literature. First, it is the first paper highlighting the negative spillover effects created by local minimum wage policies. There are a few recent papers documenting worker migration/commuting decisions are responsive to local minimum wage changes (Monras, 2015; McKinnish, 2017). However, this is the first paper linking labor flows with negative externalities for neighboring area workers. The insight that local policies may create externalities in the neighboring area through policy-induced migration is also discussed in the fiscal-federalism literature. For example, Serrato and Zidar (2016) studies the incidence of state corporate taxes on the welfare of workers, landowners and firm owners. In their model, a state tax cut reduces the tax liability and the cost of capital, attracting more establishments to move in. Cohen et al. (2011) studies the effects of marginal tax rates on migration decisions in the U.S., while Young and Varner (2011) and Moretti and Wilson (2017) focus on the geographic locations of top earners. Although policy-induced migration has already drawn significant attentions in the tax competition literature, my paper is the first application in the minimum wage context.

My paper also contributes to the structural minimum wage literature. It extends Flinn (2006) by allowing for location-specific minimum wages and spatial mobility. Previous minimum wage studies usually assume one universal minimum wage for the whole labor market. (Eckstein and Wolpin, 1990; Van den Berg and Ridder, 1998; Flinn, 2006; Mabli and Flinn, 2007; Eckstein et al., 2011; Flinn and Mullins, 2015; Flinn et al., 2017) By extending the framework to multiple connected sub-markets, my model is able to incorporate geographical minimum wage variation for identification and evaluate the externalities of local minimum wage laws. The spatial search framework in my paper is similar to that of Meghir et al. (2015), which develops an equilibrium wage-posting model with formal and informal sectors. Their paper focuses on firm heterogeneity while I focus on worker heterogeneity. Other relevant spatial equilibrium frameworks include Coen-Pirani (2010); Baum-Snow and Pavan (2012); Kennan and Walker (2011); Schmutz and Sidibe (2016). By embedding local minimum wage policy into a spatial equilibrium model, my model allows examination of the effects of minimum wages on labor mobility, local employment, migration, wages and welfare.

This paper also explores the methodological implications for minimum wage studies that use adjacent counties as the control group. Starting with Card and Krueger (1994), cross-border comparisons became a popular method of studying the employment effects of minimum wage increases. For example, Dube et al. (2007, 2010, 2016) generalize this strategy to all contiguous county pairs and find small disemployment effects, consistent with Card and Krueger (1994). Although the cross-border design is persuasive, because of the geographic proximity between the treatment and control areas, there are concerns about the assumption that adjacent counties are unaffected, particularly when the minimum wage discrepancy between counties is large. I find that ignoring labor mobility leads to an underestimation of disemployment effects for two reasons. First, the unemployed workers move out of the "treated" area when they can not find jobs, and second, they move into neighboring areas, contaminating the control group.

Lastly, this paper contributes to the recent local labor market policy literature, emphasizing the potential externalities caused by place-based policies.[6] I show that low type workers, who are the intended beneficiaries of minimum wage policies, are actually worse-off after minimum wage increases. The estimates of moving cost confirms that taking the neighboring

[6] See Glaeser et al. (2008) and Enrico (2011) for reviews. Other recent papers include Kline (2010); Busso et al. (2013); Kline and Moretti (2013)

job is costly in general, which is consistent with the finding in Manning and Petrongolo (2017). While their paper argues that the probability of a random distant (at least 5km away) job being preferred to random local (less than 5km away) job is only 19% based on data from UK. Using county-level U.S. data, I find a slightly higher probability of 22.2%.

The structure of the paper is as follows. The next section presents a spatial job search equilibrium model. Section 3 describes the multiple data sources I will use to estimate the model. Section 4 discusses the identification and estimation strategy. Section 5 present the estimation results. Section 6 discusses the counterfactual experiments. Section 7 concludes.

1.2. Model

I develop a dynamic spatial search model where individuals live and work in one of the paired counties (j, j'). A job seeker in one county may receive either a local offer or a neighboring offer at certain rates. When a worker meets a firm in county j, they bargain over the wage subject to the minimum wage policy in county j. Local minimum wage changes would potentially affect labor market conditions in the neighboring county due to labor mobility.

1.2.1. Framework

I consider a continuous time model, where infinitely lived, risk neutral workers maximize their expected utility (income) with discount rate ρ. The economy consists of two adjacent local markets, a pair of counties (j, j'). The economy has a fixed number of potential workers with different types a. $N(a, j)$ represents the number of workers with type a in county j. Type is discrete, taking n different values $a \in A = \{a_1, ..., a_n\}$.[7] The number of workers for each type is exogenous. However, their working and living status are determined by the endogenous job searching process. $U(a, j)$, $L(a, j)$, and (a, j) represent the number of unemployed workers, local workers, and mobile workers with type a in county j. I focus on job search and labor mobility behavior in the steady state.

[7]For computational tractability, I consider two types: high (a_h) and low (a_l) in the empirical analysis.

1.2.2. Worker's problem with wage w

A job seeker of type a in county j may receive wage offers from county j or j'. Upon meeting a firm, the productivity is given by

$$y = a\theta$$

where θ is the random matching quality, which is assumed to be an i.i.d. draw from the distribution function $G(\theta)$.[8] Given the job offers from the local county arrive at rate λ_j and the job offers from the neighboring county arrive at rate $\lambda_{j'}$, the value of unemployment can be written recursively as:

$$\begin{aligned} \rho V_u(a,j) &= \underbrace{ab_j}_{\text{(1) flow value}} + \lambda_j \underbrace{\int_{m_j}^{\infty} \{V_e(w,j) - V_u(a,j)\}^+ dF(w|a,j)}_{\text{(2) option value of accepting a local offer}} \\ &+ \lambda_{j'} \underbrace{\int_{m_{j'}}^{\infty} \{V_e(w,j') - c(a,j) - V_u(a,j)\}^+ dF(w|a,j'))}_{\text{(3) option value of accepting a neighbouring offer}} \end{aligned} \tag{1.1}$$

The notation $\{x\}^+ \equiv \max\{x,0\}$. m_j represents the minimum wage level in county j. In the continuous time setting, at most one job offer arrives in each moment. Equation 1.1 decomposes the value of unemployed workers into three components: (1) ab_j represents the flow utility of unemployment;[9] (2) the option value when a local offer with wage w is better than staying unemployed; (3) the option value when an offer with wage w from neighboring county, net of the the moving cost $c(a,j)$, is better than staying unemployed.

Following Baum-Snow and Pavan (2012) and Schmutz and Sidibe (2016), I use the moving cost $c(a,j)$ as the key function to distinguish the local labor market from the neighboring labor market. If $c(a,j) = 0$, the workers in county j and county j' will have exactly the

[8] The assumption that the flow productivity $y_{ij} = a_i\theta_j$ is the multiplicity of a firm type θ_j and a worker type a_i is standard in the literature (Postel-Vinay and Robin (2002); Cahuc et al. (2006)). Following this spirit, the distribution of matching productivity should be location-specific (firm-specific) $G_j(\theta)$. Since labor market conditions in county pairs should be similar, I assume the matching productivity $G_j(\theta)$ is the same for these two counties.

[9] The assumption that a worker's unemployment utility "at home" ab_j and productivity at work $a\theta$ are both proportional to type a is widely used in the literature. e.g. Postel-Vinay and Robin (2002); Flinn and Mullins (2015).

same working opportunities, which means paired counties are essentially one united labor market. If $c(a,j) = +\infty$, the paired counties are totally isolated markets. As pointed out by Schwartz (1973) and Greenwood (1975), this moving cost summarizes both the psychic cost of losing local social connections with family and friends and the physical transportation cost, which depends on the moving distance. The specifically parametric form of the moving cost will be shown in section 1.4.1.

I assume no on-the-job search. Therefore, the worker who accepts a job with wage w will never voluntarily quit the current job. Thus the existing matches only dissolve with a constant exogenous rate η_j. The value of employment, V_t^e, has the the following form[10]:

$$V_e(w,a,j) = \frac{w + \eta_j V_u(a,j)}{\rho + \eta_j} \tag{1.2}$$

1.2.3. Bargaining with a minimum wage constraint

In this section, I specify how the wage between the worker and the firm is determined. I first consider the case without a minimum wage. If a worker with type a meets a firm in location j and draws a matching quality θ in period t, the bargained wage is assumed to be derived from a Nash bargaining solution. The wage $\hat{w}(a,j,\theta)$ maximizes the weighted product of the worker's and firm's net return from the match. To form the match, the worker gives up the value of unemployment $V_u(a,j)$, and the firm gives up the unfilled homogeneous vacancy, which has zero value according to the free entry condition.[11]

$$\hat{w}(a,j,\theta) = \arg\max_{w} \left(V_e(w,a,j) - V_u(a,j)\right)^{\alpha_j} V_f(w,a,\theta,j)^{1-\alpha_j}$$

[10]The derivations of equations 1.1 and 1.2 are described in Appendix 1.8.1.

[11]I do not model different outside options for local workers and mobile workers for two reasons. First, it is unclear whether moving costs are a credible "threat point" for mobile workers because they have to pay the moving cost before they can work in the other county. Second, due to menu costs, it is not economic for firms to make a separate wage offers for mobile workers who are a minority of new hires.

where location specific bargaining weight α_j is strictly between 0 and 1, representing the relative bargaining strength of the labor side. V_f is the present value of the filled vacancy for the firm. As derived in Appendix 1.8.2, the bargained wage offer function is:

$$\hat{w}(a,j,\theta) = \rho V_u(a,j) + \alpha_j(a\theta - \rho V_u(a,j)) \tag{1.3}$$

The interpretation of this bargained wage is intuitive. The workers receive their reservation wage $\rho V_u(a,j)$ and a fraction of bargained share α_j of the net surplus of the current match, which is the total production $a\theta$ minus what workers give up $\rho V_u(a,j)$.

Following Flinn (2006), the introduction of a minimum wage in area j is treated as a "side constraint" to the original bargaining problem.

$$w(a,j,\theta) = \arg\max_{w \geq m_j} \left(V_e(w,a,j) - V_u(a,j)\right)^{\alpha_j} V_f(w,a,\theta,j)^{1-\alpha_j}$$

The minimum wage constraint $w \geq m_j$ is imposed by local policy maker and applies to all potential job matches. Before considering the case when the minimum wage binds, I solve for the critical value of matching quality where the worker receives exactly the minimum wage based on the original surplus decision rule (Equation 1.3).

$$\hat{\theta}(a,j) = \frac{m_j - (1-\alpha_j)\rho V_u(a,j)}{a\alpha_j}$$

If $\hat{\theta}(a,j) \leq \frac{m_j}{a}$, the minimum wage has no effect on the bargained wage because the reservation value is so high that all acceptable matches for workers actually give them wages equal or larger than m_j. (i.e. $a\theta^*(a,j) \geq m_j$). If $\hat{\theta}(a,j) > \frac{m_j}{a}$, the minimum wage is binding when $\theta \in [\frac{m_j}{a}, \hat{\theta})$. The firms in this scenario would pay workers m_j, which is more than the worker's "implicit" reservation wage $\hat{w}(a,j,\theta)$. Although payroll expenditure expands, it is still in firms' best interests to hire these workers, because destroying the jobs would reduce profits to zero. The binding minimum wage creates a wedge between the worker's wage and their "implicit" reservation wage, making the latter unobservable. Following Flinn (2006), I introduce the reservation matching quality $\theta^*(a,j)$, which is the lowest matching quality of a local match that a worker with type a will accept. In other words, the worker is indifferent between accepting a local job with matching quality $\theta^*(a,j)$ and staying unemployed.

$$V^e(\hat{w}(a,j,\theta^*(a,j)),a,j) = V^u(a,j)$$

$$\theta^*(a,j) = \frac{\rho V_u(a,j)}{a}$$

This reservation matching quality would be "implicit" in the case when the minimum wage binds ($m_j > \rho V_u(a,j)$).In this way, I obtain an affine mapping between the cumulative distribution of the matching quality, $G(\theta)$, and the cumulative wage distribution $F(w|a,j)$:

$$f_t(w|a,j) = \begin{cases} \frac{(a\alpha)^{-1}g(\tilde{\theta}(w,a,j))}{\tilde{G}(\frac{m_j}{a})} & w > m_j \\ \frac{\tilde{G}(\hat{\theta}(a,j))-\tilde{G}(\frac{m_j}{a})}{\tilde{G}(\frac{m_j}{a})} & w = m_j \\ 0 & w < m_j \end{cases} \tag{1.4}$$

where $f(w|a,j)$ is the probability distribution function(PDF) of $F(w|a,j)$, $g(\theta)$ is the PDF of $G(\theta)$, and $\tilde{G}(\theta) = 1 - G(\theta)$ is the complementary function of the cumulative distribution function $G(\theta)$. $\tilde{\theta}(w,a,j) = \frac{w-(1-\alpha_j)\rho V_u(a,j)}{a\alpha_j}$ denotes the matching quality whose bargained wage is equal to w. The observed wage distribution consists of a point m_j with mass $\frac{G(\hat{\theta}(a,j))-\tilde{G}(\frac{m_j}{a})}{\tilde{G}(\frac{m_j}{a})}$ and a continuous function (assuming $G(\theta)$ is continuous) when $\theta > \hat{\theta}$. Thus the bargained wage can be summarized as:

$$w(a,j,\theta) = \max\{m_j, \alpha_j a\theta + (1-\alpha_j)\rho V_u(a,j)\} \tag{1.5}$$

It is worth to point out that a binding minimum wage affects the wages of all workers, but through different channels. For the workers with matching quality $\theta \in [\frac{m_j}{a}, \hat{\theta}(a,j))$, the minimum wage directly benefits them by boosting their wage to m_j. For workers with even higher matching quality $\theta \in [\hat{\theta}(a,j), \infty)$, the minimum wage changes their value of unemployment $\rho V_u(a, j)$.[12] To summarize, introducing the minimum wage as a side restriction on Nash-bargained wages converts a continuous underlying productivity distribution into a mixed continuous-discrete accepted wage distribution, with a mass point at the minimum wage.

1.2.4. Migration/commuting trade-off

Next, I characterize the spatial strategies of the workers. To capture the different types of labor mobility observed in the data, I distinguish commuting from migrating by specifying a choice-specific moving cost $cc_h(a, j), h = \{0, 1\}$. The timing is as follows: (1) an offer from neighboring area j' arrives at rate $\lambda_{j'}$. (2) After the matching quality θ is realized, the worker decides to accept/reject the offer based on the trade-off between the wage offer $w(a, j', \theta)$ net of the ex-ante moving cost $c(a, j)$ and the value of unemployment, $V_u(a, j)$. (3) If the worker accepts the offer, the preference shock ε_h is realized and the worker chooses whether to commute or migrate.

A worker with type a continues to receive job offers from the neighboring county at rate $\lambda_{j'}$. The expected moving cost $c(a, j)$, is a function of the worker's type and location-specific characteristics. Following Schmutz and Sidibe (2016), I introduce a *"implicit" mobility compatible indifferent matching quality* $\theta^{**}(a, j)$, fulfilling the following condition:

$$V_u(a, j) + c(a, j) = V_e(\theta^{**}(a, j), a, j')$$

where j represents the worker's place of residence and thus j' will be the worker's place of work. The worker will accept the neighboring offer if and only if the matching quality of the offer exceeds the mobility compatible threshold $\theta \geq \theta^{**}(a, j)$. This match will also be sustainable for firms as long as $\theta \geq \frac{m_{j'}}{a}$. To summarize, the worker whose residence is in county j will accept a neighboring offer if and only if $\theta \geq \max\{\frac{m_{j'}}{a}, \theta^{**}(a, j)\}$.

[12] However, the sign of this change is ambiguous, depending on the trade-off between the increase of expected income and the reduction of expected working opportunities.

After accepting the neighboring offer, workers have two alternatives. They can either work as migrants ($h = 1$), pay a lump-sum cost $cc_{h=1}(a, j)$, and become a native worker in county j' or work as commuters ($h = 0$) and pay a recurring commuting cost $cc_{h=0}(a, j)$. I use $cc_h(a, j)$ to represent the lump-sum equivalent cost. The choice-specific moving cost $cc_h(a, j)$ is a function of both the worker's type and physical distance between counties, as well as the cost differences of house renting between paired counties. Its exact parametric form will be discussed in Section 1.4.1.

In additional to the moving cost $cc_h(a, j)$, workers also receive an unobserved preference shock ε_h. The workers thus choose their lowest cost mobility option, $h(a, j)$:

$$h(a,j) = \begin{cases} 0 & \text{if } \varepsilon_{a0} - cc_0(a,j) > \varepsilon_{a1} - cc_1(a,j) \\ 1 & \text{if } \varepsilon_{a0} - cc_0(a,j) \leq \varepsilon_{a1} - cc_1(a,j) \end{cases}$$

Assuming the preference shock ε_{ah} follows an i.i.d. type I extreme value distribution with a location parameter 0 and a common scale parameter σ_a, then the ex-ante expected cost has the following analytic formula (Rust 1987):

$$\begin{aligned} c(a,j) &= \max\{\varepsilon_{a0} - cc_0(a,j), \varepsilon_{a1} - cc_1(a,j)\} \\ &= \sigma_h \log(\textstyle\sum_{h=0}^{1} \exp(-cc_h(a,j))/\sigma_a) + \sigma_a \gamma \end{aligned}$$

The probability of each choice is specified as:

$$P_h(a,j) = \frac{exp(-cc_h(a,j)/\sigma_a)}{exp(-cc_0(a,j)/\sigma_a) + exp(-cc_1(a,j)/\sigma_a)} \tag{1.6}$$

1.2.5. Worker's optimal strategies

The worker's optimal strategies consists of the local job taking strategies and sequential strategies for the neighboring job offers. The local decision is fully described by the implicit reservation matching quality $\theta^*(a, j)$, while the moving decisions are summarized by both the implicit mobility compatible matching quality $\theta^{**}(a,j)$ and migration/commuting choice probability $P_h(a,j)$.

Proposition 1. *OPTIMAL STRATEGIES*

For unemployed workers of type a in county j, the optimal strategy is:

- *accept any local job with matching quality higher than* $\max\{\theta^*(a,j), \frac{m_j}{a}\}$
- *accept any neighboring job with matching quality higher than* $\max\{\theta^{**}(a,j), \frac{m_{j'}}{a}\}$
  - *with probability $P_1(a,j)$, the workers choose to commute*
  - *with probability $P_0(a,j)$, the workers choose to migrate*

In the last part of this section, I describe the fixed point equation system that is used to solve for $\theta^*(a,j)$ and $\theta^{**}(a,j)$. By applying both the reservation matching quality $\theta^*(a,j)$ and mobility compatible matching quality $\theta^{**}(a,j)$ to Equation 1.1, I get the following system of equations:[13]

$$\begin{aligned}
a\theta^*(a,j) \quad = \quad & \underbrace{ab_j}_{\text{(1) Flow utility}} \quad + \frac{\lambda_j}{\rho+\eta_j}[\underbrace{\mathbf{I}\left(\theta^*(a,j) < \frac{m_j}{a}\right)(m_j - a\theta^*(a,j))\left(\tilde{G}(\hat{\theta}(a,j)) - \tilde{G}(\frac{m_j}{a})\right)}_{\text{(2) Local offer with wage } m_j} \\
+ \quad & \underbrace{\int_{\max\{\hat{\theta}(a,j),\theta^*(a,j)\}} a\alpha_j(\theta - \theta^*(a,j))dG(\theta)]}_{\text{(3) Local offer with wage } w_j > m_j} \\
+ \quad & \frac{\lambda_{j'}}{\rho+\eta_{j'}}[\underbrace{\mathbf{I}\left(\theta^{**}(a,j) < \frac{m_{j'}}{a}\right)(m_{j'} - a\theta^*(a,j'))\left(\tilde{G}(\theta^{**}(a,j)) - \tilde{G}(\frac{m_{j'}}{a})\right)}_{\text{(4) Neighbouring offer with wage } m_{j'}} \\
+ \quad & \underbrace{\int_{\max\{\hat{\theta}(a,j'),\theta^{**}(a,j)\}} a\alpha_j(\theta - \theta^*(a,j'))dG(\theta)}_{\text{(5) Neighbouring offer with wage } w_{j'} > m_{j'}} \\
+ \quad & \underbrace{(\rho+\eta_{j'})\left(\frac{a(\theta^*(a,j) - \theta^*(a,j'))}{\rho} + c(a,j)\right)\tilde{G}(\theta^{**}(a,j))]}_{\text{(6) The unemployed value difference between staying/moving}}
\end{aligned} \tag{1.7}$$

[13]The derivation of equation 1.7 can be found in Appendix 1.8.3

with

$$\hat{\theta}(a,j) = \tfrac{m_j-(1-\alpha_j)a\theta^*(a,j)}{a\alpha_j}$$
$$\hat{\theta}(a,j') = \tfrac{m_{j'}-(1-\alpha_{j'})a\theta^*(a,j')}{a\alpha_{j'}}$$
$$\theta^{**} : V_u(a,j) + c(a,j) = V_e(\theta^{**}(a,j), a, j')$$

In equation 1.7, the value of the implicit matching quality $a\theta^*(a,j)$ consists of six components: (1) the instant flow utility ab when unemployed; (2) the expected value associated with a local offer with binding minimum wage m_j; (3) the expected value associated with a local offer with wage $w_j > m_j$; (4) the expected value associated with an acceptable neighboring offer with binding minimum wage $m_{j'}$; (5) the expected value associated with an acceptable neighboring offer with wage $w_{j'} > m_{j'}$; (6) the unemployed utility difference between staying and moving, which includes both the moving cost $c(a,j)$ and the change of the option value of being unemployed $a\theta^*(a,j) - a\theta^*(a,j')$.

The intuition of equation 1.7 is straightforward. The value difference between accepting the lowest acceptable job and remaining unemployed $a\theta^*(a,j) - ab_j$ reflects an opportunity cost, which is perfectly compensated by the expected premium of finding a better job in the future. This job could either be a local one or a neighboring one after paying the moving cost $c(a,j)$.

1.2.6. Endogenous contact rate

In this section I consider how the contact rates $\lambda_j, j = 1,2$, are determined in general equilibrium. I assume that firms randomly encounter workers with the same probability. This assumption captures the idea that workers applying for the same position may have different productivity but are easily to substitute with each other. I adapt the Mortensen and Pissarides (1994) framework and allow firms to post vacancies K_j in county j with constant marginal cost ψ_j which is open to all workers in both counties. The matching technology is assumed to be constant returns to scale. Let $N = \sum_{a\in A}(U(a,j) + U(a,j'))$ be the number of all job seekers in the economy, where $U(a,j)$ is the number of unemployed workers with type a in county j. If the firms in county j creates K_j vacancies, then the total number of potential matches created in county j, M_j, is given by

$$M_j = N^{\omega_j} K_j^{1-\omega_j}$$

where ω_j is the matching elasticity parameter in market j.

I use a Cobb-Douglas matching function with constant return to scale and total factor productivity equal to 1. It then only requires one parameter ω_j to characterize the heterogeneity of matching functions in each local labor market j.

The contact rate per job in county j, $q_j(k_j)$, can be represented as:

$$q_j(k_j) = k_j^{\omega_j}$$

where $k_j = \frac{N}{K_j}$ captures the market tightness. The correlation between market tightness and job arrival probability λ_j is

$$\lambda_j = k_j(K_j, N)^{\omega_j - 1} \tag{1.8}$$

It is important to emphasize that although workers in both counties have the exact same opportunities to meet with the same firm, their willingness to accept the same job is different due to moving costs. For workers living in the neighboring county, they are more picky about neighboring jobs because the job premium has to compensate for the additional moving cost. The total number of matches created by the firms in county j is:

$$\text{Total Hires} = \frac{M_j}{N} \sum_{a \in A} \left(\underbrace{U(a,j) G\left(\max\{\theta^*(a,j), \frac{m_j}{a}\}\right)}_{\text{Local Hires}} + \underbrace{U(a,j') G\left(\max\{\theta^{**}(a,j'), \frac{m_j}{a}\}\right)}_{\text{Neighboring Hires}} \right)$$

The firm's value of a match can be represented as:

$$V_f(\theta, a, j) = \frac{a\theta - w(a,\theta,j)}{\rho + \eta_j} \tag{1.9}$$

The expected value of creating a vacancy for firms V_v in county j is:

$$V_v = -\psi_j + \frac{k_j(K_j, N)^{\omega_i}}{N} \sum_{a \in A} \left[\underbrace{U(a, j) \int_{\max\{\theta^*(a,j), \frac{m_j}{a}\}} V_f(\theta, a, j) dG(\theta)}_{\text{Profit from local workers}} + \underbrace{U(a, j') \int_{\max\{\theta^{**}(a,j'), \frac{m_j}{a}\}} V_f(\theta, a, j) dG(\theta)}_{\text{Profit from neighboring workers}} \right]$$

Assuming each county has a population of potential entrants with an outside option equal to 0, firms will continue to create vacancies until the expected profit is equal to 0, $V_v = 0$. Under the free entry condition (FEC), the endogenous contact rate is determinate by the following equation

$$\psi_j = \frac{k_j(K_j, N)^{\omega_i}}{N} \sum_{a \in A} \left[U(a, j) \int_{\max\{\theta^*(a,j), \frac{m_j}{a}\}} V_f(\theta, a, j) dG(\theta) + U(a, j') \int_{\max\{\theta^{**}(a,j'), \frac{m_j}{a}\}} V_f(\theta, a, j) dG(\theta) \right] \tag{1.10}$$

1.2.7. Definition of a steady-state spatial equilibrium

Let $\theta \in \mathbf{R}_+$, $a \in \mathbf{A} = \{a_1, a_2, ..., a_n\}$, $j \in \mathbf{J} = \{1, 2\}$, and let $\mathbf{S_1} = \mathbf{R}_+ \times \mathbf{A} \times \mathbf{J}$ and $\mathbf{S_2} = \mathbf{A} \times \mathbf{J}$. Let $\mathbf{B}(\mathbf{R}_+)$ be the Borel $\sigma-$algebra of $\mathbf{R}_+$ and $\mathbf{P}(\mathbf{A})$, $\mathbf{P}(\mathbf{J})$ the power sets of $\mathbf{A}$ and $\mathbf{J}$, respectively. Let $\aleph = \mathbf{B}(\mathbf{R}_+) \times \mathbf{P}(\mathbf{A}) \times \mathbf{P}(\mathbf{J})$, and $\mathbf{M}$ be the set of all finite measures over the measurable space $(\mathbf{S_1}, \aleph)$

Definition 1.2.1. *A steady-state spatial equilibrium is a set of individual functions for workers $V_u : \mathbf{S_1} \to \mathbf{R}_+$ and $V_e, \theta^*, \theta^{**}, P_h : \mathbf{S_2} \to \mathbf{R}_+$, a set of the functions for firms $V_f : \mathbf{S_1} \to \mathbf{R}_+$ and $\{K_j\}_{j=1,2}$, a set of contact rates $\{\lambda_j\}_{j=1,2}$ and wage rates $w : \mathbf{S_1} \to \mathbf{R}_+$ and a set of aggregate measures of different working status $U, L, M : \mathbf{S_2} \to \mathbf{R}_+$, the following conditions hold:*

1. Worker's problem: given the contact rate, wage and initial condition, V_u and V_e are the solutions of Eqs. 1.1 and 1.2, respectively. The optimal strategies θ^*, θ^{**} are described in Proposition 1 and $\{P_h\}_{h=0,1}$ are described in Eq. 1.6. The functions $\{V_u, V_e, \theta^*, \theta^{**}, P_h\}$ are measurable with respect to $\aleph$.

2. Firm's problem: given the contact rate, wage and initial condition, V_f is solved by Eq. 1.9 and K_j is solved by Eq. 1.10.

3. The bargained wage: the bargained wage with a minimum wage constraint is defined by Eqs. 1.4 and 1.5.

4. Endogenous contact rate (labor market clear): the contact rate λ_j is solved by Eq. 1.8.

5. The aggregate measures of working status keep constant

$$
\begin{aligned}
\underbrace{\lambda_j \left(U(a,j)\tilde{G}(\max\{\theta^*(a,j), \frac{m_j}{a}\}) + U(a,j')P_0(a,j')\tilde{G}(\max\{\theta^{**}(a,j'), \frac{m_j}{a}\}) \right)}_{\text{Inflow to L}} &= \underbrace{L(a,j)\eta_j}_{\text{Outflow from L}} \\
\underbrace{U(a,j) \left(\lambda_j \tilde{G}(\max\{\theta^*(a,j), \frac{m_j}{a}\}) + \lambda_{j'} \tilde{G}(\max\{\theta^{**}(a,j), \frac{m_{j'}}{a}\}) \right)}_{\text{Outflow from U}} &= \underbrace{L(a,j)\eta_j + M(a,j)\eta_{j'}}_{\text{Inflow into U}} \\
\underbrace{\lambda_j U(a,j')P_1(a,j')\tilde{G}(max\{\theta^{**}(a,j'), \frac{m_j}{a}\})}_{\text{Inflow to M}} &= \underbrace{M(a,j)\eta_j}_{\text{Outflow from M}}
\end{aligned}
$$

1.3. Data and descriptive statistics

This paper primarily uses two data sets: the Quarterly Workforce Indicators (QWI) for local labor market information and the American Community Survey (ACS) for labor mobility information. QWI provides the number of job stocks and flows, and average earnings by industry, worker demographics, employer age, and size. The QWI comes from the Longitudinal Employer-Household Dynamics (LEHD) linked employer-employee micro data, which are collected through a unique federal-state sharing collaboration between the U.S. Census Bureau and state labor market agencies.[14] Compared to the CPS and JOLTS, the QWI has near-universal worker-employer paired information, covering 96% of all private-sector jobs. Second, QWI provides worker-side demographic information such as age, sex, race/ethnicity, and education.[15] This feature allows me to analyze the demographics of a particular industry or specific local market.[16] Lastly, QWI has labor flow information,

[14]Data for Massachusetts, Puerto Rico, and the US Virgin Islands are still under development.

[15]Workers are identified by their Social Security number and linked with a variety of sources, including the 2000 Census, Social Security Administrative records, and individual tax returns to get their demographic information.

[16]While CPS contains similar information based-on household surveys, it generates small sample sizes when analyzing individual industries or areas.

including hires, separations, and turnovers, which are important because the direct impacts of minimum wage hikes are on job turnovers rather than employment stocks.[17] I focus on 2005-2015 primarily because the states missing from QWI before 2005 are not random - smaller states are under-represented. By 2005, all states except Massachusetts have joined the QWI program.[18]

In addition to QWI data, I also use the ACS from 2005-2015 to identify the commuting and migration flows between different jurisdictions.[19] Commuters are defined as people whose place of work is different from their place of residence, while migrants are defined as those who have changed their place of residence in the past year, according to the ACS. The basic geographic units in the ACS are "Public Use Micro Areas"(PUMAs) which are special non-overlapping areas that partition each state into contiguous geographic units containing between 100,000 to 300,000 residents. There were a total of 2,071 PUMAs in the 2000 census.

1.3.1. Contiguous border county pairs and their associated geographic minimum wage variations

Following the contiguous county-pair design proposed by Dube et al. (2010, 2016), I divide all counties in the U.S. into two sub-samples: counties that border another state (border counties), and counties that do not (interior counties). Out of 3,124 counties, 1,139 counties are border counties and I construct 1,181 unique pairs.[20] Figure 1 shows the locations of all counties along with their associated minimum wage policies. Between 2005 and 2015, there were 332 minimum wage adjustments (see 13 for details of minimum wage policies). While 78 changes are driven by the federal minimum wage law, the Fair Minimum Wage Act of 2007,[21] the other 164 events were due to state ordinances. Two observations are highlighted

[17] See Dube et al. (2010, 2016) for detailed discussions.

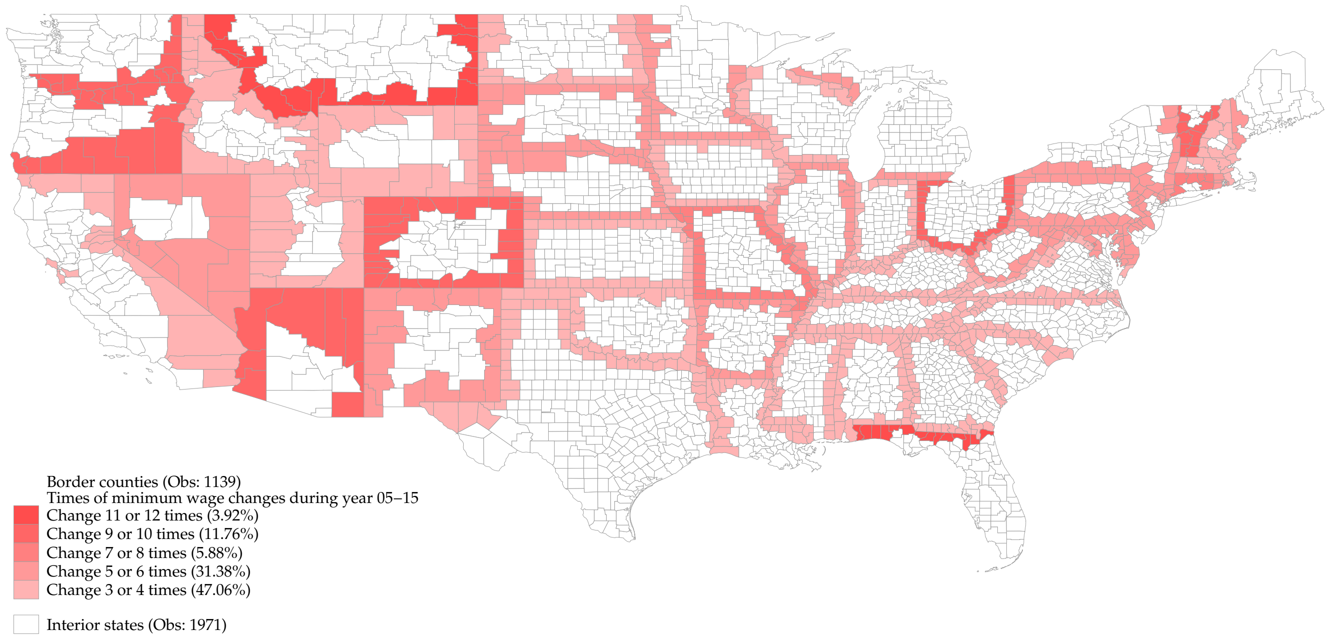
[18] Massachusetts does not join the QWI until 2010.

[19] I combine the 2005-2007, 2008-2010, and 2011-2015 ACS.

[20] Counties may border more than one county in the adjacent state, resulting in more pairs than border counties.

[21] The Act raised the federal minimum wage in three stages: to \$5.85 60 days after enactment (2007-07-24), to \$6.55 one year after that (2008-07-24), then finally to \$7.25 one year after that (2009-07-24).

Figure 1: Frequency of Minimum Wage Adjustments for Border Counties (2005-2015)



on the map. First, border counties frequently adjust their minimum wages. Between 2005 and 2015, all counties (except for those in Iowa) changed their local minimum wage at least three times, which gives me adequate variation to identify the effects of minimum wage hikes. Second, western counties are larger than other counties. Thus, the workers in those counties may suffer higher moving costs when working in a neighboring county.

In a given year, about half of the county pairs have different minimum wages. These differences average about 10%, but there is substantial heterogeneity across years (see Table 1). Overall, the substantial variation between county minimum wages is useful for identifying the effect of minimum wage hikes.

1.3.2. Migration and commuting flows

I use the American Community Survey (ACS) Public Use Microdata Sample (PUMS) data between 2005-2015 to distinguish commuters and migrants. Each respondent provides information about their place of residence one year ago, their current residence, and their current working address. To perform policy analysis, I convert PUMAs into pseudo-counties using

Table 1: Differences in County Pair Minimum Wages (2005-2015)

Year	Share of pairs with minimum wages differential	Percent difference in minimum wages
2005	27.6%	18.6%
2006	33.6%	19.1%
2007	66.0%	15.6%
2008	63.7%	11.1%
2009	52.2%	8.7%
2010	31.8%	5.8%
2011	36.2%	6.0%
2012	37.8%	7.7%
2013	44.1%	7.4%
2014	49.0%	8.6%
2015	68.5%	9.4%
Average	46.4%	10.7%

the Michigan Population Studies Center PUMA-to-County crosswalk.[22]

To construct a sample of workers most sensitive to minimum wage changes, I restrict my sample to individuals between 16 and 30 that live in the continental U.S. and are not currently in the military. I divide this sample into two groups based on education: the low educated group (high school dropouts group) and the high educated group (the high school graduates and above). These restrictions are commonly used in the literature because young people and least-educated people are more likely to be minimum wage workers (Deere et al. (1995); Burkhauser et al. (2000); Neumark (2001)). If the minimum wage effect is not significant for this group, then it is unlikely to be significant for other groups.

Local governments prioritize their residents over residents of neighboring counties and as a result, I carefully distinguish between migrants (who have moved out of a county) and commuters (who might work in neighboring counties). Descriptive statistics for both commuting outflows to other states and migration inflows from other states are provided in

[22] I do this for two reasons. First, since PUMAs are population-based, they are not natural jurisdictions for local policy analysis. In urban areas, a single county may contain multiple PUMAs. For example, Los Angeles County, California is comprised of 35 PUMAs. Likewise, a PUMA will consist of several counties in less population areas. Second, I want to match the ACS to county-based statistics from the QWI. See Appendix 1.10.2 and http://www.psc.isr.umich.edu/dis/census/Features/puma2cnty/for details.

Table 2: Summary Statistics of Migrants and Commuters (2005-2015)

		Interior counties		Border counties		Difference	
		Count	Rate	Count	Rate	Count	Rate
ALL workers							
Migrants	Mean	231	0.040	266	0.051	35.0	0.011
	S.D.	749	0.038	829	0.047	(8.96)	(0.0005)
Commuters	Mean	44.9	0.019	210	0.066	165	0.047
	S.D.	138	0.078	718	0.127	(6.48)	(0.0013)
Low educated group							
Migrants	Mean	28.5	0.024	31.5	0.030	3.00	0.006
	S.D.	95.4	0.033	87.2	0.040	(1.01)	(0.0004)
Commuters	Mean	4.49	0.021	20.1	0.047	15.6	0.026
	S.D.	20.2	0.090	74.7	0.118	(0.681)	(0.0012)
High educated group							
Migrants	Mean	203	0.045	235	0.058	22.0	0.013
	S.D.	674	0.043	770	0.053	(8.25)	(0.0005)
Commuters	Mean	40.4	0.031	189	0.070	149	0.039
	S.D.	125	0.097	656	0.130	(5.92)	(0.0013)
Observation		22,033		12,518			

*Data Source: ACS. Note: All statistics are reported at the county level. The count of migrations reports the number of individuals in each county whose place of residence last year differs from the place this year. The rate (a value between 0 and 1) is the percent of migrants in the local population. The count of commuters is the total number of workers whose state of work differs from the state of current residence. The rate (a value between 0 and 1) represents the percent of these commuters among the people who are currently in the labor force. Difference is border minus interior. * for 10%. ** for 5%, and *** for 1%.*

Table 2.[23] Migrants are defined as individuals whose county of residence last year differs from their current county of residence. Commuters are defined as workers whose state of work differs from their state of residence. The rate (a value between 0 and 1) represents the share of commuters in the labor force. All statistics are on county-level and are grouped by whether they are border or interior counties. Border counties have higher migration and commuting rates, likely because commuting and moving costs are lower (See Table 2)

I further estimates some regression models to explore how migration flows and commuting flows respond to the local minimum wage hikes. The results suggest that low educated workers tend to move away from rather than move towards counties with minimum wage increases, either by commuting or migration. In contrast, the high educated workers, who are served as the control group, are less responsive to the minimum wage changes. And these mobility patterns are robust to the following sensitivity analysis: (1) use alternative migration flows based on addresses on the income tax returns provided by the Internal Revenue Service (IRS); (2) using only the minimum wage changes caused by federal minimum wage laws; (3) restricting to county pairs whose centriods are within 75 kilometers. The detailed regression results are reported in Table 11 in Appendix 1.9.1.

1.3.3. Local labor market outcomes

From the QWI, I extract four quarterly variables: average monthly earnings, employment, hire rates, and separation rates. To make the QWI sample comparable to the ACS sample, I restrict worker's age to be between 19-34.[24] Labor force participation is extracted independently from the ACS. Overall, border and interior counties are similar across labor market statistics (Table 3).

[23] The other two potential measures of labor mobility patterns are commuting inflows and migrating outflows. They are in principle able to be calculated by summarizing all workers who migrate from/commute into the targeted PUMA in the sample. However, this calculation suffers from serious measurement error because the migrants from the particular PUMA and the commuters working in the particular PUMA are a small minority in other PUMAs and thus unlikely to be sampled.

[24] The division of age groups in QWI are 19-21, 22-24, 25-34, 35-44, 45-54, 55-64, and 65-99. To match with the selected ACS sample whose ages are between 16-30, I combine the first four age spans 14-18, 19-21, 22-24, and 25-34.

Table 3: County-Level Labor Market Summary Statistics (2005-2015)

	Interior counties		Border counties	
	Mean	SD	Mean	SD
Monthly earnings	1932	739	1930	739
Employment	14883	54878	13045	45968
Separation rates	0.299	0.111	0.301	0.103
Hire rates	0.326	0.171	0.326	0.128
Labor force participation rate				
All	0.618	0.199	0.623	0.197
High educated	0.701	0.222	0.704	0.219
Low educated	0.394	0.161	0.399	0.162

Note: All statistics are quarterly and from Quarterly Workforce Indicators except labor force participation, which is from the American Community Survey. Monthly earnings are in nominal dollars.

In Appendix 1.9.1, I estimates a regression models following Dube et al. (2007) and Dube et al. (2016) to examine the magnitude of disemployment in response to minimum wage increases. When using a common time fixed effect in column (1) in Table 11, the estimated disemployment elasticity is -0.068. However, this disemployment effect shrinks to -0.039 in column (2) when I replace the common time fixed effect with a pair-specific time fixed effect as the control. I attribute this change to the existence of spatial spillover effect. After the local county increases its own minimum wage, unemployed workers may seek their jobs in the neighboring county (either by migration or by commuting), which causes disemployment in the neighboring county. As a result, this spillover effect generates a common trend between counties in one pair. When this pair-specific co-movement is controlled by the pair-specific time effect, the estimates of local disemployment effect become less substantial.

1.4. Estimation strategy

1.4.1. Parametrization

To estimate the model, I need to make parametric assumptions for the types and moving costs. To be consistent with the data, I assume workers are of two types: a_h and a_l. High type workers are workers with high school diplomas and above while low type workers are high school dropouts. The proportion of these two types of workers are p_h and p_l.

I assume moving costs depend on a linear combination of worker's type a, the physical distance $d_{jj'}$ as well as the amenity difference $\gamma_j - \gamma_{j'}$ between the two counties.

$$cc_h(a,j) = \begin{cases} \beta_{0j} + \beta_{0d} d_{jj'} + \beta_{0a} I(a = a_h) + \beta_{0\gamma}(\gamma_j - \gamma_{j'}) & \text{if } h = 0 \\ \beta_{1j} + \beta_{1d} d_{jj'} + \beta_{1a} I(a = a_h) + \beta_{1\gamma}(\gamma_j - \gamma_{j'}) & \text{if } h = 1 \end{cases} \tag{1.11}$$

Equation 1.11 follows the standard gravity equation for migration. β_{hj} measures the relative openness of labor market j, which is county-specific and differs by the mobility choice h. The different impacts of distance on migrants and commuters are captured by β_{0d} and β_{1d}.[25] I also assume the moving costs to be differ by types a. The coefficientsβ_{0a} and β_{1a} represent the additional costs paid by high type workers. Lastly, I attribute the asymmetry between the cost $cc_h(a,j)$ and the cost $cc_h(a,j')$ to different housing rental prices γ_j and $\gamma_{j'}$, which are proxies of local living cost.

Assuming parametric distribution for matching quality is necessary for identification purposes. As Flinn and Heckman (1982), only a certain class of distributions satisfies the "recovery condition" necessary for identification. Following Flinn (2006) and Flinn and Mullins (2015), I assume the matching quality distribution $G(\theta)$ follows a log-normal distribution. Given the above assumptions, the economy is characterized by the vector S which combines a set of general parameters and a set of county-specific parameters.

$$\begin{aligned} S &= \{\rho, \mu_G, \sigma_G, a_h, a_l, \beta_{0d}, \beta_{0a}, \beta_{0r}, \beta_{1d}, \beta_{1a}, \beta_{1\gamma}, \sigma_0, \sigma_1\} && \text{General} \\ &\bigcup \{m_j(n), b_j(n), \eta_j(n), \psi_j(n), \alpha_j(n), \omega_j(n), \gamma_j(n), d_{jj'}(n), \beta_{0j}(n), \beta_{1j}(n), p_h(n), p_l(n)\}_{(j,n)\in\{1,2\}\times N} && \text{County} \end{aligned}$$

The county-pair specific parameters are unique for every $n \in N$, while the general parameters are shared by all counties. Although the general parameters simplify the estimation, the model remains computationally demanding if the county-pair specific parameters are recovered non-parametrically. For tractability, I impose random coefficient assumptions the

[25]While the distance between centroids is only a proxy for the real commuting time between two counties, some evidence shows the correlation between these two measures is quite high (Phibbs and Luft (1995);Boscoe et al. (2012)).

unobserved county-specific variables $s_j(n) \in \{b_j(n), \psi_j(n), \beta_{0j}(n), \beta_{1j}(n)\}$.[26] Given the close connection between the paired counties, I draw $s_1(n)$ and $s_2(n)$ from a multivariate normal distribution modeled for each $s_j(n) \in \{b_j(n), \psi_j(n), \beta_{0j}(n), \beta_{1j}(n)\}$:

$$\begin{pmatrix} x_{s1} \\ x_{s2} \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_s \\ \mu_s \end{bmatrix}, \begin{bmatrix} \sigma_{s1}^2 & \rho_s \sigma_{s1} \sigma_{s2} \\ \rho_s \sigma_{s1} \sigma_{s2} & \sigma_{s2}^2 \end{bmatrix} \right)$$

where the correlation ρ_s captures the similarity between these two counties. The random variables $s_j(n), j = 1, 2$ are the mapping from the $n - th$ draw of the following one-to-one mapping F (which is 6×1),

$$\begin{aligned}
\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} &\sim N\left(\begin{bmatrix} \mu_b \\ \mu_b \end{bmatrix}, \begin{bmatrix} \sigma_b^2 & \rho_b \sigma_b^2 \\ \rho_b \sigma_b^2 & \sigma_b^2 \end{bmatrix} \right) \\
\begin{pmatrix} \log \psi_1 \\ \log \psi_2 \end{pmatrix} &\sim N\left(\begin{bmatrix} \mu_\psi \\ \mu_\psi \end{bmatrix}, \begin{bmatrix} \sigma_\psi^2 & \rho_\psi \sigma_\psi^2 \\ \rho_\psi \sigma_\psi^2 & \sigma_\psi^2 \end{bmatrix} \right) \\
\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} &\sim N\left(\begin{bmatrix} \mu_{\beta 0} \\ \mu_{\beta 1} \end{bmatrix}, \begin{bmatrix} \sigma_{\beta 0}^2 & \rho_\beta \sigma_{\beta 0} \sigma_{\beta 1} \\ \rho_\beta \sigma_{\beta 0} \sigma_{\beta 1} & \sigma_{\beta 1}^2 \end{bmatrix} \right)
\end{aligned}$$

Thus, the joint distributions of these six variables are fully characterized by 11 parameters: 4 means, μ_s; 4 variances, σ_s^2; and 3 correlations, ρ_s. These parameters $(\mu_s, \sigma_s, \rho_s : s \in \{b, \psi, \beta_0, \beta_1\})$, in addition to those general parameters $\{\rho, \mu_G, \sigma_G, a_h, a_l, \beta_{0d}, \beta_{0a}, \beta_{0r}, \beta_{1d}, \beta_{1a}, \beta_{1\gamma}, \sigma_0, \sigma_1\}$, constitute the primitive parameters Ω of the model.

1.4.2. The method of simulated moments

My model is estimated by the method of simulated moments (MSM). When combining moments from multiple databases, MSM is a more natural estimation approach than maximum likelihood estimation (MLE).

Given Ω, I draw the unobserved variables $\left\{ b_j^r, \psi_j^r, \beta_{0j}^r, \beta_{1j}^r \right\}_{j=1,2}$ R times from the distri-

[26] The other county-specific parameters $\{m_j(n), \alpha_j(n), \gamma_j(n), \eta_j(n), d_{jj'}(n), p_h(n), p_l(n)\}_{j=1,2}$ are directly observed in the data.

butions of F for each county pair n. Combined with other observed county-level variables $\{m_j(n), \alpha_j(n), \gamma_j(n), \eta_j(n), d_{jj'}(n), p_h(n), p_l(n)\}_{j=1,2}$ and general parameters $\{\rho, \mu_G, \sigma_G, a_h, a_l, \beta_{0d}, \beta_{0a}, \beta_{0r}, \beta_{1d}, \beta_{1a}, \beta_{1\gamma}, \sigma_0, \sigma_1\}$, I then compute the vector of moments $\tilde{M}_{N,R}(\Omega)$ from the simulation. Model parameters are estimated by minimizing the weighted difference between those simulated moments $\tilde{M}_{N,R}(\Omega)$ and the actual data moments M_N, using the following quadratic distance function

$$\hat{\Omega}_{N,R,W} = \arg\min_{\Omega} \left((M_N - \tilde{M}_{N,R}(\Omega))' W_N (M_N - \tilde{M}_{N,R}(\Omega)) \right)$$

where M_N denotes the data moments for all county pairs in the data set, and $\tilde{M}_{N,R}(\Omega)$ represents the simulated moment evaluated at Ω based on R simulations of N county pairs. W_N is a symmetric, positive-definite weight matrix constructed using the resampling method of Del Boca et al. (2014). In particular, the resampled moment vector $M_N^g, g = 1, ..., Q$ is calculated by bootstrapping the original data Q times.[27] Then, the weight matrix is the inverse of the covariance matrix of M_N:

$$W_n = Q^{-1} \left(\sum_{g=1}^{Q} (M_N^g - M_N)(M_N^g - M_N) \right)^{-1}$$

Del Boca et al. (2014) show the consistency of this type of estimator for large simulations, $plim_{R\to\infty} \tilde{M}_{N,R}(\Omega_0) = M_N(\Omega_0)$.[28] Given identification and these regularity conditions,

$$plim_{N\to\infty} plim_{R\to\infty} \hat{\Omega}_{N,R,W} = \Omega \text{ for any positive definite } W$$

1.4.3. Identification and selection of moments

My model is not nonparametrically identified, for reasons related to those given in Flinn and Heckman (1982) and Flinn (2006). However, it is useful to briefly discuss the identification

[27] In practice, I set Q equal to 200.

[28] Compared with directly calculating the optimal weighting matrix, this method simplifies computation significantly. Altonji and Segal (1996) discuss that gains from using an optimal weighting matrix may be limited.

in the model of Flinn (2006) given its close relationship with this paper. The model in Flinn (2006) can be regarded as a special case of my model when there is only one type of worker ($a_l = a_h$), one pair of counties and no labor mobility ($M(a, j) = 0$). The only job search decision for the worker is θ^*. Even in this specific case, the model is still unidentified because accepted wage and duration information is not enough for nonparametric identification. He further shows that a center class of parametric distributional assumption G, referred to as the "recoverability condition", is required.[29] In my model, given the assumed log-normal distribution of matching quality, all parameters are identified except for the set of discount factor and unemployment utility(ρ, b) because those parameters enter into the likelihood function through the critical value θ^*. Parameters b, η, G, λ will be identified given a fixed value of discount factorρ. Although I use the moments-based estimator rather than the likelihood-based estimator in Flinn (2006), their identification argument can be carried over in this paper given the same log-normal distribution assumption of θ and ex-ante fixed value of ρ.[30]

This paper extends Flinn (2006) in two dimensions by incorporating multiple worker types and multiple connected markets. As a result, instead of one critical value θ^*, individuals make two optimal decisions: accept local offer if $\theta \geq \theta^*(a, j)$ and accept neighboring offer if $\theta \geq \theta^{**}(a, j)$. Now I focus my attention on the log-wage distribution in one local county j. There are four different group of workers: high type natives, low type natives, high type movers, and low type movers. Given Equation 1.3, the log-wage distribution of local workers and the distribution of mobile workers only differ in the truncated values of their distributions. Besides the truncated log normal distribution, there is also a mass point as the left end at value m_j when the minimum wage is binding. As a result, the log wage distribution R should be a left truncated normal distribution with a potential mass point at its left end. We use R_0 to represent the distribution for natives and R_1 to represent the

[29] A comprehensive discussion about this "recoverability condition" can be found in Flinn and Heckman (1982).

[30] The identification depends on the proper selection of moments to characterize the wage distribution. I discuss this in Table 4.

distribution for movers.

$$\begin{array}{lll} \text{Natives} & \log w(\theta, a, j) \sim & R_0(\log w; a, \mu_\theta, \sigma_\theta, \alpha_j, m_j, \theta^*) \\ \text{Movers} & \log w(\theta, a, j) \sim & R_1(\log w; a, \mu_\theta, \sigma_\theta, \alpha_j, m_j, \theta^*, \theta^{**}) \end{array}$$

Since the fractions of local workers $L(a,j)$ and Mobile workers $M(a,j)$ are observed for the four groups of workers, it is straightforward to verify that the parameters $\mu_\theta, \sigma_\theta, a_l, a_h, \alpha_j, \theta^*$ are identified directly. To identify θ^{**}, one additional support condition $\theta^{**}(a,j) > \frac{m_j}{a}$ should be satisfied. Otherwise, the mobile worker's wage distribution R_1 would be identical to local workers' wage distribution, leaving $\theta^{**}(a,j)$ unidentified.

Therefore, I use the fraction of movers Fr, to help identify $\theta^{**}(a,j)$ as well as the moving cost term $c(a,j)$. First of all, I note that

$$Fr(a,j) = \frac{\tilde{G}(\max\{\theta^{**}(a,j), \frac{m_{j'}}{a}\})}{\tilde{G}(\max\{\theta^{*}(a,j), \frac{m_{j}}{a}\})}$$

Given that $\tilde{G}$ and $\theta^*(a,j)$ are already identified, the critical value $\theta^{**}(a,j)$ is identified directly from the observed $Fr(a,j)$.[31] Moving costs can then be backed out from the following one-to-one mapping:

$$c(a,j) = \frac{\alpha_{j'} a(\theta^{**}(a,j) - \theta^*(a,j'))}{\rho + \eta_{j'}} + \frac{a(\theta^*(a,j') - \theta^*(a,j))}{\rho}$$

Given the identified $c(a,j)$ and observed migration/commuting choices $P_0(a,j)$ and $P_1(a,j)$, the choice-specific moving cost $cc_0(a,j)$ and $cc_1(a,j)$ are identified by the logit assumption of equation 1.11.

Althrough the bargaining power α_j can be identified from R_0 and R_1, Flinn (2006) uses a Monte Carlo experiment to show its practical power is tenuous. Because of this, I use the average payroll share of firms' expenditures from the Economy Wide Key Statistics (EWKS), which is the U.S. government's official five-year measure of American business

[31] $\theta^{**}(a,j)$ is potentially not identified when $Fr(a,j) > 1$, which means the number of movers are larger than the number is local workers. However, this situation rarely happens empirically.

and the economy. This payroll share is calculated at the county level and provides cross-sectional variation of the labor share α_j.

The identification of the vacancy cost ψ_j follows from Equation 1.10 as long as the matching technology ω_j is known. Flinn (2006) uses multiple cross sections with different minimum wages to identify ω_j based on the assumption that the economy is in a steady-state in both measurements and the vacancy cost is constant.[32] In this paper, I use a market tightness index (job demand/labor supply) constructed from the Conference Board Help Wanted OnLine (HWOL) data, which is widely used in the macroeconomic literature as a direct measure of matching technology that does not impose any additional assumptions.[33][34]

Table 4 summarizes the empirical moments used to identify the model parameters.

1.4.4. Model fit

My model reproduces many features of the data (Table 5). It predicts a higher employment rate and higher average hourly wage for high-type workers compared with those for low-type workers. The fraction of migrants for both low-type workers and high-type workers are also well matched. Although the fraction of high-type commuters is almost perfectly predicted, the fraction of low-type commuters is over-predicted. My model also correctly predicts the correlation between labor mobility patterns and the geographic characteristics (rent prices and physical distance between paired counties). My model replicates the negative correlation between mobility rates and housing prices. I observe low numbers of migrants and commuters in counties with relatively high rental prices. On the other hand, both simulation and data find a positive correlation between migration and distance but a negative correlation between commuting and distance.

[32] See the discussion of Condition C-Coherency in Flinn (2006) for more details.

[33] Beginning in 2005, HWOL provides a monthly series that covers the universe of vacancies advertised on about 16,000 online job boards and online newspaper editions. While HWOL only collects the job openings advertised online, its pattern is quite similar with the general pattern measured by JOLTS, especially before 2013. A detailed comparison between HWOL and JOLTS can be found in Şahin et al. (2014).

[34] See Petrongolo and Pissarides (2001) for a survey of these studies.

Table 4: Selection of Moments

Empirical moments	County j		County j'		Identified
	Mean	S.D.	Mean	S.D.	Parameters
Moments from mean and S.D. in county pair $p(j,j')$					
Employment rate (high type)	0.881	0.083	0.886	0.078	$\mu_b, \sigma_b, \mu_\psi, \sigma_\psi$
Employment rate (low type)	0.785	0.127	0.761	0.127	$\mu_b, \sigma_b, \mu_\psi, \sigma_\psi$
Average hourly wage (high type)	13.63	2.47	-	-	$\mu_b, \sigma_b, a_h, \mu_G, \sigma_G$
Average hourly wage (low type)	9.23	2.57	-	-	$\mu_b, \sigma_b, a_l, \mu_G, \sigma_G$
Proportion of migrants (high type)	0.073	0.050	0.070	0.046	$\mu_{\beta 0}, \sigma_{\beta 0}, \beta_{0a}$
Proportion of migrants (low type)	0.042	0.037	0.037	0.039	$\mu_{\beta 0}, \sigma_{\beta 0}, \beta_{0a}$
Proportion of commuters (high type)	0.113	0.127	0.096	0.106	$\mu_{\beta 1}, \sigma_{\beta 1}, \beta_{1a}$
Proportion of commuters (low type)	0.084	0.102	0.072	0.091	$\mu_{\beta 1}, \sigma_{\beta 1}, \beta_{1a}$
Correlation between migrants and commuters	0.630	-	0.523	-	ρ_β
Correlation between migrants and distance	0.149	-	0.014	-	β_{0d}
Correlation between commuters and distance	0.008	-	-0.168	-	β_{1d}
Correlation between migrants and rent cost	-0.103	-	-0.056	-	$\beta_{0\gamma}$
Correlation between commuters and rent cost	-0.116	-	-0.110	-	$\beta_{1\gamma}$
Correlation between migrants and rent cost					
Correlation between employments (high type)	0.318	-	-	-	ρ_b, ρ_ψ
Correlation between employments (low type)	0.211	-	-	-	ρ_b, ρ_ψ
Correlation between separation rate	0.599	-	-	-	ρ_ψ
Correlation between wage rate	0.498	-	-	-	ρ_b, ρ_ψ
Moments directly measure parameter values					
Separation rates in county j(quarterly)	0.353	0.130	0.358	0.132	η_j
Bargaining power α_j in county j	0.311	0.044	0.310	0.043	α_j
Matching technology ω_j in state $s(j)$	1.36	0.385	1.41	0.406	ω_j
Centroid distance $d_{jj'}$ between j and j'	66.6	45.9	66.6	45.9	$d_{jj'}$
The median rent cost (local amenity γ_j) in j	683	168	683	178	γ_j

Note: (i) For details about the construction of the empirical moments, see Appendix 1.10. (ii) County j represents the county which increases its minimum wage, while county j' is the county keeps the minimum wage fixed.

Table 5: Model fit

Empirical moments	County 1		County 2	
	Data	Sim	Data	Sim
Employment rate (high type)	0.883	0.829	0.888	0.827
Employment rate (low type)	0.754	0.789	0.765	0.786
Hire rate	0.375	0.354	0.361	0.348
Average hourly wage (high type)	13.630	13.385	-	-
Average hourly wage (low type)	9.230	9.156	9.230	9.156
Proportion of migrants (high type)	0.074	0.075	0.069	0.070
Proportion of migrants (low type)	0.043	0.047	0.038	0.045
Proportion of commuters (high type)	0.109	0.114	0.094	0.107
Proportion of commuters (low type)	0.082	0.143	0.071	0.131
Correlation between migrants and commuters	0.612	0.695	0.510	0.732
Correlation between migrants and distance	0.123	0.079	0.066	0.031
Correlation between commuters and distance	-0.079	-0.011	-0.155	-0.071
Correlation between migrants and rent cost	-0.101	-0.069	-0.063	-0.103
Correlation between migrants and rent cost	-0.099	-0.029	-0.098	0.011

1.5. Estimation results

In this section, I present the parameter estimates and discuss their magnitudes. I then compare the elasticities of migrants and comments with respect to minimum wage changes predicted by the model with the elasticities estimated in the previous regressions. Finally, I quantify the downward bias in the estimation of disemployment effect when ignoring labor mobility.

1.5.1. Understanding the model estimates

Table 46 provides model estimates for both the general parameters and the parameters in the moving equation (Equation 1.11). The estimated value of unemployment, $\left(b_j, b_{j'}\right)$, is relatively homogeneous across counties. However, the vacancy cost ψ displays considerable heterogeneity across counties. Its mean value is 428, which is equivalent to \$68,480 if the filled worker is required to work 160 hours/month. Furthermore, the large standard error suggests substantial spatial diversification in vacancy costs. In addition, I find the productivity of high educated workers is on average significantly higher than the productivity of low educated workers ($a_h = 3.106$ vs. $a_l = 1.406$). When comparing mobility costs,

Table 6: Parameter estimates

General parameters				
Parameters	Notation	Mean μ	S.D. σ	Corr. τ
Matching quality	θ	1.963	0.162	-
Unemployed flow utility	b	-23.8	0.123	0.949
Vacancy cost	ψ	428	211	0.196
High type productivity	a_h	3.106	-	-
Low type productivity	a_l	1.406	-	-
Commuting cost	β_0	48.5	1.217	0.458
Migration cost	β_1	78.4	9.709	
Coefficients in equation $cc_h(a, j)$				
	Commuting ($h = 0$)		Migration ($h = 1$)	
Additional cost for high type	β_{0a}	2.222	β_{1a}	-4.927
Coefficient for different local amenity	$\beta_{0\gamma}$	2.884	$\beta_{1\gamma}$	7.709
Coefficient for different distance	β_{0d}	0.697	β_{1d}	-2.051
Scale of preference shock (low type)	σ_l	15.0		
Scale of preference shock (high type)	σ_h	25.0		

migrating is more costly ($\beta_1 = 78.4$) than commuting ($\beta_0 = 48.5$), which explains why the fraction of commuters is on average larger than the fraction of migrants.

The lower panel in Table 46 reports the the determinants of choice-specific moving costs $cc_h(a, j)$. The positive sign of β_{0a} and negative sign of β_{1a} indicate that, compared to low educated workers, high educated workers are more likely to migrate when accepting the job offers from a neighboring county. These two coefficients rationalize the observation that when looking at commuting behavior, 40% of high educated workers are commuters whereas only 34% of low educated workers are commuters. The next two coefficients, $\beta_{0\gamma}$ and $\beta_{1\gamma}$ link the moving cost with the local housing rental price, which is regarded as a proxy of local amenities . The positive values of $\beta_{0\gamma}$ and $\beta_{1\gamma}$ mean that high housing costs are associated with high moving costs. Workers are less likely to take neighboring jobs when the housing price in neighboring county is high. Even accepting the neighboring jobs, the mobile workers are less likely to choose migration as their preferred moving option. The coefficients β_{0d} and β_{1d} capture the correction between physical distance and moving cost. The positive sign of β_{0d} and negative sign of β_{1d} reflects the pattern that more mobile workers would choose migration over commuting as county pairs are farther apart. Lastly, the scale parameters

Table 7: Moving costs and neighboring county preference

	Lump-sum ex-ante moving cost (unit: \$)				Indifferent opportunity			
	Low educated		*High educated*		*Low educated*		*High educated*	
	County j	County j'	County j	County j'	County j	County j'	County j	County j'
10th	7,749	7,773	7,648	7,626	0.010	0.008	0.022	0.014
25th	8,400	8,286	8,602	8,459	0.047	0.036	0.092	0.067
Median	8,794	8,760	9,262	9,210	0.222	0.181	0.229	0.211
75th	9,194	9,109	9,818	9,741	0.880	0.823	0.561	0.495
90th	9,691	9,544	10,410	10,365	1.023	1.018	0.997	0.974
Mean	8,693	8,683	9,098	9,067	0.407	0.373	0.367	0.336
SD	1,046	931	1,330	1,242	0.405	0.402	0.351	0.333

Note: the dollar value of ex-ante moving cost $c(a, j)$ is estimated based on a representative full time worker whose working time is 160 hours/month.

for low-educated workers is smaller than that for high educated workers.

Table 7 reports the distributions of moving costs. The left panel displays summary statistics of the ex-ante moving costs $c(a, j)$ for workers differentiated in their types and locations. According to my estimates, the ex-ante moving cost is on average \$8,700 for low type workers and \$9,100 for high type workers. These costs are summary statistics of relocation costs, housing market transaction costs (for migrants only) and psychic costs.(Schwartz (1973), Greenwood (1975)) The estimated moving costs are lower than previous ones reported in the literature. For example, Kennan and Walker (2011) estimate a moving cost value of \$312,000 for an average movers across states in the US. And Schmutz and Sidibe (2016) find the average moving cost among French cities is around € 15,000. The moving costs in my paper are lower for two reasons: first, I focus on the migration/commuting flows between two contiguous counties. The geographic proximity could greatly reduce the potential moving costs. Secondly. I focus on younger workers who are more likely to be affected by minimum wages. The opportunity costs of moving for those workers are relatively low.

Moreover, the moving cost can be equivalently measured using the openness of the local labor market. The right panel illustrates this idea and calculates the possibility that a random job from a neighboring county is preferred to a random job from the local county. The probability distribution for low type workers are less diversified than that for high type workers. The probability of preferring a neighboring county ranges from 0.010 (the 10th percentile) to 1.023 (the 90th percentile) for low educated workers in county j; this probability shrinks to a range of 0.020 (the 10th percentile) to 0.997 (the 90th percentile) for high educated workers. I also find this distribution is right skewed. In the median county, the probability for a low-skilled worker to receive a preferred job from neighboring county is 22.2%. This effect is comparable to the results of Manning and Petrongolo (2017). Using UK data, they find that the probability of a random job 5km distant being preferred to random local job is only 19%.

1.5.2. Out-of-sample validation: comparing model-based predictions with regression results

In this section, I use the model to predict the minimum wage elasticities of commuters and migrants and then compare the predicted elasticities to actual elasticities estimated by regression 1.14. This comparison is treated as an extra out-of-sample validation since the elasticities of commuters and migrants with respect to minimum wage are not used as targeted moments when estimating the baseline model.

Given county specific parameters and local minimum wage levels, the model allows me to calculate the fraction of migrants and commuters in each county . Specifically, the fraction of migrants in county j given minimum wage pair $MI(a, j; m_j, m_{j'})$ is expressed as the number of migrants from county j' to county j, divided by the sum of local hires in county j, and total mobile hires from county j', i.e.

$$MI(a, j; m_j, m_{j'}) = \frac{P_1(a, j')U(a, j')\tilde{G}(\max\{\theta^{**}(a, j'), \frac{m_j}{a}\})}{U(a, j)\tilde{G}(\max\{\theta^{*}(a, j'), \frac{m_j}{a}\}) + U(a, j')\tilde{G}(\max\{\theta^{**}(a, j'), \frac{m_j}{a}\})}$$

Meanwhile, the fraction of commuters in county j given the local minimum wage pair $CM(a, j; m_j, m_{j'})$ is given by the total number of commuters from county j, divided by the sum of local hires in county j' and all mobile workers (both commuters and migrants) from county j, i.e.

$$CM(a,j;m_j,m_{j'}) = \frac{P_0(a,j)U(a,j)\tilde{G}(\max\{\theta^{**}(a,j),\frac{m_{j'}}{a}\})}{U(a,j')\tilde{G}(\max\{\theta^{*}(a,j'),\frac{m_{j'}}{a}\}) + U(a,j)\tilde{G}(\max\{\theta^{**}(a,j),\frac{m_{j'}}{a}\})}$$

When the minimum wage in county j increases from m_j to $m_j + \Delta m_j$ but the minimum wage in county j' remains unchanged, I calculate new fractions of commuters $CM(a,j;m_j + \Delta m_j, m_{j'})$ and migrants $MI(a,j;m_j,m_{j'})$ in the new steady-state. The percentage changes in labor mobility are defined as:

$$\begin{aligned}
\Delta \log MI(a,j) &= \log(MI(a,j;m_j+\Delta m_j,m_{j'}) - \log MI(a,j;m_j,m_{j'}) \\
\Delta \log CM(a,j') &= \log(CM(a,j';m_j+\Delta m_j,m_{j'}) - \log CM(a,j';m_j,m_{j'})
\end{aligned}$$

Using data on minimum wage changes, I predict $\Delta \log MI(a,j)$ and $\Delta \log CM(a,j')$. Figure 2 displays the distributions of $\Delta \log MI(a,j)$ and $\Delta \log CM(a,j')$ for different types. First, all distributions show substantial heterogeneity across county-pairs, suggesting local markets are diversified. Minimum wage hikes decrease the chance of finding a job but increase the expected wages once hired. When the cost exceeds the benefit, the local labor market becomes less attractive, and workers either move away or stop moving in. The mean value of $\Delta \log CM(low,j)$ is positive (0.034) whereas the average value of $\Delta \log MI(low,j)$ is negative (-0.034), both of which indicate that low-skilled workers are more likely to leave areas with higher minimum wages in the majority of county pairs. Second, the distributions for low-skill workers are more dispersed than those for high-skilled workers. This is in line with the observation that low-skill workers are more responsive to minimum wage changes.

Next, I check the out-of-sample validation by comparing the model generated $\Delta \log MI(a,j)$ and $\Delta \log CM(a,j')$ with the data. In the sample, the percentage changes of migrants and commuters are directly calculated by comparing the fractions of mobile workers before minimum wage changes with those after minimum wage changes. Then I run the following regression to compute the minimum wage elasticity from model predictions ("Model-based

Figure 2: The distribution of $\Delta \log MI(a, j)$ and $\Delta \log CM(a, j')$ after minimum wage hikes

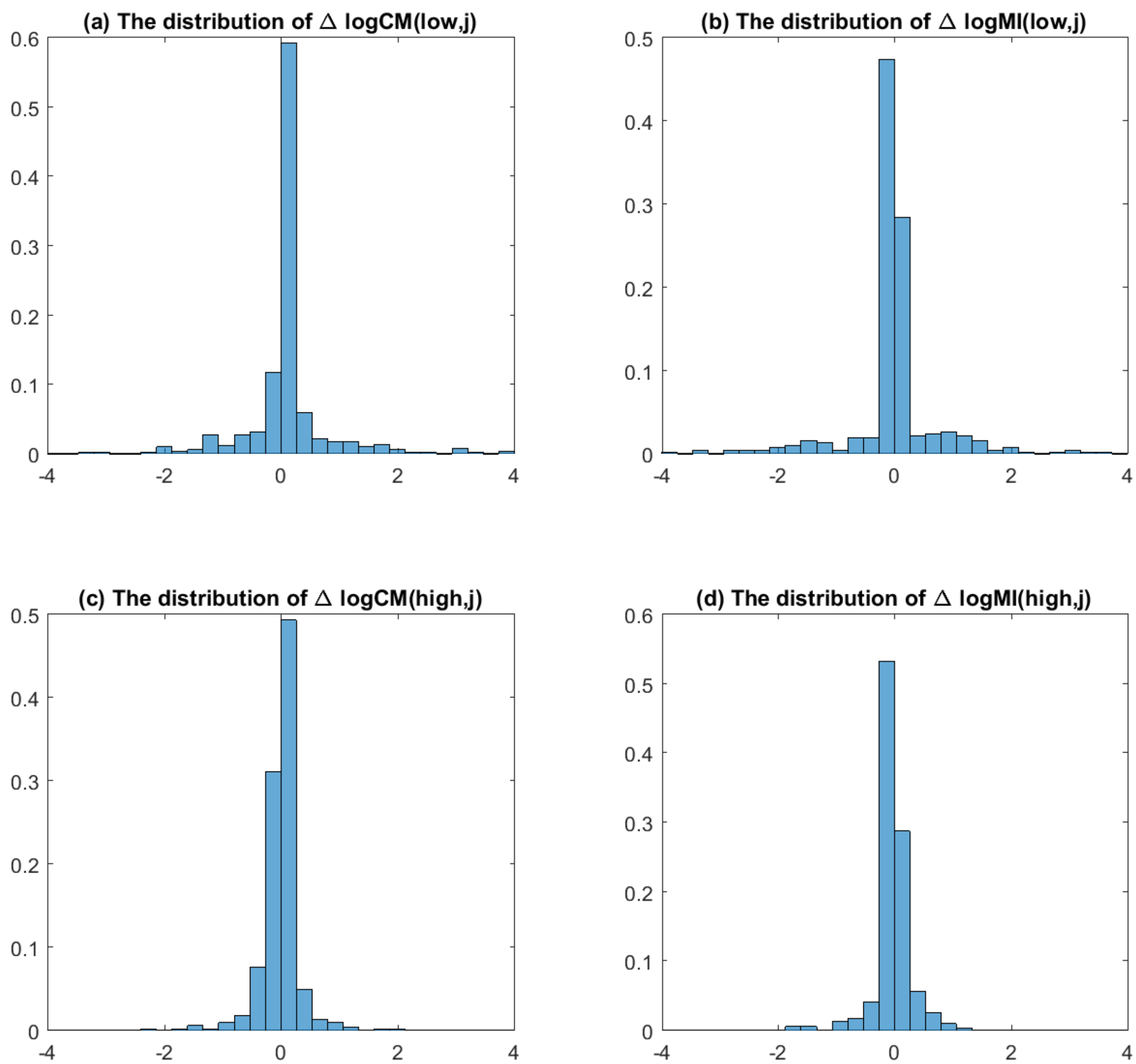


Table 8: The comparison between model predictions and regression results

	Model-based (β_1^*)	Data-based (β_1)
Low skilled Commuters	0.741***	0.458**
	(0.234)	(0.215)
Low skilled Migrants	-0.590**	-0.589***
	(0.260)	(0.160)
High skilled Commuters	-0.282***	0.263**
	(0.082)	(0.133)
High skilled Migrants	-0.081	-0.101
	(0.080)	(0.112)

*Note: The regression column is directly from Table 10. Standard errors are displayed in parentheses. * for 10%. ** for 5%, and *** for 1%.*

elasticity β_1^*") and from data observations ("Data-based elasticity β_1") separately:

$$\begin{aligned} \Delta \log MI(a,j) &= \beta_1^* \Delta logMW_j + \Delta\epsilon_0 \\ \Delta \log CM(a,j) &= \beta_1^* \Delta logMW_j + \Delta\epsilon_1 \end{aligned} \tag{1.12}$$

The regression results based on the data were previously calculated in Table 10 since regression 1.12 is a simplified version of Equation 1.14 that ignores the county fixed effect and restricts the observational period. Table 8 shows that the model-based β_1^* and the data-based β_1 are comparable. For low educated mobile workers, both estimates suggest that they exit counties with minimum wage hikes. And the magnitudes of both elasticities are very similar (within a 90% confidence interval). In addition, both estimates find the elasticities for low educated workers(absolute value) are larger than the elasticities for high skilled workers. This is consistent with the intuition that low educated workers are more responsive to the minimum wage adjustments.

The model-based elasticity for more highly educated commuters is less consistent with data-based elasticity. This discrepancy can be attributed to the distinction between short- and long- run effects. Althrough the data-based β_1 captures the immediate response after the minimum wage change, the model-based β_1^* demonstrates cumulative changes between two steady states.

This distinction between the short- and long-run effects is also emphasized in Sorkin (2015).

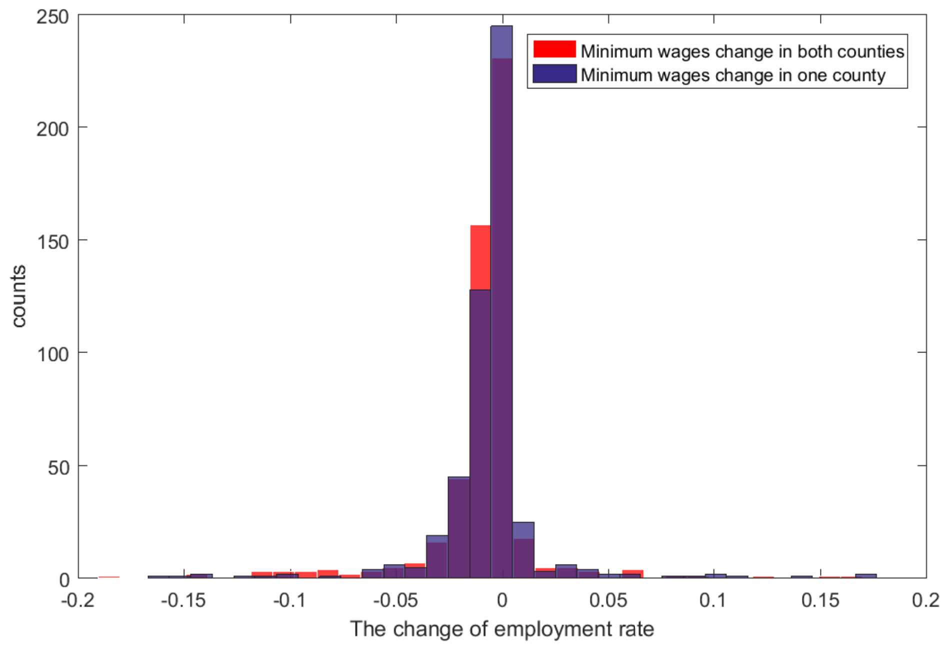
He argues that the reduced-from effects are essentially uninformative about the true long-run elasticity. In my case, the key reason is the sorting of workers provides additional feedback effects in the long run. When a local county increases its minimum wage, the fraction of low type workers decreases in local county but increases in the neighboring county. As the average worker quality improves in the local market, firms have more incentive to post vacancies in local county rather than in neighboring county. However, this feedback effect is hard to be observed in the short run since the adjustment of local worker quality is slow. I will further explore this mechanism in Section 1.6.

1.5.3. Quantifying the underestimation of disemployment effects when ignoring labor mobility

Starting with Card and Krueger (1994), cross-border comparisons became a common method of studying the disemployment effects of the minimum wage. Dube et al. (2010) and Dube et al. (2016) generalize this strategy to all county pairs and find limited disemployment effects, which is consistent with Card and Krueger (1994). Although the cross-border design allows one to assume similarity between the treated area and control area, it may be problematic. As pointed out by Neumark et al. (2014b), "spillover effects can certainly contaminate the control observations. If workers displaced by the minimum wage find jobs on the other side of the border, employment will expand in the control areas". Based on my model, I quantitatively evaluate two sources of the underestimation of disemployment effects. First, unemployed workers who leave are "missing" from the treated county. Second, they may "reappear" in the neighboring county, contaminating the control group.

To evaluate the first channel, I compare the disemployment effect from two different minimum wage increases. In case 1, both counties increase their minimum wage by the same percentage $(m_j, m_{j'}) \rightarrow (m_j + \Delta m_j, m_{j'} + \Delta m_{j'})$. In case 2, only one county increases its minimum wage $(m_j, m_{j'}) \rightarrow (m_j + \Delta m_j, m_{j'})$. In case 1, the geographical minimum wage differences are more compressed since the minimum wage increases in both counties rather than increase only in one local county. Therefore, the opportunity to arbitrage relative

Figure 3: The disemployment effect under different minimum wage hikes



minimum wage differences are largely eliminated in case 1 compared with case 2. The disemployment effect caused by minimum wage hikes is defined as the change of the log employment rate under the steady-state before minimum wage change and the new steady-state after the minimum wage change:

$$
\begin{aligned}
\text{Case 1:} \quad \Delta \log \mathit{Emp}_j &= \log \mathit{Emp}_j(m_j + \Delta m_j, m_{j'} + \Delta m_{j'}) - \log \mathit{Emp}_j(m_j, m_{j'}) \\
\text{Case 2:} \quad \Delta \log \mathit{Emp}_j &= \log \mathit{Emp}_j(m_j + \Delta m_j, m_{j'}) - \log \mathit{Emp}_j(m_j, m_{j'})
\end{aligned}
$$

Figure 3 compares the distribution of $\Delta \log \mathit{Emp}_j$ under case 1 and case 2. The average value of $\Delta \log \mathit{Emp}_j$ in case 1 is more negative than that in case 2 while the distribution of $\Delta \log \mathit{Emp}_j$ in case 1 (red histogram) is more right-skewed than in case 2 (blue histogram). 86.9% of counties in case 1 experience negative employment changes due to minimum wage hikes compared to only 82.0% in case 2. This comparison confirms that the spillover effect actually attenuates the disemployment effect.

Next, I calculate the minimum wage elasticity of employment by estimating the following regressions:

$$\text{Case 1:} \quad \Delta \log \mathit{Emp}_j = \beta_1 \Delta \mathit{logMW}_j + \Delta \epsilon_1$$
$$\text{Case 2:} \quad \Delta \log \mathit{Emp}_j = \beta_2 \Delta \mathit{logMW}_j + \Delta \epsilon_2$$

To include the potential bias caused by the contamination of control group, I recalculate the disemployment elasticity in case 2 using the neighboring county as the control group. This calculation mimics the diff-in-diff approach:

$$\text{Case 3:} \quad \Delta \log \mathit{Emp}_j - \Delta \log \mathit{Emp}_{j'} = \beta_3 \Delta \mathit{logMW}_j + \Delta \epsilon_3$$

Table 9 reports the minimum wage elasticity of employment in all three cases. "Case 1" reports the elasticity of employment when both counties increase their minimum wages by the same proportion. "Case 2" reports the elasticity of employment when only the local county increases its minimum wage. Finally, "Case 3" displays the alternative elasticity if the neighboring county is used as the control group. Workers in case 1 have less incentive to arbitrage the minimum wage difference between two counties compared with their incentive in case 2. Therefore, changes in labor mobility after minimum wage hikes in case 1 is smaller than changes in case 2. Consequently, I observe a larger disemployment effect in case 1 (-0.0733) compared with case 2 (-0.0421). Furthermore, when using the neighboring county as the control group, the disemployment effect continues to shrink from -0.0421 to -0.0341. This shares the same pattern with the different disemployment effect estimated in Table 11. In Table 11, the minimum wage elasticity of employment changed from -0.068 to -0.039 after controlling for pair-specific time trends instead of a common time trend. Dube et al. (2016) argue that this change is driven by spatial heterogeneity. My findings suggest that such changes are driven by labor mobility rather than by spatial heterogeneity. This result highlights the concern that neighboring counties, despite their geographic proximity, may not be the appropriate control group due to the contamination caused by labor mobility.

If labor mobility is causing underestimation of the disemployment effect, then the bias

Table 9: Elasticity of employment with respect to the minimum wage

	Case 1 (β_1)	Case 2 (β_2)	Case 3 (β_3)
Whole Sample	-0.0733***	-0.0421***	-0.0341***
	(0.0069)	(0.0075)	(0.0096)
Below bottom quartile	-0.0957***	-0.0445***	-0.0153
of moving cost	(0.0163)	(0.0136)	(0.0190)

*Note: "Case 1" reports the elasticity of employment when both county increase their minimum wages by the same proportion. "Case 2" reports the elasticity of employment when only the local county increase its minimum wage. "Case 3" displays an alternative elasticity if the neighboring county is wrongly picked as the control group. Standard errors are displayed in parentheses. * for 10%. ** for 5%, and *** for 1%.*

should be larger for counties with lower moving costs. To verify this conjecture, I conduct an additional placebo test for a sub-sample of counties whose moving costs are in the bottom quartile. My estimates, reported in the second row of Table 9, are in line with this conjecture. First, the difference of the elasticities between case 1 and case 2 becomes larger when using the restricted sample. The main reason is that the disemployment effect in case 1 is larger (-0.0957) compared with the previous effect (-0.0733) using the full sample. Second, using the neighboring county as the control group creates more severe downward bias. Although the elasticities in case 2 are robust to different sub-samples, it goes down sharply to -0.0153 in case 3 when using the neighboring county as the control group. To summarize, ignoring labor mobility and potential spillover effects cause the disemployment effect to be underestimated.

1.6. Policy experiments

In this section, I use the estimated model to examine the distributional impacts of local minimum wage hikes. There are (at least) two criteria to evaluate the welfare consequences of the minimum wage polices. The first natural welfare candidate is the value of unemployment $V_u(a, j)$, which can also be interpreted as the ex-ante welfare of heterogeneous workers with different types a and locations j. This is my primary measure because my goal is to understand the distributional effects for heterogeneous workers under minimum wage hikes. A second welfare criteria is defined for the local government, which is of particular interest when considering the total spillovers of local minimum wage policy to the neighboring county. Following Flinn (2006), I assume that the minimum wage is the only policy instrument available to the local government and the welfare function of local government defined as follows:

$$
\begin{aligned}
W_j(m_j) &= \textstyle\sum_{a\in\{a_l,a_h\}}[\underbrace{L(a,j)\bar{V}_e(\theta,a,j,\theta^*(a,j))}_{\text{(1) Local employed workers}} + \underbrace{MI(a,j)\left(\bar{V}_e(\theta,a,j,\theta^{**}(a,j')) - c(a,j')\right)}_{\text{(2) Migrants from neighbouring county}} \\
&+ \underbrace{CM(a,j')\left(\bar{V}_e(\theta,a,j',\theta^{**}(a,j')) - c(a,j)\right)}_{\text{(3) Commuters to the neighbouring county}} + \underbrace{U(a,j)V_u(a,j)}_{\text{(4) Unemployed workers}} \\
&+ \underbrace{E(a,j)\bar{V}_f(a,j)}_{\text{(5) Revernue from filled vacancies}} \quad] - \underbrace{K_j\psi_j}_{\text{(6) Total cost of vacancies}}
\end{aligned}
$$

where (1) $L(a,j)$ is the population of local employed workers with $\bar{V}_e(\theta,a,j,\theta^*(a,j))$ denoting their average welfare. (2) $MI(a,j)$ is the population of migrants who move from county j', with $\bar{V}_e(\theta,a,j,\theta^{**}(a,j')) - c(a,j')$ as their average net welfare. (3) $CM(a,j')$ is the population of migrants who commute to work in county j', with $\bar{V}_e(\theta,a,j',\theta^{**}(a,j)) - c(a,j)$ as their average net welfare. (4) $U(a,j)$ is the population of local unemployed workers (all unemployed workers have same welfare level $V_u(a,j)$). On the demand side of the market, while there are K_j vacancies in county j, only $M_j = \sum_a E(a,j)$ are filled with workers and generate positive revenue. The free entry condition guarantees that the revenue generated from the filled vacancy is equal to the total cost of posted vacancies in the steady state. Thus the total contribution of terms (5) and (6) is equal to 0.

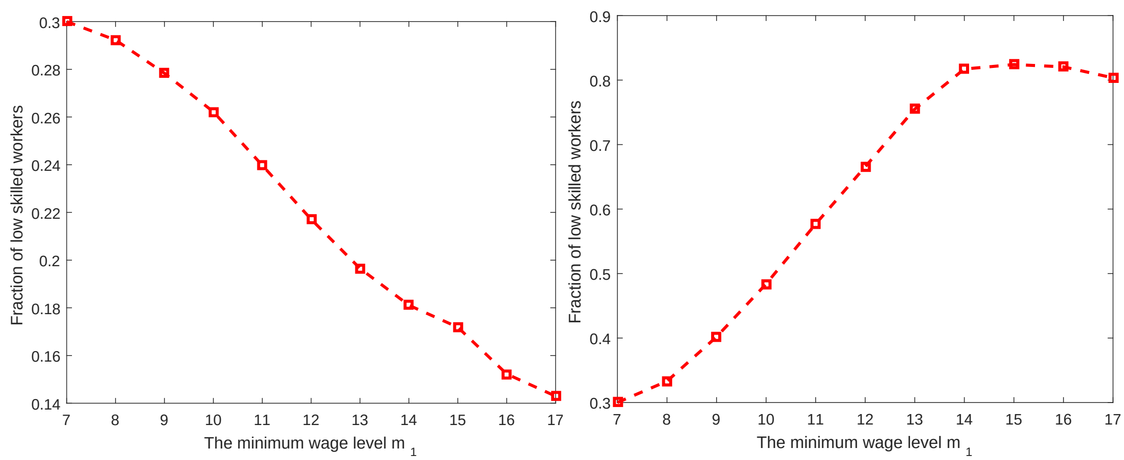
To understand the distributional effects of local minimum wage hikes, it is important to recognize the different forces at play. Assume county 1 changes its minimum wage while county 2 keeps its minimum wage unchanged. The direct effect in county 1 depends on the trade-off between the decrease in working opportunities (*"disemployment effect"*) and the increase in expected income (*"wage enhancement effect"*). Because the productivity distribution of high type workers first-order stochastically dominates that of low type workers , the working opportunity of the high type is less hurt by the same minimum wage increase compared with that of low type workers. As a result, low type workers have stronger incentives to move out of the country to avoid welfare losses caused by the minimum wage hike. Besides the direct effects, there is an additional general equilibrium effect through the change in a firm's incentive to post vacancies. First, the share of matching surplus decreases when firms are constrained by a higher minimum wage (*"share reduction effect"*). Secondly, due to the assumption of random search, firms are unable to screen workers' type when they post vacancies. Thus, vacancies (per capita) will be negatively correlated with the proportion of low type workers in their local county. Worker sorting decreases the composition of high type in county 2. As a result, the local workers in county 2 suffer additional welfare losses because of the decrease in hiring probability (*"composition changing effect"*). The additional force is the moving costs which generates welfare differences between the same type workers in different locations. Compared with local workers, mobile workers have to pay additional moving costs to work in the same job, *ceteris paribus*. This friction is traditionally referred to as the *lock-in effect*.

After understanding the distributional effects of heterogeneous workers, I conduct two counterfactual policies aiming for the reduction of spillover externalities. In the first policy experiment, I completely restrict labor mobility between counties by increasing the moving cost to infinity. This experiment captures the extreme case when no labor mobility is allowed. In the second policy experiment, the central government preempts the local minimum laws. In other word, the two paired counties follow universal minimum wage hikes rather than setting up their local minimum wages. In reality, the preemption of local minimum wage laws is a popular policy intervention for state legislation to avoid "patchwork" of wage levels within a state. So far, 27 states have passed such laws.

1.6.1. The distributional effect of local minimum wage hikes

In this section I explore how the welfare of workers (differentiated by their type a and location j) changes with respect to local minimum wage changes in county 1. To better exclude the effect of local minimum wage hikes from other disturbances such as geographic asymmetry, I consider symmetric county pairs where the geographic parameters in both counties take the mean values of the distributional estimates. The distributional effects

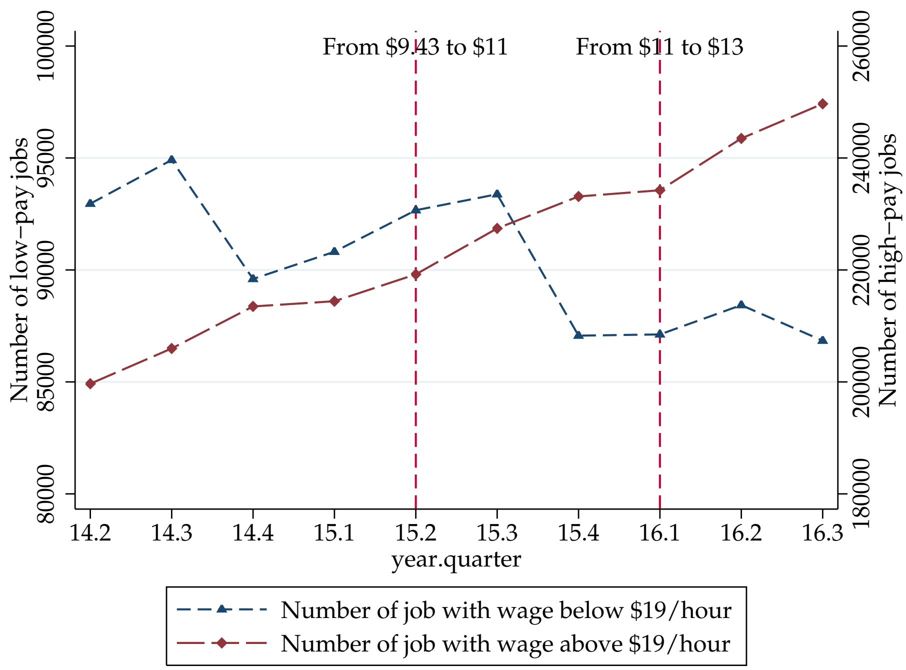
Figure 4: Worker composition under different minimum wages



depend critically on the magnitude of local minimum wage increases. I assume the initial hourly minimum wage in both counties is \$7 and consider welfare changes when increasing the minimum wage in county 1 to an amount between \$7 and \$17. Most of my results are presented in graphical form. I will first report the change in local economic conditions (e.g. contact rates, the composition of heterogeneous workers). Then, I will compute welfare changes with respect to changes in minimum wages for different workers. Lastly, I show welfare changes of local governments with changes in minimum wages.

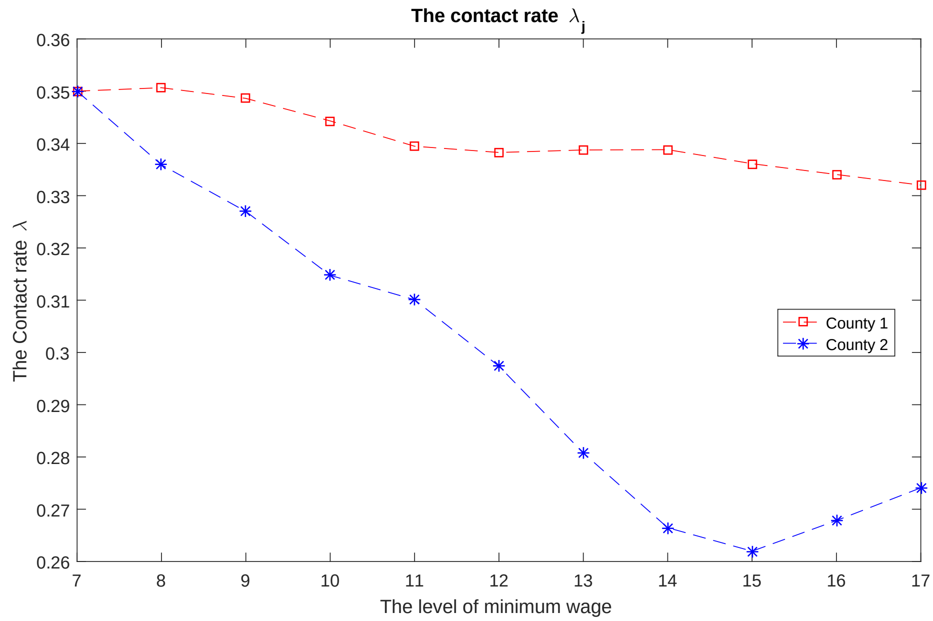
Figure 4 display changes in worker composition in both counties under different minimum wage increases. As the local minimum wage in county 1 increases from \$7 to \$17, the fraction of low type worker in county 1 monotonically decreases to 0.15 while the fraction of low type workers in county 2 has a hump shape with a peak of 0.8 when $m_1 = \$14$. These two patterns suggest that local minimum wage policy serves as a worker selection device. By setting a higher minimum wage, the local government extracts high type workers from the neighboring county while also dumping low type workers on the neighboring county. The prediction in my model is consistent with the real changes happening in Seattle after its city-level minimum wage increases from \$9.43/hour to \$13/hour.(Jardim et al. (2017)) As showed in figure 5, the number of low-pay job ($wage < 19/hour$) decreases while the number of high-pay job ($wage > \$19/hour$) increases.

Figure 5: Changes in Seattle jobs after increasing minimum wage from $9.43 to $13



Data source: administrative employment records from the Washington Employment Security Department, reported in Jardim et al. (2017), table 3.

Figure 6: Contact rates under different minimum wages



Next I consider the changes of firms' incentive to post vacancies. Figure 6 displays changes of contact rates in both counties and suggests two channels of changing the profit of posted vacancies. First, for the same match, firms get less value per vacancy when the minimum wage is higher. A higher minimum wage decreases both the probability that a given match is acceptable and makes the sustainable match less profitable. This channel explains why contact rates in both counties experience a downward change when minimum wage in county 1 increases. Second, the sorting of workers increases the concentration of high types in county 1 but decreases their concentration in county 2 because firms tend to post relatively fewer vacancies in the county with higher fraction of low type workers. The second channel explains why the contact rate in county 2 is systematically lower than that in county 1. Furthermore, the fraction of low type in county 2 reaches its peak at \$14 and starts to decrease after that, which explains the rebound of the contact rate in county 2 when $m_1 \geq \$15$.

The most crucial results are the distributional effects of local minimum wage policies on heterogeneous workers. This heterogeneity is not well explored in the previous literature because workers are often considered to be ex-ante identical (e.g. Flinn (2006)). Let $V_u(a, j; m_1, m_2)$ be the ex-ante welfare for a worker with type a and in location j when m_1 is set at m and m_2 is set at \$7, then the change of welfare is defined as

$$\Delta W_0(a, j; m, 7) = V_u(a, j; m, 7) - V_u(a, j; 7, 7)$$

Figure 7 shows the results. The top left panel displays welfare changes of low type workers in both county 1 (the blue line) and county 2 (the red line). The low type is severely harmed by higher minimum wages. As noted previously, this is driven by a combination of two effects. First, the higher m_1 rules out previously acceptable wages. Second, the higher minimum wage policy in county 1 pushes low type workers to county 2, diminishing their probability to be hired. The top right panel displays the welfare changes of high type workers in both county 1 (the blue line) and county 2 (the red line). The hump shape in high type welfare shows the existence of countervailing effects. Although raising the minimum wage increases workers' welfare by increasing the return of a match, previously acceptable matches become unacceptable. The latter effect dominates the previous effect when local minimum wage in county 1 exceeds \$14.

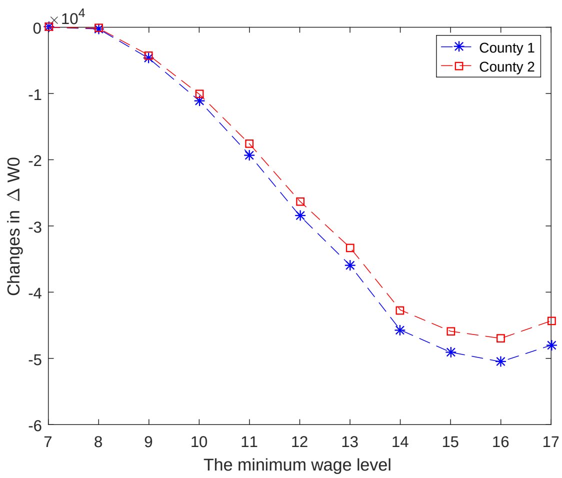
The lower panel of Figure 7 reports the change of inequality between high type and low type as minimum wage increase in county 1. Because the welfare of the low type is a convex curve whereas the welfare of the high type is a concave curve, the inequality curve expands and then reaches its peak when $m_1 = 15$. This result reveals that local minimum wage policy could actually increase inequality between high and low type workers, completely opposite of the intended policy effect.

Lastly, the welfare difference between same type workers in two counties indicates the "lock-in" effects due to the existence of moving costs, I will continue to explore this effect in the next section.

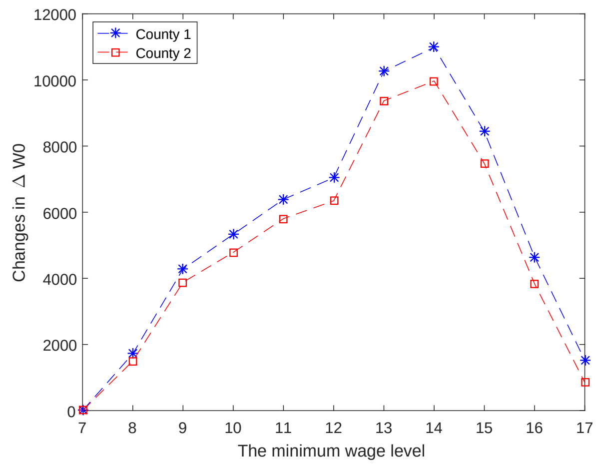
Figure 8 plots the change of total welfare in each county with respect to a change in the

Figure 7: Welfare changes across heterogeneous workers under different minimum wage increases

(a) Welfare changes - low type



(b) Welfare changes - high type



(c) Welfare inequality between high type and low type

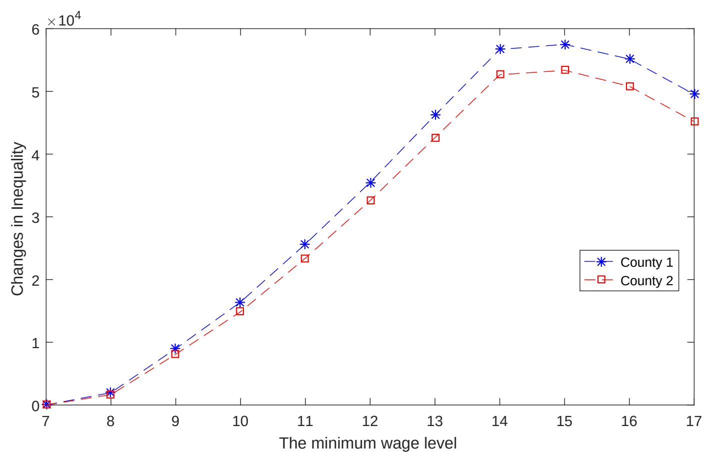
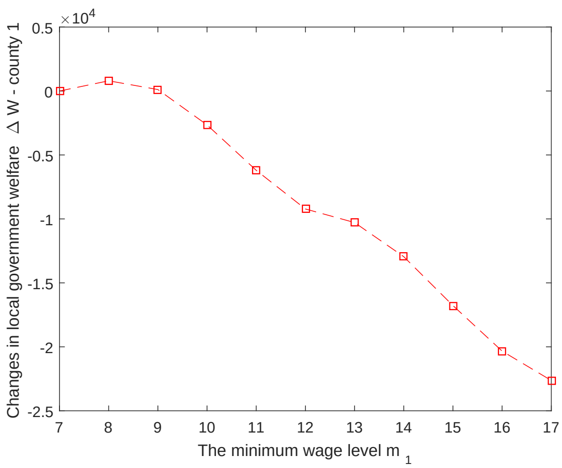
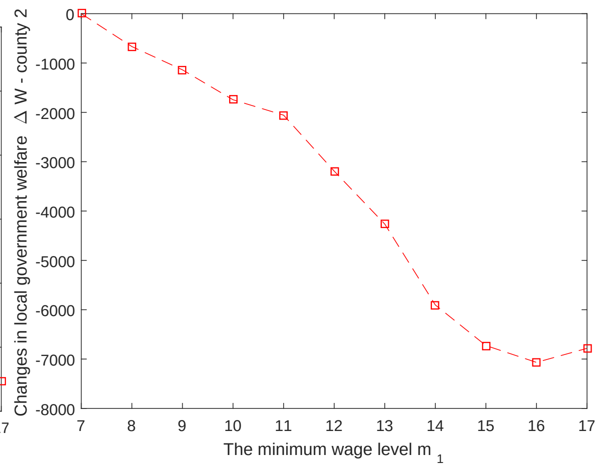


Figure 8: Changes in local government welfare as minimum wage changes in county 1

(a) Changes in local government welfare - county 1



(b) Changes in local government welfare - county 2



local minimum wage. The total welfare in county 1 has a single peak at $m_1 = 8$, while the total welfare in county 2 declines until $m_1 = 16$. An increase in m_1 almost always harms the total welfare in county 2. Put another way, the increases in local minimum wages generate negative externalities to neighboring counties.

1.6.2. Restricting labor mobility between counties

In this session, I want to reduce spillover externalities by completely blocking labor mobility between counties. This can be treated as an extreme way of implementing mandatory local hiring requirements. For example, the public infrastructure projects in San Francisco require that at least 50% of their job hours to go to San Francisco residents. I achieve this moving barrier in the model by setting the moving cost to be infinite ($c(a, j) = +\infty$) so that the two labor markets are totally disconnected, which is referred as "Autarky case".

Figure 9 compares the ex-ante welfare across different types of workers in the "Baseline" and "Autarky" cases. In the "Autarky" case, a minimum wage increase in county 1 has no effect on the workers in county 2, because these two labor markets are totally segregated. Therefore, the welfare of worker in county 2 (green line) is a horizontal line in the "Autarky" case. The welfare in the "Baseline" case and in the "Autarky" case differ because of two effects. First, workers in the "Baseline" case have additional working opportunities from the neighboring county, which generate welfare gains for all types of workers. Secondly, the sorting of workers discourages firms from posting vacancies in county 2. This reduction of contact rates in county 2 has a negative effect on all workers, but particularly on lower type workers, because they are more concentrated in county 2 when m_1 increases. Taken together, welfare increases for everyone except for the low skill worker in county 2. When $m_1 > 10$, they would prefer to stay in the "Autarky" case to avoid the negative spillover effects.

1.6.3. Preempting local minimum wage laws (universal (federal) minimum wages vs. local minimum wages)

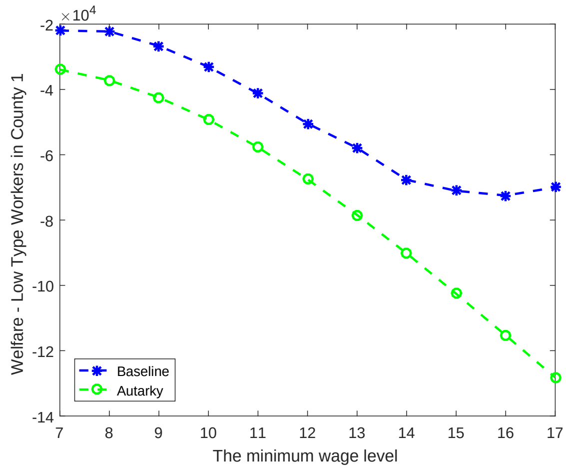
In this section, I perform a second counterfactual experiment: preempting local minimum wage laws. In reality, the preemption of local minimum wage laws is a popular policy intervention adopted by either federal or state legislatures to avoid a "patchwork" of minimum wage levels within their justification. Sometimes, progressive legislatures offer a statewide/nationwide raise to avoid more aggressive local level minimum wage changes. To understand the trade-off between universal level minimum wage hikes and local level minimum wage hikes, I consider the case in which both counties have an identical increase of their same minimum wages. Thus welfare changes of heterogeneous workers when setting a universal federal minimum wage at m is defined as

$$\Delta W_0(a, j; m, m) = V_u(a, j; m, m) - V_u(a, j; 7, 7)$$

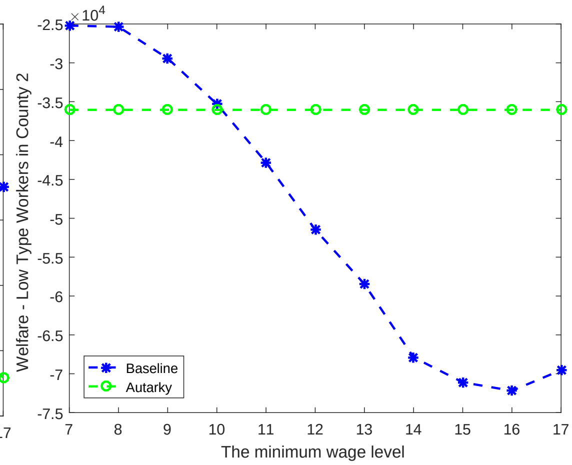
Figure 10 compares welfare changes under local minimum wage regulation and welfare changes under universal minimum wage regulation. Rather than keeping m_2 unchanged, a universal minimum wage policy equalizes the minimum wages in both counties, $m_1 = m_2$. Compared with the "Baseline" case, the increase of m_2 generates two offsetting effects. On

Figure 9: Change in worker welfare both in "Baseline" case and "Autarky" case

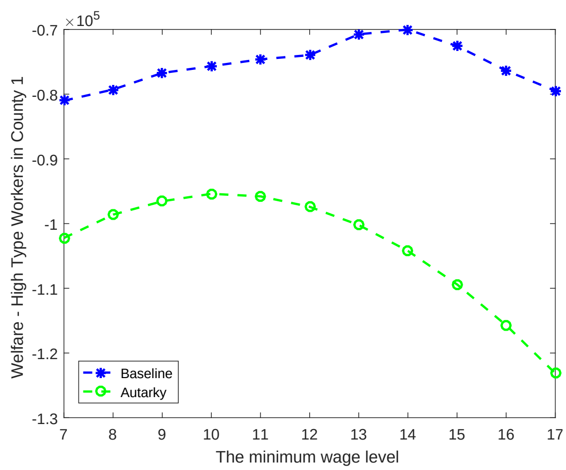
(a) Welfare - Low type in county 1



(b) Welfare - Low type in county 2



(c) Welfare - High type in county 1



(d) Welfare - High type in county 2

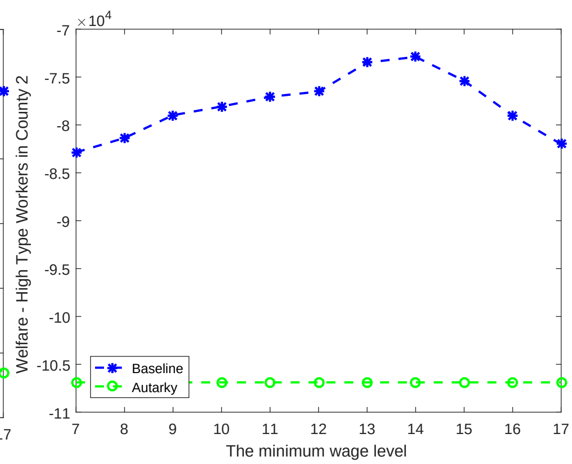
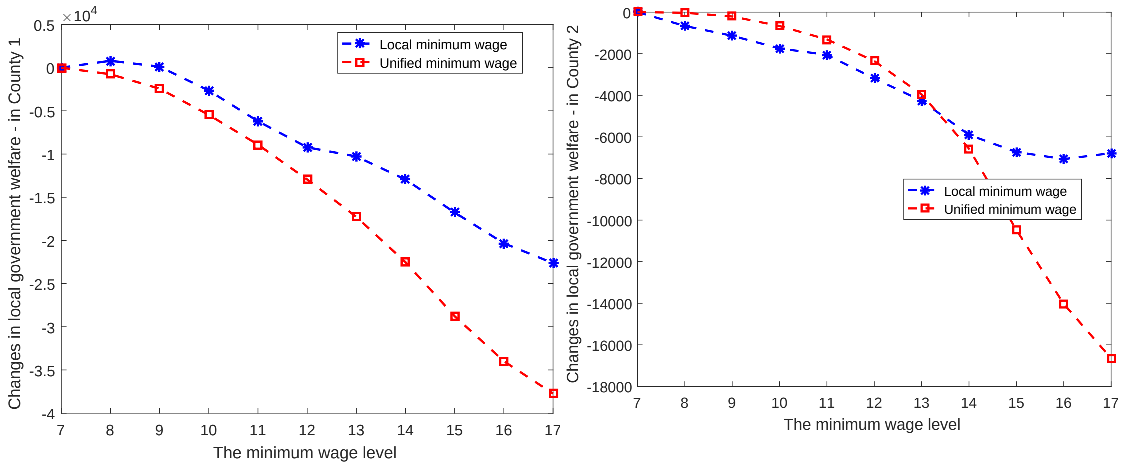


Figure 10: Changes in total welfare under local and under universal (federal) minimum wages

(a) Changes in local government welfare - county 1

(b) Changes in local government welfare - county 2

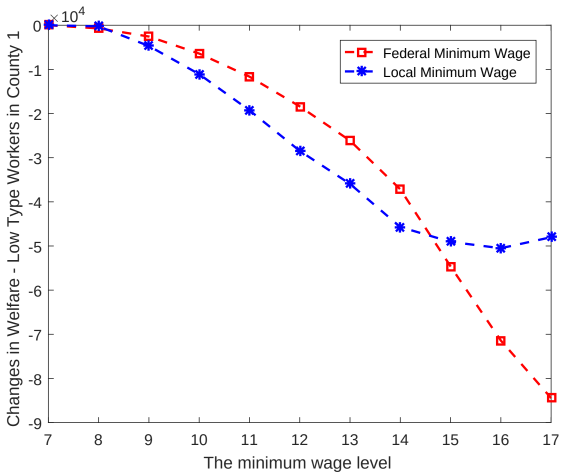


one hand, the minimum wage hikes in county 2 dissolves previously acceptable matches. On the other hand, the increase of m_2 prevents the sorting of workers between two counties, encouraging firms to post more vacancies. As shown in the right panel of Figure 10, the benefit of preventing negative spillovers dominates the cost of losing acceptable matches when $m < 13.5$. When minimum wage is not dramatically high, the total welfare in county 2 is actually higher under universal minimum wage policy. When the minimum wage exceeds \$13.5, the total welfare in both counties is reduced, because the loss of sustainable matches becomes the dominant effect.

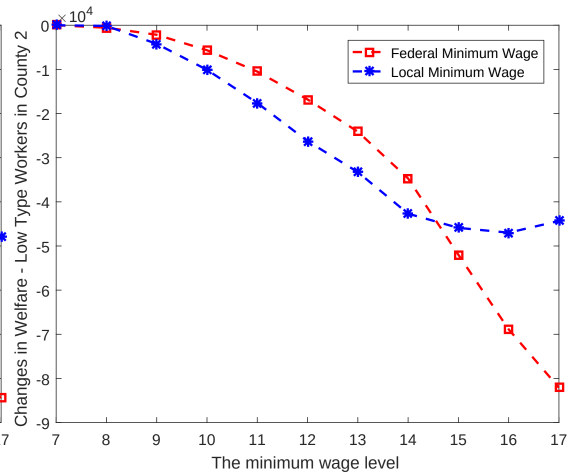
When decomposing total local welfare by worker types, I find preferred minimum wage regulation (universal vs. local) in county 2 is driven by low type workers. Thus, a planner that cares for low type workers should opt for universal rather than local minimum wage intervention when the change is moderate ($m <$ \$14.5). However, this welfare gain is accompanied with a welfare loss for high type workers.

Figure 11: Change in worker welfare under local and universal (federal) minimum wage changes

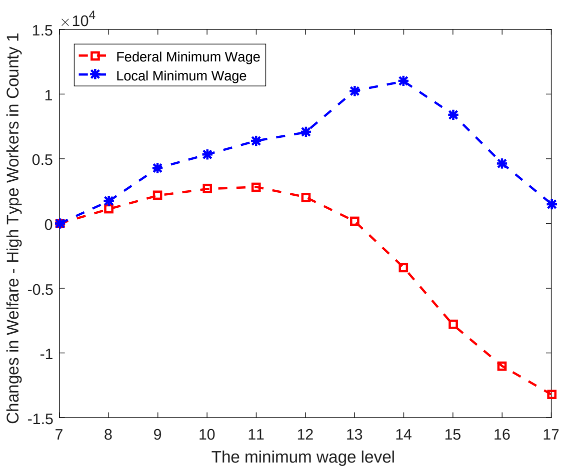
(a) Changes in welfare - low type in county 1



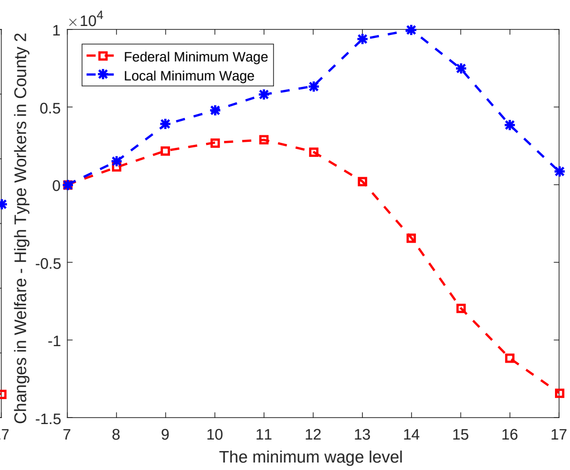
(b) Changes in welfare - low type in county 2



(c) Changes in welfare - high type in county 1



(d) Changes in welfare - high type in county 2



1.7. Conclusions

In this paper, I developed a spatial search model to study the effect of both local and universal (federal) minimum wage policies. In the model, firms endogenously choose where to post vacancies. Workers, differentiated by their type and location, engage in random search and can either accept a local job or migrate/commute to work in the neighboring county. My model captures three important effects associated with the minimum wage increases. First, conditional on being employed, a higher minimum wage shifts profits from firms to workers and increases workers' earnings. Second, a higher minimum wage also creates a disemployment effect by dissolving previously acceptable matches. This disemployment effect is more for low type worker. Third, firms reduce their vacancy postings in response to changing county-level worker composition and because they receive a smaller share of the matching surplus. Although the reduction in contact rates affects both counties, it has a larger effect on the neighboring county.

My analysis yields a number of interesting empirical findings when simulating the effects of minimum wage increases in county 1 with no change in county 2. First, minimum wage increases up to $14/hour increase the welfare of high type workers but lower the welfare of low type workers, leading to an increase in inequality. Minimum wage increases in excess of $14/hour lower the welfare of all workers, because the wage increases do not compensate for the disemployment effects. Second, the welfare of same type workers differs by locations ("lock-in effect") due to migration/commuting costs. Lastly, I find the disemployment effect of a minimum wage increase is underestimated if one ignores labor mobility. With the model, I obtain with the model a minimum wage elasticity of employment equal to -0.073; ignoring labor mobility cuts this value in half to -0.034. The bias is most severe for the counties with higher fractions of mobile workers.

I examine two counterfactual policies aiming for reducing spillover externalities: restricting labor mobility and preempting local minimum wage laws. In the experiment restricting labor mobility, the low type workers in neighboring county (the county without minimum wage change) prefer "Autarky" labor markets when the increase of local minimum wage is large ($m > \$10$). In the experiment of preempting local minimum wage laws, low type workers prefer a universal (federal) minimum wage rather than local minimum wages when the increa se of minimum wage is moderate ($m < \$14.5$). In contrast, the welfare of high type reduces unambiguously under both policies.

There are several ways to extend my analysis for future research. First, my model only compares the change between two steady states with minimum wage hikes. Adding transitional dynamics could capture the immediate effect of minimum wage hikes, which might differ from the long-term steady-state. Second, although I emphasize the worker selection and reallocation consequences of the local minimum wage policy, the local government is not a strategic player in my current model. Examining the competitive behavior of policy makers could be interesting. Third, local minimum wages also affect labor force participation. With higher minimum wages, individuals who were out of the labor force may also start to look for jobs in the labor market. This feature could be added into the model where government not only cares about the working population, but also the sub-population out of the labor force.

1.8. Appendix 1: equation expressions

1.8.1. Deducing the expressions of $V_u(a,j)$ and $V_e(w,a,j)$

I now consider individual's search problem

$$
\begin{aligned}
V_u(a,j) \quad &= (1+\rho\epsilon)^{-1}[\quad ab_j\epsilon + \lambda_j\epsilon \underbrace{\int_{m_j}^{\infty} \max\{V_e(w,j), V_u(a,j)\} dF(w|a,\theta,j)}_{\text{A local offer arrives}} \\
&+ \quad \lambda_{j'}\epsilon \underbrace{\int_{m_{j'}}^{\infty} \max\{V_e(w,j') - c(a,j), V_u(a,j)\} dF(w|a,\theta,j'))}_{\text{A neighbouring offer arrives}} \\
&+ \quad +(1-\lambda_j\epsilon - \lambda_{j'}\epsilon)V_u(a,j) + o(\epsilon)]
\end{aligned}
$$

Multiplying $1 + \rho\epsilon$ then subtracting $V_u(a, j)$ from both sides, I get

$$\begin{aligned}
\rho\epsilon V_u(a, j) &= ab_j\epsilon + \lambda_j\epsilon \underbrace{\int_{m_j}^{\infty} \max\{V_e(w, j), V_u(a, j)\} dF(w|a, \theta, j)}_{\text{A local offer arrives}} \\
&+ \lambda_{j'}\epsilon \underbrace{\int_{m_{j'}}^{\infty} \max\{V_e(w, j') - c(a, j), V_u(a, j)\} dF(w|a, \theta, j'))}_{\text{A neighbouring offer arrives}} \\
&+ -(\lambda_j\epsilon + \lambda_{j'}\epsilon)V_u(a, j) + o(\epsilon)
\end{aligned}$$

Dividing both sides by ϵ and taking limits $\epsilon \to 0$, I arrive at

$$\begin{aligned}
\rho V_u(a, j) &= ab_j + \lambda_j \underbrace{\int_{m_j}^{\infty} \{V_e(w, j) - V_u(a, j)\}^+ dF(w|a, \theta, j)}_{\text{A local offer arrives}} \\
&+ \lambda_{j'} \underbrace{\int_{m_{j'}}^{\infty} \{V_e(w, j') - c(a, j) - V_u(a, j)\}^+ dF(w|a, \theta, j'))}_{\text{A neighbouring offer arrives}}
\end{aligned}$$

The value of employment with wage w is

$$V_e(w, a, j) = (1 + \rho\epsilon)^{-1}\{w\epsilon + \eta_j\epsilon V_u(a, j) + (1 - \eta_j\epsilon)V_e(w, a, j) + o(\epsilon)\}$$

Multiplying $1 + \rho\epsilon$ then subtracting $V_e(a, j)$ from both sides, I get

$$\rho\epsilon V_e(w, a, j) = w\epsilon + \eta_j\epsilon V_u(a, j) - \eta_j\epsilon V_e(w, a, j) + o(\epsilon)$$

Dividing both sides by ϵ and taking limits $\epsilon \to 0$, I arrive at

$$V_e(w, a, j) = \frac{w + \eta_j V_u(a, j)}{\rho + \eta_j}$$

1.8.2. Solving for the bargained wage equation without the minimum wage constraint

Follow the same deduction procedure, the firm's value for a match with wage w, $V_t^f(w,a,\theta,j)$, is(I assume that the effective discount fact $\rho+\eta_j$ is the same as worker's):

$$V_f(w,a,\theta,j) = \frac{a\theta - w}{\rho+\eta_j}$$

Then the Nash bargaining $\hat{w}(\theta,a,j)$ without considering possible binding minimum wage is:

$$\begin{aligned}
\hat{w}(a,j,\theta) &= \arg\max_w (V_e(w,a,j) - V_u(a,j))^{1-\alpha_j} V_f(w,a,\theta,j)^{1-\alpha_j} \\
&= \arg\max_w (\tfrac{w+\eta_j V_u(a,j)}{\rho+\eta_j} - V_u(a,j))^{1-\alpha_j} (\tfrac{a\theta-w}{\rho+\eta_j})^{\alpha_j} \\
&= \arg\max_w (\tfrac{w-\rho V_u(a,j)}{\rho+\eta_j})^{1-\alpha_j} (\tfrac{a\theta-w}{\rho+\eta_j})^{\alpha_j} \\
&= \alpha_j a\theta + (1-\alpha)\rho V_u(a,j)
\end{aligned} \tag{1.13}$$

1.8.3. The derivation of fixed point system of $\theta^*(a,j)$ and $\theta^{**}(a,j)$

I start from the expression of unemployed value $V_u(a,j)$, equation 1.1:

$$\begin{aligned}
\rho V_u(a,j) &= ab_j + \underbrace{\lambda_j \int_{m_j}^{\infty} \{V_e(w,j) - V_u(a,j)\}^+ dF(w|a,\theta,j)}_{\text{A local offer arrives}} \\
&+ \underbrace{\lambda_{j'} \int_{m_{j'}}^{\infty} \{V_e(w,j') - c(a,j) - V_u(a,j)\}^+ dF(w|a,\theta,j'))}_{\text{A neighbouring offer arrives}}
\end{aligned}$$

Now, I replace the term $V_e(a,j,\theta)$ in the above equation using the following step-wise function:

$$V_e(a,j,\theta) = \begin{cases} \frac{m_j+\eta_j V_u(a,j)}{\rho+\eta_j} & \theta \in [m_j, \hat{\theta}(a,j)) \\ \frac{\alpha_j(a\theta-\rho V_u(a,j))}{\rho+\eta_j} + V_u(a,j) & \theta \in [\hat{\theta}(a,j), \infty) \end{cases}$$

Then I replace $\rho V_u(a,j)$ with its equivalent definition $a\theta^*(a,j)$ then get:

$$
\begin{aligned}
a\theta^*(a,j) &= ab_j + \frac{\lambda_j}{\rho+\eta_j}[\underbrace{\mathbb{I}\,(\theta^*(a,j) < \frac{m_j}{a})(m_j - a\theta^*(a,j))\left(\tilde{G}(\hat{\theta}(a,j)) - \tilde{G}(\frac{m_j}{a})\right)}_{\text{Local offer with wage } m_j} \\
&+ \underbrace{\int_{\max\{\hat{\theta}(a,j),\theta^*(a,j)\}} a\alpha_j(\theta - \theta^*(a,j))dG(\theta)]}_{\text{Local offer with wage } w_j > m_j} \\
&+ \frac{\lambda_{j'}}{\rho+\eta_{j'}}\underbrace{[\mathbb{I}(\theta^{**}(a,j) < \frac{m_{j'}}{a})(m_{j'} - a\theta^*(a,j'))\left(\tilde{G}(\theta^{**}(a,j)) - \tilde{G}(\frac{m_{j'}}{a})\right)}_{\text{Neighbouring offer with wage } m_{j'}} \\
&+ \underbrace{\int_{\max\{\hat{\theta}(a,j'),\theta^{**}(a,j)\}} a\alpha_j(\theta - \theta^*(a,j'))dG(\theta)}_{\text{Neighbouring offer with wage } w_{j'} > m_{j'}} \\
&+ \underbrace{(\rho+\eta_{j'})\,(\frac{a(\theta^*(a,j) - \theta^*(a,j'))}{\rho} + c(a,j))\tilde{G}(\theta^{**}(a,j))]}_{\text{The unemployed value difference between staying/moving}}
\end{aligned}
$$

1.9. Appendix 2: suggestive regression results

1.9.1. Both migrants and commuters are responsive to minimum wage hikes

This section presents the responses of migrants and commuters to minimum wage hikes. I find that low educated workers tend to commute/migrate away from states with higher relative minimum wage (compared to its neighboring state) rather than towards them. More specifically, the fraction of workers commuting out of the state increases and the number of individuals migrating into the local county from other states decreases.

I use the following regression to measure the effect of the relative minimum wage ratio on worker's migration and commuting behaviors:[35]

$$
\log y_{c,t} = \beta_0 + \beta_1 \log \frac{MW_{s(c),t}}{MW_{s'(c),t}} + \epsilon_{c,t} \tag{1.14}
$$

[35] Ideally, I would distinguish the effect of the own state's minimum wages from the effect of the neighboring state's minimum wages by using the following regression:

$$
\log y_{c.t} = \beta_0 + \beta_1 \log MW_{s(c),t} + \beta_1^{'} \log MW_{s'(c),t} + \epsilon_{c,t}
$$

However, due to the high correlation between $MW_{s(c),t}$ and $MW_{s'(c),t}$, the estimates suffer multicollinearity and become too sensitive to model specification. Therefore, I put the restriction $\beta_1 = -\beta_1^{'}$ to deliver more stable estimates.

Here $y_{c,t}$ is the ratio of migrants or commuters in county c, at time t. I estimate separate regressions for each education group. The minimum wage ratio $\frac{MW_{s(c)t}}{MW_{s'(c)t}}$ compares the minimum wage of $s(c)$, the state containing county c, to the minimum wage of $s'(c)$, the neighboring state of county c. The coefficient β_1 is the primary parameter of interest, which is the elasticity of outcomes y_{it} with respect to the relative minimum wage ratio.

Regression estimates are reported in Table 10. Column (1) reports the elasticity of the flows of migrants and commuters with respect to the change of relative minimum wage ratio. I use the relative minimum wage ratio rather than the absolute minimum wage levels to allow the flexibility that migration and commuting could be driven by either the own state's minimum wage hikes or the neighboring state's minimum wage increases. I find that minimum wage changes have a statistically significant negative effect for low educated migrants. In response to a 1% hike in the relative minimum wage ratio, the flows of low-educated migrants decrease by 0.539%. For commuters, these flows increase by 0.458% in response to a 1% increase in the relative minimum wage ratio.

However, observed commuting and migration changes could respond to other factors happening simultaneously with minimum wage increases. For example, if the local economic conditions are declining for the states with minimum wage increases, I would misattribute these changes to minimum wage changes instead of local economic conditions. Column (2) estimates the same regression model for high educated workers. If local conditions were underlying the observed changes of labor mobility, then high educated workers should present similar patterns, but that is not the case. There is no statistically significant migration response and only moderate commuting response to the same minimum wage increase.[36] While the evidence above does not prove causality, it is consistent with the view that minimum wage policy should have asymmetric effects on workers with different educational levels. Compared with low educated workers, the high educated group receives a higher

[36] I ran the same regression only for the high-school graduates, which are more closely related to high-school dropouts. The estimates are very close to the estimates for the whole high educated group. (This regression result is not reported in table 10)

Table 10: Migrant and Commuter Flows in Response to Minimum Wage Ratio Changes

	Baseline sample			*Restricted sample*		*Alternative*	*Extended sample*
y_{it}	*(1) Low education*	*(2) High education*	*(3) Whole sample*	*(4) Low education*	*(5) High education*	*(6) Whole sample*	*(7) Low education*
Migrants	**-0.589*****	-0.101	-0.093	**-0.682****	0.082	**-0.148*****	**-0.417*****
	(0.160)	(0.112)	(0.107)	(0.315)	(0.156)	(0.026)	(0.140)
	5,828	8,266	8,330	1,711	2,664	10,459	7,123
Commuters	**0.458****	**0.263****	**0.278****	**0.678***	**0.378****	**0.212*****	**0.442****
	(0.215)	(0.133)	(0.134)	(0.379)	(0.139)	(0.079)	(0.205)
	4,501	7,270	7,129	934	1,794	6,491	5,117
Controls							
Pair FEs	Y	Y	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y	Y	Y
Centriods <75mi	Y	Y	Y	Y	Y	Y	

The table reports coefficients associated with the log of relative minimum wage ratio ($\log \frac{MW_{st}}{MW_{st}}$) *on the log of the dependent variables noted in the first column. All regressions include both county fixed effects and year fixed effects. Columns (1)-(3) provide estimates for all individual between 16-30 based on pseudo county-level variation constructed by ACS PUMS between year 2005-2015. Column (6) uses IRS data. The variable "Migrants" is collected by the IRS Statistics of Income Division (SOI), year 05-15. The variable "Commuters" is extracted from the county-level ACS (09-15) through the interface called American FactFinder (web:https://factfinder.census.gov/) In Column (5)-(6), the sample is restricted to year 2007-2009 when the the Fair Minimum Wage Act of 2007 is enforced. For Column (7), the sample is extended to all county-pairs. Robust standard errors, in parentheses, are clustered at the the paired-county levels. * for 10%. ** for 5%, and *** for 1%. Sample sizes are reported below the standard error for each regression.*

wage on average, yielding a lower probability to be bound by minimum wage increases. Another concern is that the state-level minimum wage policy may move in tandem with other redistribution policies, such as unemployment insurance benefits, which may also cause asymmetric effects on workers with different levels of education. To minimize this concern, I restrict my sample to the period covered by The Fair Minimum Wage Act of 2007.[37] It is worth noting that the federal minimum wage compresses the minimum wage difference between contiguous counties. Therefore, the federal minimum wage should generate the opposite effect for states bound by the federal minimum wage: the commuting flows increase while the migration flows decrease. Columns (4) and (5) report values that are slightly higher (-0.682 and 0.678 compared to -0.589 and 0.458) than my baseline estimates for the low-education group, but not significantly different. The estimates for the high educated group are also similar to my baseline estimates. The elasticity of migration is not statistically significant and the elasticity of commuting is significantly positive but moderate in its magnitude. To sum up, my results are robust to the restricted sample only using the federal-level minimum wage variation, which supports the hypothesis that the potential endogenity of state-level minimum wage change does not bias the estimates.

Another concern is that pseudo county-based statistics may be imprecise. To mitigate this concern, I re-run the same regressions using different data sources in Column (6). The alternative migration data comes from the Internal Revenue Service (IRS) which collects the year-to-year address changes reported on individual income tax returns between 2005-2015.[38] The alternative commuting data comes from the 2009-2015 aggregated county-level ACS.[39] Unfortunately, these two alternative data sets lacks workers' demographic

[37]The Fair Minimum Wage Act of 2007 was implemented by three stages. Stage one increased the minimum wage from \$5.15 to \$5.85 in 2007. Stage two continued to increase it to \$6.55 in 2009. Then the final stage finalized the minimum wage in the level of \$7.25 in 2009. Thus I restrict my sample to year 2007-2009 to include the total effect of federal minimum wage change.

[38]IRS data is more robust than other data for a few reasons. First, IRS data covers 95 to 98 percent of the individual income tax filing population. Furthermore, the IRS and ACS display similar declines in migration after 2005, which is not true for other data such as the Panel Study of Income Dynamics (PSID), the Survey of Income and Program Participation (SIPP), and the Current Population Survey (CPS). A detailed discussion comparing different migration data sets can be found in Molloy et al. (2011).

[39]This is collected from the American FactFinder which only provides aggregate moments. Thus it is impossible to further disaggregate moments to get conditional ones on workers' characteristic. See

characteristics. Therefore, I can only compare estimates based on the full sample rather than estimates of subgroups classified by their education levels. The estimates using alternative data have similar values but different level of statistical significance. The elasticities of migration and commuting when using alternative data sets are -0.148 and 0.212 compared to my baseline estimates of -0.093 and 0.278.[40]

Lastly, I do another robustness check on the selection of contiguous county pairs. Following Dube et al. (2016), the baseline regression includes county pairs whose centriods are within 75 kilometers because the counties with closer centriods have more similar labor markets. In column (7), I run the same specification using all county pairs. Compared with the estimates using the baseline sample, the elasticities for low educated group are smaller but not different from my baseline estimates. This makes sense because as physical distance increases, workers have less incentive to take opportunities in the neighboring market since moving costs are higher. The regression results in this section suggest that low educated workers tend to move away from counties with minimum wage increases, either by commuting or migration.

1.9.2. The disemployment effect of local minimum wage hikes

In this section, I show additional evidence that the increase of outflows in response to a minimum wage increase is caused by the decline of local working opportunities. Following Dube et al. (2007) and Dube et al. (2016), I run the following regression:

$$\log y_{c,t} = \beta_0 + \beta_1 \log MW_{s(c),t} + \beta_2 X_{c,t} + \phi_c + \eta_{p(c),t} + \epsilon_{c,t} \tag{1.15}$$

where $y_{c,t}$ refers to the local labor market variables, including earnings, employment, separations and hires, in county c and period t. $X_{c,t}$ is the log of the total local population.

https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtmlfor details.

[40]The larger variance of my baseline estimates is due to the imputation process. One PUMA usually contains several counties, which washes away the inter-county variation when converting the PUMA-based statistics into the county-based statistics. Consequently, the “pseudo” county-level variation should be smaller than the “true” county-level variation, which results in less significant estimates.

Table 11: Minimum wage elasticity for employment stocks and flows

y_{it}	(1)	(2)
Hires	**-0.156*****	0.012
	(0.017)	(0.045)
	84,140	83,280
Separations	**-0.190*****	-0.024
	(0.017)	(0.022)
	84,120	83,246
Employment	**-0.068*****	**-0.039****
	(0.017)	(0.017)
	84,140	83,280
Earnings	**0.056*****	-0.016
	(0.015)	(0.015)
	84,140	83,280
Controls		
County fixed effect	Y	Y
Common time effects	Y	
Pair-specific time effects		Y
Centriods <75mi	Y	Y

*Data source: 2005-2015 Quarterly Workforce Indicator (QWI). This table reports the elasticity of the labor market outcomes listed in the first column. The regression sample is restricted to the counties from 964 county-pairs whose centriods are within 75 miles and includes all workers whose age is between 14-34. Robust standard errors, in parentheses, are clustered at the the paired-county level. * for 10%. ** for 5%, and *** for 1%.*

The coefficient β_1 is the primary variable of interest representing the elasticity of y_{it} with respect to the local minimum wages. Table 11 reports two regressions which only differ in their specification of the time-fixed effect. In Column (1), I restrict the time fixed effect to be common across all county pairs ($\eta_{p(c),t} = \eta_t$) and I find statistically significant disemployment effects in response to local minimum wage changes. The estimated elasticity of employment stock is -0.156. Meanwhile, the elasticities of employment flows are also substantial with minimum wage increases. The hire elasticity and separation elasticity are -0.190 and -0.156, both of which are statistically significant. The fact that the separation elasticity is larger than the hire elasticity is consistent with the negative effect of minimum wage on employment stock. However, when I account for the pair-specific time fixed effect (to control for time-varying, pair-specific spatial confounders), the estimates for the hire elasticity and separation elasticity are not distinguishable from zero. I attribute this change to the existence of spatial spillover effect. After the local county increases its own minimum wage, unemployed workers may seek their jobs in the neighboring county (either by migration or by commuting), which causes disemployment in the neighboring county. As a result, this spillover effect generates a common trend between the counties in one pair. When this pair-specific co-movement is teased out by pair-specific time effect, the estimates of local disemployment effect become less substantial.

1.10. Appendix 3: sample construction

1.10.1. Minimum wage policies between 2005-2015

In this section, I consider changes of minimum wage policies both on the state and federal level (See Table 12).[41] Between 2005 and 2015, there was only one change to federal minimum wage law, the Fair Minimum Wage Act of 2007.[42] While 78 changes in minimum wage resulted from the Act, the other 164 events were due to state ordinances. Table 12

[41] David et al. (2016) document all minimum wage law changes between 1979-2012. My table differs slightly from David et al. (2016) because I extend the sample through 2015 and include DC. Additionally, I have corrected errors in the minimum wages of Pennsylvania and Colorado.

[42] The Act raised the federal minimum wage in three stages: to \$5.85 60 days after enactment (2007-07-24), to \$6.55 one year after that (2008-07-24), then finally to \$7.25 one year after that (2009-07-24).

highlights two important patterns. First, at least 5 states change their effective minimum wage every year. Second, there is significant variation in how often states change their minimum wages. For example, Georgia only changed its minimum wage three times in line with federal minimum wage policy. On the contrary, its neighbor, Florida, makes the most minimum wage adjustments, changing 11 times.[43] Overall, the effective minimum wage increases $0.54 per change on average, but with substantial variation (Table 13). The largest change ($1.90) happened in Michigan in 2005, while the smallest increment ($0.04) happened in Florida in 2010.

One limitation is the scarcity of city-level minimum wage ordinances. Before 2012, only five localities had their own minimum wage laws. As of September 2017, 39 counties and cities have passed local minimum wage ordinances. Due to limited data, I evaluate the effect of county-level minimum wage indirectly. I estimate the baseline model using state-level minimum wage variation but focus on the resulting county-level labor market outcomes. Then, the effect of the county-level minimum wage will be inferred using contiguous border county pairs.

1.10.2. The raw ACS 2005-2015 PUMA database cleanup

First, I merge the three raw ACS 2005-2007, 2008-2010 and 2011-2015 data files into one that contains all the relevant variables between 2005-2015. The raw ACS files are downloaded directly from the US Census Bureau, following https://www.census.gov/programs-surveys/acs/data/pums.html. From year 2012, the ACS starts to use the 2010 version of Public Use Microdata Areas (PUMAs). Therefore, I further use the 2000-2010 PUMA crosswalk (https://usa.ipums.org/usa/volii/puma00_puma10_crosswalk_pop.shtml) to map the 2010 PUMA definitions to 2000 PUMA definitions for all the years after 2010. The variables obtained from the raw database are reported in Table 14. The wage measures are adjusted for inflation to be "2015 dollars" equivalent. I further put an age restriction $16 \leq age \leq 30$ on the population.

[43] Two changes happened in 2009.

Table 12: Variation in State Minimum Wages (2005-2015)

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Changes
Federal MW	5.15	5.15	5.15	5.85	6.55	7.25	7.25	7.25	7.25	7.25	7.25	3
Alabama												3
Alaska	7.15	7.15	7.15	7.15	7.15	7.75	7.75	7.75	7.75	7.75	8.75	3
Arizona			6.75	6.90	7.25		7.35	7.65	7.80	7.90	8.05	8
Arkansas			6.25	6.25							7.50	4
California	6.75	6.75	7.50	8.00	8.00	8.00	8.00	8.00	8.00	9.00	9.00	3
Colorado			6.85	7.02	7.28	7.28	7.36	7.64	7.78	8.00	8.23	8
Connecticut	7.10	7.40	7.65	7.65	8.00	8.25	8.25	8.25	8.25	8.70	9.15	6
Delaware	6.15	6.15	6.65	7.15	7.15					7.75	8.25	5
D.C.	6.60	7.00	7.00	7.00	7.55	8.25	8.25	8.25	8.25	9.50	10.5	7
Florida	6.15	6.40	6.67	6.79	7.21		7.31	7.67	7.79	7.93	8.05	11
Georgia												3
Hawaii	6.25	6.75	7.25	7.25	7.25							3
Idaho												3
Illinois	6.50	6.50	7.00	7.63	7.88	8.13	8.25	8.25	8.25	8.25	8.25	5
Indiana												3
Iowa			6.20	7.25	7.25							2
Kansas												3
Kentucky												3
Louisiana												3
Maine	6.35	6.50	6.75	7.00	7.25	7.50	7.50	7.50	7.50	7.50	7.50	5
Maryland		6.15	6.15	6.15							8.25	4
Massachusetts	6.75	6.75	7.50	8.00	8.00	8.00	8.00	8.00	8.00	8.00	9.00	3
Michigan			7.05	7.28	7.40	7.40	7.40	7.40	7.40	8.15	8.15	4
Minnesota		6.15	6.15	6.15						8.00	9.00	5
Mississippi												3
Missouri			6.50	6.65	7.05				7.35	7.50	7.65	7
Montana			6.15	6.25	6.90		7.35	7.65	7.80	7.90	8.05	10
Nebraska											8.00	4
Nevada			6.24	6.59	7.20	7.55	8.25	8.25	8.25	8.25	8.25	5
New Hampshire												3
New Jersey		6.15	7.15	7.15	7.15				7.25	8.25	8.38	5
New Mexico				6.50	7.50	7.50	7.50	7.50	7.50	7.50	7.50	4
New York	6.00	6.75	7.15	7.15	7.15					8.00	8.75	6
North Carolina			6.15	6.15								3
North Dakota												3
Ohio			6.85	7.00	7.30	7.30	7.40	7.70	7.85	7.95	8.10	8
Oklahoma												3
Oregon	7.25	7.50	7.80	7.95	8.40	8.40	8.50	8.80	8.95	9.10	9.25	10
Pennsylvania			6.70	7.15	7.15							6
Rhode island	6.75	7.10	7.40	7.40	7.40	7.40	7.40	7.40	7.75	8.00	9.00	5
South Carolina												3
South Dakota											8.50	4
Tennessee												3
Texas												3
Utah												3
Vermont	7.00	7.25	7.53	7.68	8.06	8.06	8.15	8.46	8.60	8.73	9.15	10
Virginia												3
Washington	7.35	7.63	7.93	8.07	8.55	8.55	8.67	9.04	9.19	9.32	9.47	10
West Virginia			6.20	6.90	7.25						8.00	4
Wisconsin	5.70	6.50	6.50	6.50								4
Wyoming												3
Changes	12	17	47	45	47	5	9	8	10	18	24	242

Note: Two minimum wage changes happened in 2009 for Florida.

Table 13: Summary Statistics of State-Level Effective Minimum Wage Changes (2005-2015)

Year	Counts	Mean	S.D.	Min	Max
2005	12	0.621	0.475	0.10	1.45
2006	17	0.605	0.463	0.15	1.85
2007	47	0.831	0.527	0.25	1.90
2008	45	0.541	0.285	0.10	1.35
2009	47	0.533	0.206	0.05	1.00
2010	5	0.548	0.234	0.04	0.70
2011	9	0.160	0.190	0.06	0.70
2012	8	0.315	0.032	0.28	0.37
2013	10	0.160	0.068	0.10	0.35
2014	18	0.362	0.321	0.10	1.00
2015	24	0.629	0.467	0.12	1.85
Total	212	0.538	0.370	0.04	1.90

Note: All units are in nominal dollars.

Table 14: Variables obtained from the raw ACS

Variables	Variable labels
serialno	Housing unit/GQ person serial number
puma	Public use microdata area code
st	State code
adjinc	Adjustment factor for income and earnings dollar amounts
agep	Age
pwgtp	Person's weight replicate
migpuma	Migration PUMA
migsp	Migration recode - state or foreign country code
powpuma	Place of work PUMA
powsp	Place of work - State or foreign country recode
schl	Educational attainment
esr	Employment status recode
wagp	Wages or salary income past 12 months
wkhp	Usual hours worked per week past 12 months
wkw	Weeks worked during past 12 months

Table 15: Converting individual-level observations to county-level moments

Individual-level variables	County-level variables	Definition	RAW ACS
High type dummy	High type fraction	Education attainment is high school graduate or above	schl
Low type dummy	Low type fraction	Education attainment is high school dropouts	schl
Employment dummy	Employment rate by types (high and low)	(1) Employed at work and (2) employed with a job but not at work	esr
Hourly wage	Average hourly wage by types (high and low)	"Wages or salary income past 12 months"(wagp) divided by the product of "usual hours worked per week past 12 months"(wkhp) and "weeks worked during past 12 months"(wkw)	wagp, wkhp, wkw
Migrants dummy	The fraction of migrants by types (high and low)	Individuals who report a migration states (not N/A)	migsp
Commuters dummy	The fraction of commuters by types (high and low)	Individuals who report the place of work different from the place of residence	powsp
Labor force dummy	Labor force participation rate by type (high and low)	(1) Employed at work, (2) employed with a job but not at work and (3) unemployed	esr

Next, I convert the individual-level observations into county-level moments, reported in Table 15. The biggest challenge in this process is that the basic geographic units for respondents in ACS is "Public Use Micro Areas"(PUMAs) rather than any jurisdiction geographic entity (i.e. county, city, etc.) in order to comply with census non-identifiable disclosure rule. Therefore, I instead construct the "pseudo" county-level statistics by the following two steps: (1) First, I construct the PUMA-level summary statistics from the corresponding individual-level variables. (2) Second, I impute the county-based measures from the corresponding PUMA-based measures following the crosswalk provided by Michigan Population Studies Center http://www.psc.isr.umich.edu/dis/census/Features/puma2cnty/. The new constructed county-level variables are reported in second column in Table 15, while the original individual-level variables are displayed in first column.

Finally, I label the adjacent counties on the state borderline, consistent with the classification showed in figure 1. Table 2 and the second panel in table 3 report conditional statistics

Table 16: County-level moments obtained from QWI

Variables	Definition	Raw QWI
Average monthly earnings	Average monthly earnings of employees who worked on the first day of the reference quarter.	EarnBeg
Employment	Estimate of the total number of jobs on the first day of the reference quarter.	Emp
Hire rate	The number of workers who started a new job at any point of the specific quarter as a share of employment	HirA/Emp
Separation rate	The number of workers whose job in the previous quarter continued and ended in the given quarter	SepBeg/Emp

both by educational types and by interior/borderline locations.

1.10.3. The raw QWI 2005Q1-2015Q4 database cleanup

The time series of county-level variables from QWI are directly obtained through LED extraction tool https://ledextract.ces.census.gov/static/data.html. The age group 19-21, 22-24, 25-34 are selected. The variables displayed in table 16 are calculated and used in this paper.

1.10.4. Creating the merged sample using multiple data sources

In this session, I will report the final step to merge multiple data sources together into the final completed sample. First, I will use QWI as the baseline data sample. Second, I will merge the ACS into QWI. Third, I will further merge other county-level moments from several different data sources.

- **Step 1: build the baseline data structure with QWI variables.** I create a balanced panel of all contiguous county-pairs with quarterly frequency between 2005Q1-2015Q4. (Obs. 43,596) Then I only keep the observations when one of the two counties changes its minimum wage at quarter t and the information for the following quarter $t + 1$ is still completed. (Obs. 3,278) I only keep one quarter observation if minimum wage changes multiple times in one year. (Obs. 2,886)

Table 17: Key moments from other data sources

Variables	Definition	Data source	Year	Completeness
Bargaining power α_j	The annual payroll expenditure in account for the employer value of sales, shipments, receipts, revenue, or business done in restaurant industry (NAICS:722)	Economy Wide Key Statistics (EWKS)	2012	2,243
Matching technology ω_j	The number of divided by the number of total ads and reflects the latest month for which unemployment data is available	The Conference board Help Wanted OnLine (HWOL)	2017.4	2,306
The centroid distance $d_{jj'}$	-	Dube et al. (2010).	-	2,314
The local amenity γ_j	Median gross rent	2011-2015 American Community Survey	2012	2,314

- **Step 2: merge with the pseudo county-level ACS 05-15 variables.** I merge the ACS into QWI using the indicator combining county-pair and year. I use the QWI data from step 1 as the master file for the merge. Then I only keep all the observation with positive shares of both migrants and commuters. (Obs. 2,314)

- **Step 3: merge additional other variables from several different databases.** I merge several key variables from other data sources which are displayed in the following table. The final sample covers 2,243 observations with all variables completed.

CHAPTER 2 : A Dynamic Model of Personality, Schooling, and Occupational Choice

Petra Todd and Weilong Zhang

2.1. Introduction

It has long been recognized that cognitive skills are important determinants of labor market success, but there is increasing evidence that noncognitive skills also play a salient role.(Becker (1964); Griliches (1977)) For example, using data from the Perry Preschool randomized experiment, Heckman et al. (2010) find that the ability to plan and to exert self-control significantly affects lifetime earnings and employment. Devising effective social policies that maximize the potential for human development requires an understanding of the mechanisms through which cognitive and noncognitive skills evolve and influence individuals' education and labor market trajectories.

This paper develops and estimates a dynamic model of schooling, work, and occupational choices that incorporates noncognitive personality traits, as measured by the so-called "Big Five." Our model allows both cognitive and noncognitive traits to influence educational and labor market outcomes through multiple channels, by affecting pecuniary or nonpecuniary returns from schooling and by affecting the reward from choosing white or blue collar occupations. Our analysis is inspired in part by the pioneering work of Keane and Wolpin (1997) that estimates a similar type of model without personality traits.

A key finding from Keane and Wolpin (1997) analysis is that 90 percent of the total variance in expected lifetime utility is explained by unobserved skill endowments at age 16. The importance of unobserved heterogeneity in explaining educational and labor market outcomes has also been emphasized in other studies. For example, Yamaguchi (2012) finds that endowment differences prior to labor market entry account for 70% of the log-wage variance in the first year and 35% after 20 years. Sullivan (2010) finds that 56% of the variance in discounted expected lifetime utility is explained by initial heterogeneity. Huggett et al.

(2011) conclude that 61.5 percent of the variation in lifetime earnings and 64.0 percent of the variation in lifetime utility is attributable to initial conditions.

Although accumulated evidence clearly points to the importance of endowment heterogeneity in explaining educational and labor market trajectories, its precise components remain unclear. Keane and Wolpin (1997) find that family background accounts for less than 10 percent of the total variation in lifetime utility and that adding cognitive ability only increases the explained variation to 14 percent. Prior studies have not considered the potential role of personality traits as a component of endowment heterogeneity because the datasets typically used do not include personality trait measurements.

In the psychology literature, personality traits have been shown to be correlated with many aspects of individuals' lives. However, study of their effects on economic outcomes is relatively scarce. (Almlund et al. (2011)) The five-factor model (so called "Big-five") is the most widely adopted measurement of personality in both psychology (Goldberg (1992a);Saucier (1994);Gosling et al. (2003)) and economics (Borghans et al. (2008)). The Big Five traits include openness to experience, conscientiousness, extraversion, agreeableness and neuroticism (OCEAN). The meaning of these traits and their determination will be further described below.

The dynamic model of schooling, work and occupational choices that we estimate assumes that individuals make one of four mutually exclusive choices: attending school, staying home, working in a white-collar job or working in a blue-collar job from ages 15 to 58. Individual endowments at age 15 consist of personality traits, cognitive ability, and family background characteristics, which include parental schooling, number of siblings, sibling order and whether the person lived with both parents at age 14. To allow for unobserved heterogeneity in a tractable way, we assume each individual is one of four types (denoted I-IV). An individual's type potentially affects their pecuniary and nonpecuniary reward from choosing particular schooling or work options. In the dynamic discrete choice literature, it is common to assume unobserved types are fixed over time (e.g. Keane and Wolpin (1997), Yamaguchi (2012), Sullivan (2010)). Our model begins at age 15. At this age, it has been shown that the ranking of cognitive ability is relatively stable. However, personality traits still evolve until the mid 30s. We incorporate personality traits into our model in a parsimonious way as a determinant of the unobserved types and we allow the distribution of unobservable heterogeneity to change over time. We therefore allow the unobserved types, which depend in part on personality traits, to change with age. We implement a likelihood ratio test for type stability over the life-cycle, which we strongly reject in our data.

The model is estimated using the Household Income and Labour Dynamics in Australia (HILDA) longitudinal data set, waves 1(2001) through 13(2013). The data have repeated measures of the Big-Five personality traits as well as measures of cognitive ability. Our estimation results show that the unobserved types are malleable, particularly at early ages. At age 15, individuals have on average a 40% probability to change type, but by age 36 their type stabilizes. Our results are broadly consistent with findings from some psychology studies on personality trait stability. For example, Terracciano et al. (2006) and Terracciano et al. (2010) report that intra-individual stability increases up to age 30 and thereafter stabilizes.

We use the estimated model to evaluate two education policies: making senior secondary school compulsory and providing a 50% cost subsidy to attend college. Both policies provide incentives to enroll in school but differ in their distributional implications. We find that individuals belonging to types I and IV have a comparative advantage in education and receive the most benefit from the college subsidy policy. Their average number of years of completed education increases by around one year, in comparison to half a year on average for types II and III. In contrast, the impacts of compulsory senior second school are concentrated on types II and III, who tend to come from lower SES backgrounds. The average increase in years of education is around one-half for these two types but close to zero for the other two types. Thus, the two policies both increase average years of education but have very different distributional effects.

To study the relevance of personality traits in assessing impacts of these educational policies, we also estimate a model with fixed types. When this channel is shut down, we find that there is less incentive for disadvantaged groups to pursue education, because they no longer have the potential to alter their disadvantaged types. The increase in annual earnings attributable to the policy intervention is significantly smaller in the fixed type model. In other words, the inequality in the distribution of the policy effect is significantly overstated in the restricted fixed type model.

This paper is organized as follows: Section II reviews the literature. Section III describes the HILDA data and the Big-Five measures. Section IV describes the general structure of our model and its econometric implementation. Section V discusses the identification strategy and estimation method. Section VI explains our estimation strategy. Section VII presents the estimation results and provides information for the goodness of model fit. The model implications are discussed in section VIII. Section IX reports results from the two policy experiments and section X concludes.

2.2. Related Literature

The "Big Five" personality traits are defined as follows: (1) extraversion: an orientation of one's interests and energies toward the outer world of people and things rather than the inner world of subjective experience; characterized by positive affect and sociability, (2) neuroticism: a chronic level of emotional instability and proneness to psychological distress. Emotional stability is predictability and consistency in emotional reactions, with absence of rapid mood changes, (3) openness to experience/intellect: the tendency to be open to new esthetic, cultural, or intellectual experiences, (4) conscientiousness: the tendency to be organized, responsible, and hardworking and (5) agreeableness: the tendency to act in a cooperative, unselfish manner.

Several studies examine the influence of personality traits on wage performance and occupational choices. For example, both Nyhus and Pons (2005) and Salgado (1997) find that emotional stability and conscientiousness are strongly correlated with wage and job performance. Cubel et al. (2016) examine whether Big Five personality traits affect productivity using data gathered in a laboratory setting where the task effort is directly measurable. They find that individuals who exhibit high levels of conscientiousness and higher emotional stability perform better on the task. Fletcher (2013) uses data on siblings and finds a robust relationship between personality traits and wages using sibling samples. Specifically, conscientiousness, emotional stability, extraversion and openness to experience were all found to positively affect wages. There are few papers that examine the correlation between personality traits and educational attainment. Personality traits are predictive of educational attainment. For example, Lundberg (2013) finds positive correlations between personality traits (such as conscientiousness, agreeableness and openness to experience) and college entrance. However, personality traits may also be changed by education experience. Dahmann and Anger (2014) and Schurer et al. (2015) note that educational experiences in secondary school and at university shape students' personality.

Our paper is also related to the burgeoning literature examining the process of noncognitive skill formation. Heckman et al. (2006) study the effect of non-cognitive skills on schooling decisions and subsequent labour market outcomes, allowing schooling and family background to influence be potential determinants of skill formation. Cunha and Heckman (2008) estimate a linear dynamic model to study the formation of cognitive and non-cognitive skill as it depends on parental investment. Heckman and Raut (2016) formulate a dynamic structural model that integrates preschool investment choices that affect skill formation with schooling and earning outcomes later in life.

2.3. Data

The analysis is based on a sample of individuals from the Household Income and Labour Dynamics in Australia (HILDA) longitudinal data set. HILDA is a representative one in one thousand sample of the Australian population. It is an ongoing annual dataset starting from the year 2001 with 19,914 initial individuals from 7,682 households. (Summerfield et al.

(2014)) Our paper makes use of the following variables: (1) labor market outcomes including occupational information (coded following the ANZSCO system[1]), annual labor earnings and working hours; (2) family background information including parental education levels, sibling number and order as well as measures of household intactness; (3) education levels ranging from senior secondary school until the highest degree; (4) cognitive ability measured in wave 12; and (5) the "Big-five" personality traits assessment repeatedly collected in wave 5, 9 and 13.

Table 18: Definitions and examples of the ANZSCO coding of occupations

Collar	Occupations	Examples
White Collar	Managers	Legislators, senior officials Corporate/general managers
	Professionals	Professionals, Physician, mathematician, Engineer and life science.
	Technicians and tradespersons	Technicians and associate professionals, Physical and engineering scientists, Life science and health association
Blue Collar	Community and personal service workers	Office clerks, Customer service clerks
	Clerical and administrative workers	Service workers and shop workers, Personal and protective service workers Models, salespersons
	Sales workers	Sales representative, insurance brokers, checkout operator, models and telemarketers,
	Machinery operators and drivers	Industrial spraypainter, sewing machinist, motion picture projectionist, crane operator, forklift driver, and train driver
	Labourers	Cleaners, steel fixer, product assembler, packer, slaughter, farm worker, kitchen hand, freight handler and handypersons

To the best of our knowledge, HILDA has the best quality measures of personality traits among all nationwide data sets. For the majority of respondents, we observe three repeated measurements of personality traits over an eight-years time window.[2] HILDA's "Big-Five"

[1]In practice, we classify all occupations into two categories: blue-collar job and white-collar job. White collar jobs includes managers, professionals, technicians and tradesperson. Blue collar jobs include community and personal service workers, clerical and administrative workers, sales workers, machinery operators and drivers as well as labourers. See table 18 for details.

[2]One alternative national-wide data set providing personality traits inventory assessment is German

Table 19: The survey illustration of personality questionnaire

B19 How well do the following words describe you? For each word, cross one box to indicate how well that word describes you. There are no right or wrong answers.

*(Cross ☒ **one** box for **each** word.)*

Does not describe me at all — *Describes me very well*

	1	2	3	4	5	6	7
talkative	1	2	3	4	5	6	7
sympathetic	1	2	3	4	5	6	7
orderly	1	2	3	4	5	6	7
envious	1	2	3	4	5	6	7
deep	1	2	3	4	5	6	7
withdrawn	1	2	3	4	5	6	7
harsh	1	2	3	4	5	6	7
systematic	1	2	3	4	5	6	7
moody	1	2	3	4	5	6	7
philosophical	1	2	3	4	5	6	7
bashful	1	2	3	4	5	6	7
kind	1	2	3	4	5	6	7
inefficient	1	2	3	4	5	6	7
touchy	1	2	3	4	5	6	7
creative	1	2	3	4	5	6	7
quiet	1	2	3	4	5	6	7
cooperative	1	2	3	4	5	6	7
sloppy	1	2	3	4	5	6	7

Does not describe me at all — *Describes me very well*

	1	2	3	4	5	6	7
jealous	1	2	3	4	5	6	7
intellectual	1	2	3	4	5	6	7
extroverted	1	2	3	4	5	6	7
cold	1	2	3	4	5	6	7
disorganised	1	2	3	4	5	6	7
temperamental	1	2	3	4	5	6	7
complex	1	2	3	4	5	6	7
shy	1	2	3	4	5	6	7
warm	1	2	3	4	5	6	7
efficient	1	2	3	4	5	6	7
fretful	1	2	3	4	5	6	7
imaginative	1	2	3	4	5	6	7
enthusiastic	1	2	3	4	5	6	7
selfish	1	2	3	4	5	6	7
careless	1	2	3	4	5	6	7
calm	1	2	3	4	5	6	7
traditional	1	2	3	4	5	6	7
lively	1	2	3	4	5	6	7

information is based on 36 personality questions.(table 19) Respondents were asked to pick a number between 1 to 7 to assess how well each personality adjective describes them. The lowest number 1 denotes a total opposite description and the highest number 7 denotes a perfect description. According to Losoncz (2009), only 28 of 36 items load well into their corresponding components when performing factor analysis. The other 8 items are discarded due to either their low loading value or their ambiguity on several traits.[3] The construction of "Big-five" in our paper follows the approach of Losoncz (2009). Big-five personality traits are available for 4,938 males aged 15-58 interviewed in wave 5. The traits are recorded for is 5,048 and 6,771 respondents in wave 9 and wave 13, respectively. We include all individuals who have at least one measure of personality traits in our estimation sample.[4]

Cognitive ability is only surveyed once in wave 12.[5] We construct a one-dimensional measure of cognitive ability from three different measurements: (i) Backward Digits Span, (ii) Symbol Digits Modalities and (iii) a 25-item version of the National Adult Reading Test.

2.3.1. Additional background variables and sample restrictions

In addition to the cognitive and noncognitive trait measures described above, we use the following information in our analysis: sibling information (including whether the person has siblings, whether he is the eldest child in the family and how many siblings), an indicator of growing up in an intact family, parental education, and parental working information.[6] We also include state of residence and cohort information.

Our estimation focuses on males between age 15-58. Women are not included to avoid additional complication of modeling marriage and fertility decisions, which may impact

---

Socio-Economic Panel (GSOEP) study. GSOEP also surveys "Big-five" three times in years 2005, 2009 and 2013.

[3] The way to check each item' loading performance is to calculate the loading value after doing oblimin rotation. The loading values of 8 abandoned items were either lower than 0.45, or did not load more than 1.25 times higher on the expected factor than any other factor.(Losoncz (2009))

[4] A detailed comparison of the personality traits measurement between three waves will be provided in section 2.3.3.

[5] According to the report of Wooden (2013), the response rate is high, approximately 93%.

[6] All the parental questions are conditional on the situation when the respondent was at the age of 14.

schooling and labor supply decisions to a larger extent. We exclude the persons age 58 or older because many people are retired by that age. Individuals serving in the military are also dropped. Lastly, we drop person-year observations that are missing information on the state space variables in our model. The remaining sample has 36,639 observations from 4,215 individuals in total.

Selected summary statistics of individual's characteristics are reported in table 20. Our sample is distributed across eight states and territories.[7] 83.4% of individuals report residing in an intact family at the age of 14, whereas 7.34% of individuals lived only with their mothers at that age. The majority (96.37%) have siblings. The cases of one, two, three and four siblings account for 25.63%, 30.71%, 18.20% and 9.94% of the total sample. About one-third of the individuals (34.32%) are the eldest child in the family. Table 20 also provides statistics on parental education and occupations when the individual was age 14. 57.98% of fathers and 37.78% of mothers have a college degree. Most fathers were employed, but only about half the sample had working mothers. Almost two-thirds of fathers' jobs were in white-collar occupations. Half of the working mothers worked in blue-collar jobs.

2.3.2. Educational and occupational choices over life cycle

During the survey, individuals report both school enrollment and employment information annually. Details include the desired education level and whether they eventually reach this level.[8] The employment information includes employment status, working hours, total annual earnings and occupational codes.

Figure 12 shows the choice distribution of schooling, staying at home, blue collar jobs and white collar jobs by age. At age 15, about 80% are enrolled in school but after age 17, this fraction drops sharply to around 30%. The majority of secondary school graduates choose to work immediately rather than to continue their tertiary education. The enrollment rate

[7] They are Australian Capital Territory(ACT), New South Wales(NSW), Victoria(VIC), Queensland(QLD), South Australia(SA), Western Australia(WA), Tasmania(TAS) and Northern Territory(NT).

[8] A rough classification of the tertiary education certificates includes 1. Certificates I-IV; 2. Diploma, Advanced Diploma, Associate Degree; 3. Bachelor degree and honors; 4. Graduate Certificate and Graduate Diploma; 5. Master degree; 6. Doctoral degree.

Table 20: Sample summary statistics

Variable	Proportion	Variable	Proportion
Geographic Information		Parental Information	
State		***Father Education***	
NSW	0.3125	College	0.5798
VIC	0.2496	Not College	0.4202
QLD	0.2009	***Mother Education***	
SA	0.0928	College	0.3778
WA	0.0871	Not College	0.6222
TAS	0.0275	***Father Working***	
NT	0.0057	Employed	0.9558
ACT	0.0240	Not Employed	0.0209
Family background		Deceased	0.0233
Family Intactness		***Father Occupation***	
Both parents	0.8341	White Collar	0.6485
Father and step	0.0107	Blue Collar	0.3515
Mother and step	0.0427	***Mother Working***	
Father only	0.0233	Employed	0.5488
Mother only	0.0734	Not Employed	0.4139
Other	0.0158	Deceased	0.0720
Sibling Info		Not Asked	0.0302
Sibling dummy		***Mother Occupation***	
Has siblings	0.9637	Not Asked	0.2113
No siblings	0.0373	White Collar	0.2889
Sibling numbers		Blue Collar	0.4990
Not Asked	0.0379	Cohort Information	
1	0.2563	***Year***	
2	0.3071	1940-1949	0.1038
3	0.1820	1950-1959	0.1919
4	0.0994	1960-1969	0.2358
5 or more	0.1173	1970-1979	0.1913
Eldest Sibling		1980-1989	0.1686
Not Asked	0.0373	1990-	0.1040
Oldest	0.3432		
Not Oldest	0.6195	Total Individuals	4215

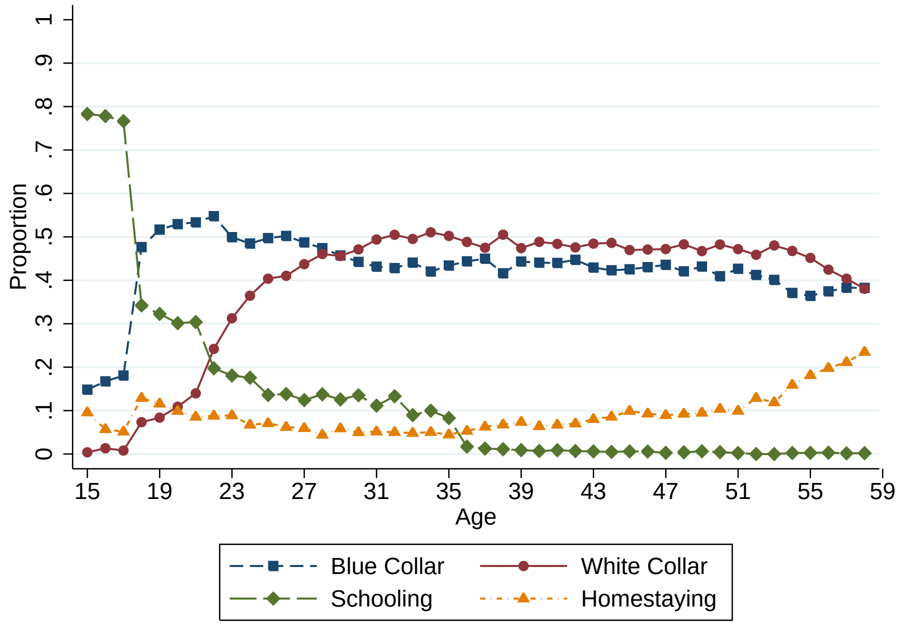


Figure 12: Work status and college attendance by age(% of the sample)

keeps decreasing from 19% at age 23 to around 9% at age 34.

An individual is defined to be "working" if reported to be working positive hours and not enrolled in school. An individual is defined to be "staying home" if he is neither working nor in school.[9] The blue-collar participation rate decreases monotonically from around 50% at age 18 to around 38% at age 58. The significant increase of the white-collar participation rate between ages 22 to 25 suggests that a college degree is a prerequisite for many white-collar occupations. The white collar participation rate continues to increase after age 26, as some workers switch from blue-collar job to white-collar jobs over time. After age 53, the option of staying home becomes more prevalent, reflecting the retirement decisions of participants.

Figure 13 reports the age-earnings profile by two occupations, between ages 18 to 58.[10] Both

[9]We do not distinguish between being unemployed and being out of labor force, as the decision to be unemployed is always considered voluntary under our model.

[10]We drop the wage observations between age 15 and age 17 because of two reasons. 1. the observations are few. 2. A Large fraction of this group are senior school students who only do some part-time jobs. Thus their choices is classified as schooling according to our definition.

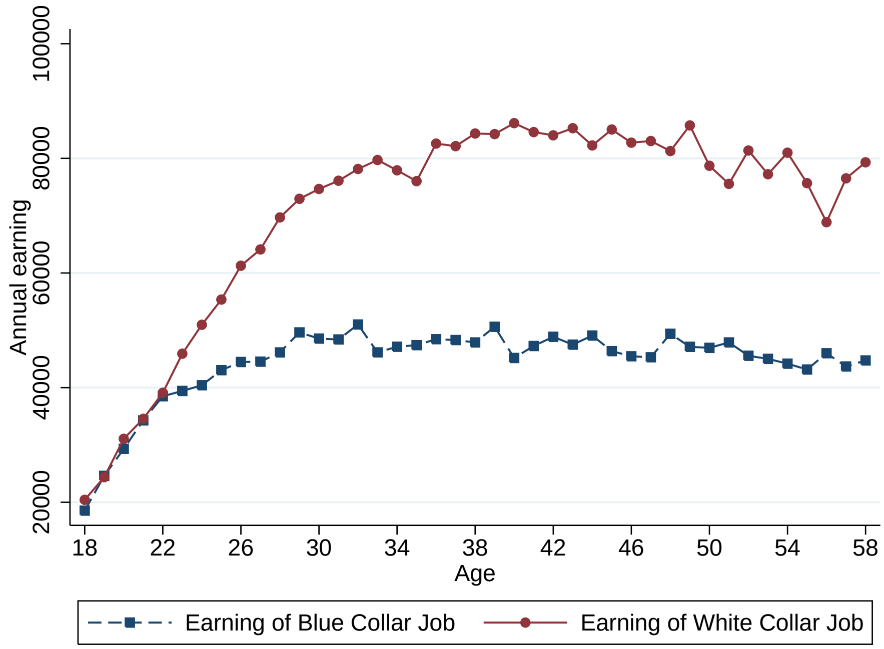


Figure 13: Average wage profile by occupations over life cycle

the white-collar and blue-collar earning profiles exhibit a hump shape, overall. Prior to age 24, earnings of white-collar and blue-collar workers are similar. Subsequently, however, the shape of the blue-collar earnings profile becomes flatter and then stops growing after age 28. The white-collar earnings profile keeps increasing until the mid-30s. Peak average earnings from blue-collar jobs is around AU$48,000, whereas the peak from white-collar jobs is around AU$85,000. Earnings in both sectors decrease slightly at older ages. Data on personality traits are gathered in 2005, 2009 and 2013. Table 21 reports the average personality trait scores for three different educational levels: senior secondary school or lower, college dropouts and college graduates. In general, the group with higher educational attainment has higher scores of emotional stability, openness to experience, conscientiousness, and agreeableness. However, this group tends to be less extraverted. Table 22 reports the difference in personality traits between workers in white-collar and blue-collar jobs. Workers in white-collar occupations are more likely to be emotionally stable, open to experience, and conscientious, but are less extraverted. The most significant differences are seen in conscientiousness and openness to experience.

Table 21: Average personality traits by educational level

Occupation	Emotional Stability	Openness to experience	Conscientiousness	Agreeableness	Extroversion
High School	-0.0478	-0.1414	-.0784	-0.0508	0.0393
or Lower	(0.0140)	(0.0139)	(0.0138)	(0.0141)	(0.0133)
College	0.0258	0.0605	0.1033	0.0765	-0.0056
Dropouts	(0.0354)	(0.0338)	(0.0349)	(0.0345)	(0.0358)
College	0.1043	0.3096	0.1430	0.0839	-0.0997
Graduates	(0.0208)	(0.0202)	(0.0217)	(0.0208)	(0.0232)

Note: Each personality trait was standardized to have mean 0, variance 1.
Source: HILDA, waves 5, 9 and 13.

Table 22: Average personality traits by occupation category

Occupation	Emotional Stability	Openness to experience	Conscientiousness	Agreeableness	Extroversion
Blue-collar	-0.0366	-0.1715	-.0464	-0.0208	0.0215
	(0.0166)	(0.0162)	(0.0162)	(0.0168)	(0.0158)
White-collar	0.0797	0.1507	0.1360	0.0573	-0.0127
	(0.0166)	(0.0164)	(0.0171)	(0.0164)	(0.0179)

Note: Each personality trait has been standardized to have mean 0, variance 1.
Source: HILDA, waves 5, 9 and 13.

2.3.3. Stability of personality traits

The stability of personality traits is an important issue discussed both in the psychology and economics literature. Some studies find that personality traits are stable for adults (Terracciano et al. (2006), Terracciano et al. (2010)). Other studies find evidence of changing personality traits, particularly during younger ages (Almlund et al. (2011), Cunha and Heckman (2007), Cunha et al. (2010)). In this section, we use the HILDA data to examine the malleability of personality traits over the life cycle. We calculate average personality trait scores over the life cycle using the wave 13 sample (figure 14). After that, we investigate how working and schooling behaviors correlate with observed changes in personality (table 23), the main channel through which personality impacts agents in the structural model.

Figures 14(a) to 14(e) present the average score of "big-five" over the life cycle using the 2013 wave. Compared with the other three traits, openness to experience and emotional stability are relatively more persistent. Conscientiousness and agreeableness increase over time, with the greatest increase observed among respondents under the age of 35. Extraversion decreases with age until age 35, and then stays stable. Overall, traits appear to be more malleable for younger respondents.

We next investigate how education and working experiences correlate with personality change. To this end, we regress the changes of the Big-Five personality scores, standardized to mean 0 and variance 1, in the medium-run and in the long-run, defined as between years 2005-2008 and years 2005-2013, on years of experience in blue-collar and white-collar jobs and on years of schooling. The estimates are shown in table 23. Occupational experience shows little relationship with personality changes but education is related. During an eight-year time window, a male with one more year of schooling becomes more agreeable (0.032 std. dev.) and more conscientious (0.066 std dev.). The intercept term captures the age trend in personality traits and shows an increase in conscientiousness (0.105 std dev.) and emotional stability (0.090 std dev.) per one-year age growth.

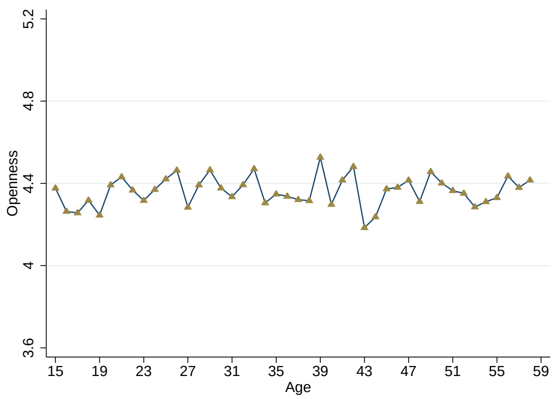
Table 23: Medium and long-run changes in Big-Five personality and education/occupation

	Extraversion		Agreeableness		Conscientiousness		Stability		Openness	
	Medium	Long	Medium	Long	Medium	Long	Medium	Long	Medium	Long
Education	-0.009	0.005	0.049**	0.032*	0.022	0.066**	0.004	0.017	0.022	0.012
	(0.022)	(0.017)	(0.023)	(0.018)	(0.023)	(0.018)	(0.026)	(0.020)	(0.023)	(0.018)
White Collar	-0.002	-0.008	0.007	0.006	-0.012	0.002	-0.010	0.001	0.000	-0.001
	(0.013)	(0.008)	(0.014)	(0.008)	(0.014)	(0.008)	(0.016)	(0.009)	(0.014)	(0.008)
Blue Collar	-0.011	−0.016**	0.014	0.011	0.003	0.001	-0.016	0.004	-0.013	-0.006
	(0.014)	(0.008)	(0.015)	(0.008)	(0.014)	(0.008)	(0.016)	(0.009)	(0.014)	(0.008)
Trend	0.004	0.031	-0.052	0.019	0.078	0.105*	0.142**	0.090	-0.039	0.044
	(0.056)	(0.053)	(0.060)	(0.056)	(0.059)	(0.057)	(0.067)	(0.064)	(0.059)	(0.056)

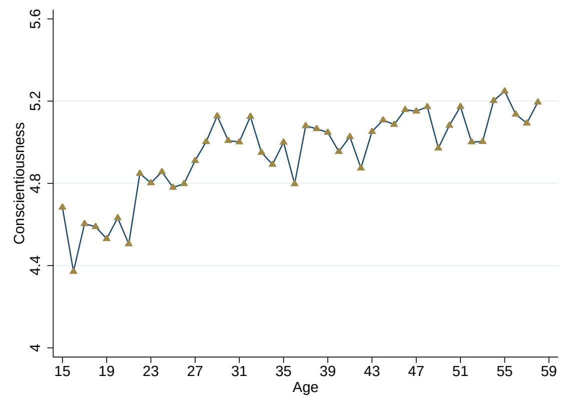
Note: * 10% significance level. ** 5% significance level. Standard errors in parentheses.
Source: HILDA, wave 5, 9 and 13.

Figure 14: The scores of "Big-Five" personality traits over time

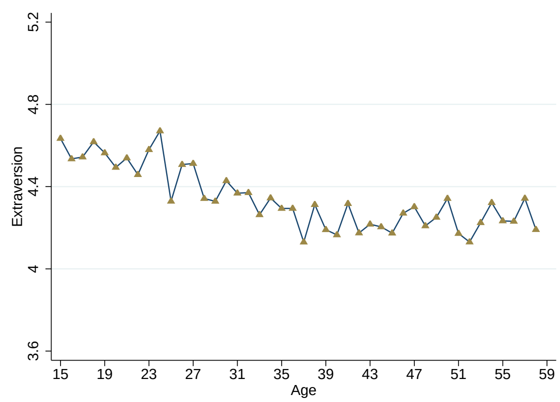
(a) Openness to experience



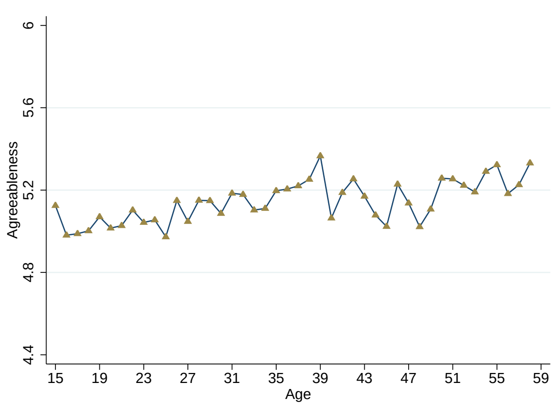
(b) Conscientiousness



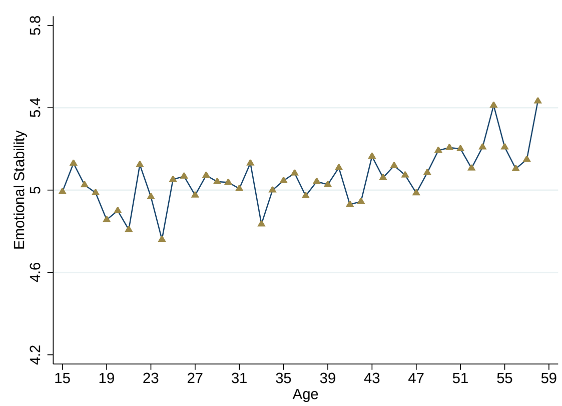
(c) Extraversion



(d) Agreeableness



(e) Emotional Stability



Source: HILDA, wave 2013.

2.3.4. Correlation between personality traits and schooling and occupational choices

We next estimate probit and linear regression models to investigate how personality traits correlate with individuals' college attendance and occupational choice decisions and wage performance. $D = 1$ if an individual has ever attended any form of college education by age 26, else $D = 0$. $D_0 = 0$ for individuals with blue-collar jobs and $D_0 = 1$ for individuals with white-collar jobs.

$$\begin{aligned}
Pr(D=1) &= \Phi(X\beta_D + \lambda'_D\theta_n + \gamma'_D C + \epsilon_D) \\
Pr(D_o=1) &= \Phi(X\beta_{D_o} + \lambda'_{D_o}\theta_n + \gamma'_{D_o} C + \epsilon_{D_o}) \\
\log w_{00} &= X\beta_{00} + \lambda'_{00}\theta_n + \gamma'_{00} C + \epsilon_{00} \quad D=0 \ \& \ D_o=0 \\
\log w_{10} &= X\beta_{10} + \lambda'_{10}\theta_n + \gamma'_{01} C + \epsilon_{10} \quad D=1 \ \& \ D_o=0 \\
\log w_{01} &= X\beta_{01} + \lambda'_{01}\theta_n + \gamma'_{10} C + \epsilon_{01} \quad D=0 \ \& \ D_o=1 \\
\log w_{11} &= X\beta_{11} + \lambda'_{11}\theta_n + \gamma'_{11} C + \epsilon_{11} \quad D=1 \ \& \ D_o=1
\end{aligned}$$

w_{D,D_o} represents the average earnings between ages 26-30 conditional on the schooling choice D and occupational choice D_o. The X variables include family background characteristics including an intact family dummy, parental occupations, parental education level, sibling number, sibling order, cohort effect and geographical locations. θ_n denotes the mean value of "Big-five" measurements in wave 5, 9 and 13.[11] C represents the value of cognitive ability.

Table 24 shows how personality traits and cognitive ability relate to college attendance decisions. Probit 1 only includes personality traits only whereas probit 2 includes additional controls for family background characteristics. The estimates indicate that the college atten-

[11] We normalized the score of each trait to be mean 0 and variance 1 before taking the average values across three data waves.

dance is mainly correlated with openness to experience, conscientiousness and extraversion. A unit standard deviation in extraversion decreases college attendance by 4.5 percentage points, whereas the same increase in openness to experience and conscientiousness increases the probability of college enrollment by 6.6 and 4.3 percentage points. For comparison, a one standard deviation increase in cognitive ability increases the probability of entering college by 15.7 percentage points.

Table 24: How personality traits and cognitive ability relate to schooling decisions

	Probit 1	Marginal	Probit 2	Marginal
Emotional Stability	0.084***	0.026	0.057*	0.017
Openness	0.228***	0.070	0.219***	0.066
Conscientiousness	0.137	0.042	0.142***	0.043
Agreeableness	-0.033***	0.010	0.028	0.008
Extraversion	-0.136***	-0.042	-0.150***	-0.045
Cognitive	0.514***	0.157	0.519***	0.157
Family Characteristics	No		Yes	
Observations	6101		4361	
R Square	0.1117		0.1255	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 25 shows the relationship between personality traits and occupational choices. The first regression only includes personality traits, whereas the second and the third regressions add family background characteristics and family background characteristics plus college choices. Conditioning on college attainment and family background, a one standard deviation increases in openness to experience increases the probability of having a white-collar job by 6.6 percentage points. At the same time, a one standard deviation increase in conscientiousness raises the probability of having a white-collar job by 2.4 percentage points.

Table 26 examines how personality traits relate to log wages for four educational and occupational groups. Among all personality traits, conscientiousness and openness to experience are the dominant traits in predicting wages. A one standard deviation increase in conscientiousness increases wages by about 10% for white-collar workers and 7% for blue-collar workers. Openness to experience, surprisingly has negative effect on wages; the values

Table 25: How personality traits and cognitive ability relate to occupation

	Probit1	Mgn1	Probit2	Mgn2	Probit3	Mgn3
Emotional Stability	-0.044	-0.015	-0.049	-0.015	-0.074	-0.022
Openness	0.205***	0.072	0.273***	0.093	0.224***	0.066
Conscientiousness	0.122***	0.043	0.103***	0.035	0.083**	0.024
Agreeableness	-0.016	-0.006	0.041	0.014	0.055	0.016
Extraversion	0.042	0.015	-0.012	-0.004	0.030	0.009
Cognitive	0.664***	0.232	0.573***	0.195	0.353***	0.105
College					1.153***	0.401
Family Characteristics	No		Yes		Yes	
Observations	4126		2855		2855	
R Square	0.1142		0.1355		0.2399	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

are statistically significantly different from zero for the blue-collar/no-college and white-collar/college groups.

Table 26: How personality and cognitive ability relate to log wages

	Blue Collar No College	White Collar No College	Blue Collar College	White Collar College
Emotional Stability	0.022	-0.045	0.024	0.001
Openness	-0.074***	-0.012	-0.078	-0.097***
Conscientiousness	0.085***	0.111***	0.067	0.092***
Agreeableness	-0.040	-0.021	-0.006	-0.046
Extraversion	0.036	0.030	0.113	0.029
Cognitive	0.017	-0.032	-0.041	0.010
Family Characteristics	Yes	Yes	Yes	Yes
Observations	1138	479	223	830
R Square	0.0593	0.0729	0.3095	0.0971

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2.4. The Model

We develop a discrete choice dynamic programming (DCDP) model of decision-making with regard to education, employment, and occupation sector over ages 15 to 58. At each age, individuals maximize their remaining discounted lifetime utility. The choice set in each year consists of four mutually exclusive options $m \in M$: working in either a blue- or white-collar occupation, attending school, or staying home. Let $d_m(a) = 1$ if the alternative m is chosen at age a, $d_m(a) = 0$ otherwise. Individual endowments at age 15 consist of personality traits, cognitive ability, and family background characteristics. These include parental schooling, the number of siblings, sibling order and whether the person lived with both parents at age 14. To allow for unobservable heterogeneity in a tractable way, we assume each individual is one of four types $k(a) = \{1, 2, 3, 4\}$. An individual's type can affect their pecuniary and nonpecuniary reward from choosing particular alternatives. One important innovation in our model that deviates from literature (e.g. Keane and Wolpin (1997)) is that it allows types to evolve over time in a way that may depend on age and changing personality traits.

We use $\Theta(a)$ to represent personality traits and $k(a)$ to denote the unobserved types at age a, which are assumed to be known by the individual but not known by econometricians. $s_o(a)$ represents all other observed state variables. At age 15, the initial type $k(15)$ is determined by the initial endowment $s_o(15)$. Then given the initial type $k(15)$ and observed state variables $s_o(15)$, the agent chooses the alternative $d_m(a)$ that gives the highest continuation value. The state variables, $s_o(16)$, are updated according to the choice $d_m(15)$, and then the new type $k(16)$ is drawn depending on $s_o(16)$ and the type of the previous period $k(15)$.

2.4.1. Laws of motion for $s_o(a)$ and $k(a)$

The time-varying part of $s_o(a)$ consists of four components so that $s_o(a) = (g(a), x_1(a), x_2(a), \Theta(a))$. $g(a)$ represents accumulated education while $x_1(a)$ and $x_2(a)$ represent accumulated blue-collar and white-collar experience at age a. We first specify the law of motion for states $g(a), x_1(a), x_2(a)$ and then discuss the transition probability functions governing the personality traits $\Theta(a)$ and types $k(a)$.

Years of schooling and occupation-specific experience evolve in a deterministic way. More specifically, the updating of $g(a), x_1(a)$ and $x_2(a)$ are defined as follows:

$$\begin{aligned} g(a) &: g(a+1) = g(a) + d_m(a) \\ x_i(a) &: x_i(a+1) = x_i(a) + d_m(a), i = \{1, 2\} \end{aligned} \tag{2.1}$$

As shown in section 2.3.3, personality traits are correlated with education but not with work experience. We assume that the true $n-th$ personality trait $\theta_n \in \Theta, \{n = 1, 2, 3, 4, 5\}$ is measured with error, which the measurement error shock denoted $\zeta_n(a)$. We adopt the following specification for the evolution of each trait:

$$\begin{aligned} \theta_n^M(a+1) &= \theta_n(a+1) + \zeta_n(a+1) \\ \theta_n(a+1) &= \theta_n(a) + \gamma_{0n} + \gamma_{1n}(a-15) + \gamma_{2n}d_3(a) + \gamma_{3n}(a-15)d_3(a) \end{aligned} \tag{2.2}$$

where $\theta_n^M(a+1)$ is the measure of the nth personality trait at age $a+1$ and $\theta_n(a+1)$ is the true trait without measurement error. γ_{0n} and γ_{1n} capture the age effects. The term $\gamma_{2n} + \gamma_{3n}(a-15)$ captures a potential age*education interaction effect.

As previously described, we allow the unobserved types to change in a way that may depend on age and on personality characteristics. We specify a Markov process through which types evolve. After the initial period, the type $k(a)$ can stay the same with probability $1-p(a)$ or change to a new type with probability $p(a)$.[12] Conditional on a type changing, we use notation $q_k(a)$ to represent the probability of becoming type $k \in \{1, 2, 3, 4\}$. Let $L(a)$ denote the Markov transition matrix of types between period a to period a+1. The matrix has the following form:

$$L(a) = p(a)\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} + (1-p(a))\begin{bmatrix} q_{k=1}(a) & q_{k=1}(a) & q_{k=1}(a) & q_{k=1}(a) \\ q_{k=2}(a) & q_{k=2}(a) & q_{k=2}(a) & q_{k=2}(a) \\ q_{k=3}(a) & q_{k=3}(a) & q_{k=3}(a) & q_{k=3}(a) \\ q_{k=4}(a) & q_{k=4}(a) & q_{k=4}(a) & q_{k=4}(a) \end{bmatrix} \tag{2.3}$$

where

$$p(a) = \frac{1}{1 + \exp(\gamma_7 + \gamma_8(a-15) + \gamma_9(a-15)^2)} \tag{2.4}$$

$$q_k(a) = \frac{\bar{v_k^a}(\Theta, c)}{\Pi_{k=1}^{K=4}\bar{v_k^a}(\Theta, c)} \tag{2.5}$$

[12] We assume the changing probability $p(a)$ does not vary by type k, so that different types have the same persistence at the same age.

$$\log \bar{v}_k^a(\Theta, c) = \gamma_{3k} + \sum_{n=1}^{N=5} \gamma_{4kn}\theta_n(a) + \gamma_{5k}c + \sum_{z=1}^{Z} \gamma_{6zk}d_z + \eta_k(a) \tag{2.6}$$

At age 15, the initial types are directly drawn from the distribution $q_k(15)$. In subsequent ages, types are updated following the Markov transition matrix $L(a)$. When $p(a)$ is close to 0, then $L(a)$ corresponds to an identity matrix $I_{4\times4}$ and the types, k, are fixed. When $p(a) = 1$, types do not persist from previous period. We estimate $p(a)$, allowing for the possibility that types become more or less persistent with age. The probability of each type $q_k(a)$ follows a multinomial logit form (equation 2.5). Equation 2.6 captures the correlation between types and their determinants, including current personality traits $\theta_n(a)$, cognitive skill c as well as background characteristics $d_z(a)$.

2.4.2. Rewards associated with each alternative

An individual can choose to work in either a blue-collar occupation or a white-collar occupation. The reward to a particular sector include the wage compensation $w_m(a)$ and any non-pecuniary reward $r_m(a)$. $\epsilon_m(a)$ is the preference shock when choosing $m-th$ alternative. $m = 1$ denotes the blue-collar alternative and $m = 2$ the white-collar alternative. This yields the following utility function at age a:

$$u_m(a) = w_m(a) + r_m(a) + \epsilon_m(a), m = \{1, 2\} \tag{2.7}$$

As in Keane and Wolpin (1997), the wage is specified as a human capital pricing equation. It is given by the product of the price per unit of human capital p_m and the amount of human capital $e_m(a)$ embodied in the individual. That is $w_m(a) = p_m e_m(a)$. Human capital is accumulated through work experience and by attending school:

$$\begin{aligned} e_m(a) = exp(e_m^k &+ \sum_{i=1}^{I} \beta_{m0i}d_i + \beta_{m1}g(a) + (\beta_{m2} + \beta_{m3}I\{x_m(a) \leq 2\})x_m(a) \\ &+ \beta_{m4}x_m^2(a) + \beta_{m5}x_m(a)g(a) + \xi_m(a)) \end{aligned} \tag{2.8}$$

which yields a log-wage equation the form:

$$\begin{aligned}\log w_m(a) = \log p_m + e_m^k + \sum_{i=1}^{I} \beta_{m0i} d_i + \beta_{m1} g(a) + (\beta_{m2} + \beta_{m3} I\{x_m(a) \leq 2\}) x_m(a) \\ + \beta_{m4} x_m^2(a) + \beta_{m5} x_m(a) g(a) + \xi_m(a)\end{aligned} \tag{2.9}$$

In (2.9), $d_i, i \in \{state \times cohort\}$ denotes a fixed effect of being a member of particular age cohort and residing in a particular state. e_m^k is the type-specific component of reward, which represents the advantage or disadvantage of type k when choosing alternative m. $g(a)$ represents the years of schooling and $x_m(a)$ denotes the working experience in sector m. The component $\beta_{m3} I\{x_m(a) \leq 2\} x_m(a)$ captures a potential differential in returns to experience when the agent is new in an occupation (has two years or less experience). The component $\beta_{m5} x_m(a) g(a)$ captures the interaction term between working experience $x_m(a)$ and education year $g(a)$, included to allow returns to experience to differ with education. $\xi_m(a)$ is a skill technology shock, which follows a i.i.d. normal distribution.

The second term in equation (2.7), $r_m(a)$, represents nonpecuniary aspects of choosing a certain occupation (such as working hours flexibility) expressed in monetary equivalent units. For the purpose of identification, we normalize the nonpecuniary utility from white-collar job $r_1(a)$ equal to 0. We allow the non-pecuniary utility from the blue-collar job $r_2(a)$ to vary with education level.

$$\begin{aligned} r_1(a) &= 0 \\ r_2(a) &= \beta_5 + \beta_6 I[g(a) \leq 12] \end{aligned} \tag{2.10}$$

If a person chooses to attend school, the per-period utility consists of two parts: a nonpecuniary component, which may reflect such as physical and mental costs when attending school, and a pecuniary component, such as tuition costs and fees. Thus, we have a school utility at age a defined by:

$$
\begin{aligned}
u_3(a) = e_3^k + \sum_{z=1}^{Z} \alpha_z d_z + \sum_{r=1}^{R} \alpha_r d_r + \alpha_0 I(age < 19) - \alpha_1 I(college) \\
- \alpha_2 I(graduate) + \epsilon_3(a)
\end{aligned}
\tag{2.11}
$$

The indicator d_z captures the potential effect of family background on a person's preference for attending school.[13] d_r is a cohort-specific effect. The term $\alpha_0 I(age < 19)$ captures the extra utility of attending school when the agent is under the age 19. α_3 and α_4 are per period schooling costs of attending college and attending graduate school. Lastly, e_3^k is the type-specific reward from attending school.

The reward from staying home, $u_4(a)$, consists of the type-specific component e_4^k, an age effect and an age squared effect, α_3 and α_4, and a home-staying preference shock $\epsilon_4(a)$, i.e.:

$$
u_4(a) = e_4^k + \alpha_3 \cdot age + \alpha_4 \cdot age^2 + \epsilon_4(a) \tag{2.12}
$$

It is worthwhile to mention that personality traits do not directly appear in the choice-specific utilities. Instead, they affect the choices indirectly through their influence on an individual's type probability. In addition, different types have different type-specific component e_m^k in each choice m. This structure reduces the dimensionality of the state space as it avoids the need to include a five-dimensional personality trait vector in the time-varying state space.

2.4.3. Information structure

In our model, individual heterogeneity comes from two sources: ex-ante endowments $s(15)$[14] and ex-post realized shocks $(\epsilon_m(a), \xi_m(a), \zeta_n(a), \eta_k(a))$. In terms of timing, we assume that the shocks governing the evolution of personality and of types are realized first, allowing

[13]The family background information includes sibling numbers, birth order and parental education level.
[14]the full list of state variable includes $s(a) = \{k(15), \Theta(15), c, Z, state, cohort\}$

individuals learn whether their type changed. After that, individuals observe preference shocks and choose their preferred sector. After this choice, wage shocks are realized.

Let $S^v(s) \subseteq S$ denote the set of visited states and $S^f(s) \subseteq S$ as the set of feasible states that can reached from s. Given the earlier time-line assumptions, we define $\iota(s)$ as the information set of the agent in state s by specifying all components known in the state, where

$$\iota(s) = \begin{cases} \epsilon_m(a); \zeta_n(a); \xi_m(a); \eta_k(a) : & \text{for all } s(a) \in S^v(s) \\ \epsilon_m(a+1) : & \text{for } s'(a+1) \in S^f(s) \\ k(15), \Theta(15), c, Z, state, cohort; \Omega : & \text{and for all s} \end{cases}$$

An individual in state s knows all state variable laws of motion, $\Pr(s(a+1)|s(a), d_m(a))$. He uses the distribution of wage shocks $F_m(\xi(s))$, idiosyncratic preference shocks $F_m(\epsilon(s))$, traits transition shocks $F_n(\zeta(s))$ and type transition shocks $F_k(\eta(s))$ to form an expectation over future states. For computational simplicity, $\xi_m(a)$ and $\zeta_n(a)$ are assumed to be uncorrelated and normally distributed, whereas $\epsilon_m(a)$ and $\eta_k(a)$ are assumed to be type I extreme value distributed. Conditional on the unobserved types, the other shocks are assumed to be iid over time.

2.5. Identification

The general procedure for incorporating multinomial types into longitudinal models dates back to Heckman (1981), Heckman and Singer (1984). The method was first used in the context of discrete choice dynamic programming (DCDP) models with fixed types in Keane and Wolpin (1997). The identification of serially correlated, unobserved types for a discrete choice model that satisfies a first-order Markov distribution is shown in Hu et al. (2015). Their identification strategy imposes one additional "limited feedback" restriction: namely that the type evolution is independent of choices m_{t-1} after conditioning on state variables s_{t-1}. This "limited feedback" assumption is satisfied in our model.[15] Following the

[15] Hu et al. (2015) also requires the stationary assumption of the Markov kernel, which is a common assumption in I/O applications(i.e. dynamic games). Their conclusion can be generalized to our case where the conditional choice probability is age-dependent.

argument in section 2.2 in Hu et al. (2015), one can nonparametrically identify both the distributions of unobserved types and the law of motions of all state variables (including unobserved types) using at least three time periods of data.

As the model is finite horizon, the case for identification of some of the model parameters can be made using the last time period. Wages in the blue collar and white collar sectors are observed. We impose a timing assumption on the model that individuals choose sectors after observing preference shocks but before observing wage shocks. Therefore, there is no selection problem in estimating the wage equation.

The utility values associated with the schooling choice and with the home choice as well as the nonpecuniary values of choosing a white or blue color job are not directly observed. In the last time period, the set-up of the choice problem is analogous that of a multinomial logit model given the types. Identification of these kinds of models is discussed in Horowitz (1981). The choices we observe allow us to infer relative but not absolute utilities, so identification requires normalizing one of the utility values. We normalize the nonpecuniary value of the white collar sector choice to be zero. Lastly, the difference in conditional choice probabilities by type identifies the type-specific components e_m^k of the flow utility functions.

Personality traits are observed in multiple time periods, so it is possible to directly estimate the transition process where personality traits at any time period are a function of lagged personality traits and of age following equation 2.2. The final parameter that we need to identify is the discount rate. The discount rate is identified through functional form assumptions that allow separation of the current period utility from future expected utility.

2.6. Estimation Strategy

2.6.1. Solving the dynamic programming problem

At the beginning of age a, an individual has the state vector $s(a)$, determined by his choices up to age a. As previously described, the evolving state variables include the accumulated sector-specific experience $x_i(a), i = 1, 2$, the completed schooling $g(a)$, personality traits $\Theta(a)$ and unobserved types $k(a)$.[16] Let $d_m(t) = 1$ denote that alternative m is chosen at age t. The value function at age a is the maximum over all possible sequences of future choices:

$$V(s(a), a, \Omega) = \max_{\{d_m(t)\}} E\left[\sum_{t=a}^{A} \delta^{\tau-a} \sum_{m=1}^{4} u_m(t) d_m(t) | s(a),\right]$$

where Ω denotes a set of parameter values. The summation over t denotes the ages and the summation over m denotes the different sector choices. The problem can be written in Bellman equation form.

The alternative specific value function is

$$V_m(s(a), a, \Omega) = \tilde{u}_m(s(a), a) + \delta E\left[V(s(a+1), a+1, \Omega) | s(a), d_m(a)\right]$$

for $a < A$, and

$$V_m(s(A), A, \Omega) = \tilde{u}_m(S(A), A)$$

in the last time period. As previously noted, to facilitate computation, we impose an assumption on the timing of the model that the sector is chosen after preference shocks are realized but before the wage shock is realized. We denote $\tilde{u}_m(s(a), a)$ to be $u_m(s(a), a)$ after integrating over the wage shock distribution (i.e. $\tilde{u}_m(s(a), a) = \int_{\xi_m(a)} u_m(s(a), a) f(w(\xi_m(a))) d\xi_m(a)$). Wages in the white and blue collar sectors are assumed to be both normally distributed and uncorrelated. The expectation in the Bellman equation is taken over future wage and preference shocks and over the random process that governs the transition of personality traits and the unobserved types. [17]

[16] The personality traits at the initial age may not directly be observable, so in some cases we infer them using the approach described in Appendix 2.11.

[17] Even though the realized wage shocks do not affect the contemporaneous utility associated with different sectors, the expected value functions will depend on the variance of the wage shocks.

The value function is the max over the alternative specific value functions:

$$V(s(a), a, \Omega) = \max_{m \in M} V_m(s(a), a, \Omega)$$

Recall that the preference shocks enter additively into $u_m(s(a), a)$ and, for computational simplicity, are assumed to follow an i.i.d. type I extreme value distribution with a location parameter 0 and a common scale σ_c.

Let $\tilde{V}_m(s(a), a, \Omega)$ denote the choice-specific value function excluding the contemporaneous sector-specific preference shock $\epsilon_m(a)$,.

$$V_m(s(a), a, \Omega) = \tilde{V}_m(s(a), a, \Omega) + \epsilon_m(a).$$

Because of the distributional assumption on the preference shocks, we have

$$\Pr(d_m(a) = 1 | s(a), \Omega) = \frac{\exp(\tilde{V}_m(s(a), a, \Omega)/\sigma_c)}{\sum_{j=1}^{4} \exp(\tilde{V}_j(s(a), a)/\sigma_c)}$$

As shown by Rust (1987), the expected value function can be written as

$$\begin{aligned} E\left[V(s(a+1), a+1, \Omega) | s(a), d_m(a)\right] &= E_{\epsilon_m(a)} \max_{d_m(a)} \sum_{m=1}^{4} d_m(a) \{\tilde{V}_m(s(a), a, \Omega) + \epsilon_m(a)\} \\ &= \sigma_c \log\left(\sum_{m=1}^{4} exp(\tilde{V}_m(s(a), a, \Omega)/\sigma_c)\right) + \sigma_c \gamma \end{aligned}$$

where γ is the Euler's constant and σ_c is the scale parameter of the preference shock.[18]

The dynamic programming problem uses backward recursion for each set of parameter values under consideration. That is, in the last period A, when there is no future expected value function and using the previous equation, one obtains $E\left[V(s(A), A) | s(A-1), d_m(A-1), A-1\right]$ for each possible point in the state space.

[18]This closed form representation of the value function is a big advantage in estimation because, without it, numerical integration over the structural errors is required to get the expected value function. It also generates an analytic one-to-one mapping between the choice probability and utility level of each choice. This tractable i.i.d. generalized extreme value (GEV) distributions assumption is also adopted in other recent DCDP papers such as Chan (2013) and Kennan and Walker (2011).

Plugging in $E\left[V(s(A),A)|s(A-1),d_m(A-1),A-1\right]$ into $\tilde{V}_j(s(A-1),A-1)$, one can then use the same expression to obtain $E\left[V(s(A-1),A-1)|s(A-2),d_m(A-2),(A-2)\right]$ and so on, back until the first time period. After solving the dynamic programming problem, one obtains the expected future value functions for all possible state points. It is then possible to use the model to simulate choices and to implement a simulated method of moments optimization algorithm to estimate the parameters.

2.6.2. Simulated Method of Moments estimation

Our model parameters are estimated by simulated method of moments. We use an unconditional simulation approach starting from age 15, because occupation-specific experience stocks are not observed at the time of sampling. The simulation process is briefly summarized as follows:

For each individual i, given a set of trial parameters Ω:

1. Solve backward for choice-specific value function $V_m(s(a),\Omega)$ and choice probability $\Pr(d_m(a)|s(a),\Omega)$ following the procedure described in the previous section.

2. Impute initial personality traits $\theta_n(15)$ following the procedure described Appendix 2.11. Initial unobserved types $k(15)$ are drawn from equation 2.5.

3. Starting from $s(15)=g(15)=0, x_i(15)=0, k(15), \theta_n(15)$, we simulate sequential shocks $\{\epsilon_m(a),\zeta_n(a),\xi_m(a),\eta_k(a)\}$ and compute the following outcomes: (1) agents' lifetime choices $d_m(a)$; (2) wage realizations $w_m(a)$ when $m=\{1,2\}$, $a=\{18,...,58\}$; and (3) personality traits $\theta_n(a), n=\{1,2,...,5\}$.

The simulation process is repeated for all i=1,2,...,N individuals, given their initial state variables. We then compute R moments using both the N simulated samples and the observed data, and then calculate the weighted difference between those R simulated moments $\tilde{M}_{N,R}(\Omega)$ and the data moments M_R, using the following objective criterion:

$$\hat{\Omega}_{N,R,W}=\arg\min_{\Omega}\Big((M_R-\tilde{M}_{N,R}(\Omega))'W_R(M_R-\tilde{M}_{N,R}(\Omega))\Big) \tag{2.13}$$

where M_R denotes the data moments, and $\tilde{M}_{N,R}(\Omega)$ represents the simulated moment evaluated at the parameter set Ω based on N repeated simulations.[19]

We use the variance information of each data moment to form the weighting matrix, W_R. Del Boca et al. (2014) show the consistency for this type of estimator for large sample sizes, $plim_{N\to\infty}\tilde{M}_{N,R}(\Omega_0) = M_R(\Omega_0)$.[20] In total, we match 505 moments to estimate 124 parameters. The following types of moments are used in estimation:

1. Sequential life-time choices:
    - The fraction of individuals in the blue-collar occupation sector by age (15-58).
    - The fraction of individuals in the white-collar occupation sector by age (15-58).
    - The fraction of individuals in school by age (15-58).
    - The fraction of individuals at home by age (15-58).
2. Earning profiles[21]
    - Average log earnings of blue-collar workers by age (18-58).
    - Average log earnings of white-collar workers by age (18-58).
    - The standard error of log earnings of blue-collar jobs by age (18-58).
    - The standard error of log earnings of white-collar jobs by age (18-58).
3. Personality traits

[19] This unconditional simulation algorithm is often used to estimate dynamic discrete choice models when some state variables are unobserved(e.g. Keane and Wolpin (2001), Keane and Sauer (2010)). The consistency and other asymptotic properties of this estimator based on unconditional simulation are discussed in Gourieroux and Monfort (1996), section 2.2.2.

[20] Compared with directly calculating the optimal weighting matrix, this method simplifies computation significantly. Altonji and Segal (1996) discusses that gains from using an optimal weighting matrix may be limited.

[21] We don't fit the earning between age 15-17 because too fewer observations have earning information at these ages.

- Mean value of openness to experience by four-year age groups and by waves.[22]
- Mean value of conscientiousness by four-year age groups and by waves.
- Mean value of extraversion by four-year age groups and by waves.
- Mean value of agreeableness by four-year age groups and by waves.
- Mean value of emotional stability by four-year age groups and by waves.

2.7. Estimates

2.7.1. Parameter Values

Tables 27-29 show the model parameter estimates along with standard errors. Table 27 shows the parameters corresponding to the per-period reward for each of the alternatives (white-collar job, blue-collar job, schooling, and home staying). An additional year of schooling increases white-collar and blue-collar wage offers by 4.47 and 3.99 percent. The reward for the first two years' work experience ($exp \leq 2$) is relatively high. One year of white-collar experience increases white collar wage offers by 16.24 percent, and one-year of blue-collar experience increases blue-collar wages by 29.32 percent. Although white-collar experience does not have a significant return in blue-collar jobs, blue-collar job experience is rewarded in the white-collar sector. The non-pecuniary terms capture the psychic difference between working in a white-collar or a blue-collar job. As previously described, we normalize the non-pecuniary utility from a white-collar job to 0. The non-pecuniary blue collar job premium is AU$23,388 for individuals who are not college graduates but only AU$3,377 for college graduates.

For the schooling option, we estimate a utility of AU$15,554 per year if an individual stays in school until age 17; a relatively high utility is needed to capture the drop-off in schooling after high school graduation. We find a net lump-sum cost of college education

[22]The four-year age groups are 15-18, 19-22, 23-26, 27-30, 31-34, 35-38, 39-42, 43-46, 47-50, 51-54, 55-58. The waves are 2005, 2009 and 2013.

Table 27: Model parameter estimates: reward functions

	1. *White-Collar*	2. *Blue-Collar*	*3. Schooling*	
	Skill Function		Tuition cost: college	13.5199(0.7885)
Mincer Equation			Additional cost:graduate school	10.6251(0.6737)
Schooling	0.0447(0.0047)	0.0399(0.0052)	Additional utility before age 19	1.5554(0.1181)
White-Collar experience	0.0105(0.0054)	-0.0013(0.0058)	**Constant:**	
Blue-Collar experience	0.0385(0.0048)	0.0244(0.0054)	Type I	8.1760(0.4348)
"Own" experience squared/100	-0.0293(0.0058)	-0.0304(0.0061)	Deviation of type 2	-3.7180(0.0059)
"Own" experience$\times$ edu	0.0103(0.0058)	0.0106(0.0054)	Deviation of type 3	-7.6862(0.0055)
"Own" experience ≤ 2	0.1624(0.0094)	0.2932(0.0185)	Deviation of type 4	-0.8058(0.0058)
Standard Error	0.4750(0.0246)	0.3773(0.0254)	**Family Background**	
Constant:			Family Intactness Dummy	0.0911(0.0060)
Type I	9.8638(0.0684)	9.3990(0.0807)	**Sibling(Omitted cat: only child)**	
Deviation of type 2	-0.0771(0.0058)	0.3893(0.0248)	multiple children, eldest one	-0.1054(0.0065)
Deviation of type 3	-0.6018(0.0061)	-0.3572(0.0063)	multiple($N < 4$), not eldest one	-0.0489(0.0068)
Deviation of type 4	-0.5935(0.0056)	-0.5772(0.0053)	multiple($N \geq 4$), not eldest one	-0.1947(0.0300)
State(Omitted cat:NSW)			**Parental Education(Omitted cat:no college)**	
VIC	-0.1267(0.0056)	-0.1593(0.0057)	One college	0.0750(0.0199)
QLD	-0.0306(0.0062)	-0.5000(0.0055)	Two colleges	0.3830(0.0090)
SA	-0.5045(0.0057)	0.5000(0.0333)	**Cohort(Omitted cat:40-49)**	
WA	0.0135(0.0060)	-0.0044(0.0060)	50-59	-0.1562(0.0058)
TAS	-0.2544(0.0053)	0.5007(0.0298)	60-69	-0.2539(0.0058)
NT	-0.5027(0.0054)	-0.5054(0.0045)	70-79	0.5064(0.0300)
ACT	0.2370(0.0153)	-0.0306(0.0059)	80-89	1.6202(0.0855)
Cohort(Omitted cat:40-49)			After 90	1.6620(0.1557)
50-59	0.1980(0.0134)	0.3038(0.0189)	*4. Home-staying*	
60-69	0.3508(0.0201)	0.4859(0.0250)	Age	0.0138(0.0056)
70-79	0.5334(0.0299)	0.6351(0.0369)	Age squared/100	0.0092(0.0059)
80-89	0.3010(0.0182)	0.5295(0.0305)	**Constant:**	
After 90	0.0003(0.0058)	0.0009(0.0058)	Type I	4.4872(0.1761)
	Non-pecuniary Values		Deviation of type 2	-1.0019(0.0059)
Constant	-	2.3388(0.1145)	Deviation of type 3	-2.2900(0.0055)
College Premium	-	-2.0011(0.1293)	Deviation of type 4	-1.1334(0.0065)
Preference Shock	0.9195(0.0594)		Discount Factor	0.8960(0.0284)

The unit for the non-pecuniary, school and home-staying columns is 10,000AU\$.

Table 28: Estimated coefficients on type probabilities

Types	I(baseline)	II	III	IV
Constant term	-	0.030	0.001	-0.010
		(.0058)	(.0060)	(.0062)
Cognitive	-	-0.508	-0.990	-1.520
		(.0063)	(.0058)	(.0060)
Openness to Experience	-	-1.500	-1.000	0.000
		(.0067)	(.0064)	(.0051)
Conscientiousness	-	-0.900	-0.520	-1.110
		(.0060)	(.0060)	(.0057)
Extraversion	-	-0.020	-0.026	-0.880
		(.0053)	(.0058)	(.0061)
Agreeableness	-	1.500	0.510	0.510
		(.0968)	(.0285)	(.0295)
Emotional Stability	-	-0.100	-0.110	-0.209
		(.0056)	(.0062)	(.0057)
Parental Background(baseline)				
Middle	0.001	0.007	0.000	-0.010
	(.0065)	(.0058)	(.0058)	(.0066)
High	0.020	0.020	-0.070	0.030
	(.0058)	(.0061)	(.0061)	(.0056)
Family Intactness	0.030	0.020	0.040	0.010
	(.0057)	(.0059)	(.0057)	(.0063)
Type Persistence	Time shift term		$Age - 15$	$\frac{(Age-15)^2}{100}$
Values	0.40		0.12	1.00
	(.0164)		(.0085)	(.0654)

Table 29: Estimated model coefficients for personality trait transitions

Traits	Edu	$Edu * (Age - 15)/100$	$Age - 15$	$(Age - 15)^2/100$
Openness to Experience	0.0022	-0.0042	-0.0022	0.0116
	(.0056)	(.0056)	(.0060)	(.0055)
Conscientiousness	0.0460	-0.1159	0.0342	-0.0694
	(.0051)	(.0052)	(.0050)	(.0070)
Extraversion	0.0049	-0.0057	-0.0167	0.0384
	(.0058)	(.0058)	(.0055)	(.0055)
Agreeableness	0.0364	-0.0968	0.0086	0.0136
	(.0059)	(.0061)	(.0053)	(.0061)
Emotional Stability	0.0079	-0.0141	0.0108	0.0075
	(.0054)	(.0057)	(.0052)	(.0062)

of AU$135,199 and a one-time cost of graduate school of AU$108,251.[23] This cost includes both tuition and living expenditures as well as potential psychological costs.

With regard to the home staying option, the flow utility is specified as quadratic in age. The utility of staying home increases from AU$13.8 at age 15 to AU$7,635 at age 58. Lastly, we estimate a discount rate parameter, β, equal to 0.8960 and preference scale parameter σ_c equal to 0.9195.

There is considerable variation in the estimated rewards across occupations for the unobserved types. For the two working options, types I and II have comparative advantages. Type I receives the highest reward in the white-collar occupation and type II receives the highest reward in the blue-collar occupation. With regard to the schooling alternative, type I gets the highest reward from attending school, followed by types IV, II and III. The benefit of type I is only slightly higher than that of type IV (AU$8,058), but much higher than for types II (AU$37,180) and III (AU$76,862). For the option of staying home, the rewards of type I-IV are AU$44,872, AU$34,853, AU$21,972 and AU$33,538.

Table 28 shows how estimated type probabilities relate to cognitive ability, personality traits and family background. Both personality traits and cognitive ability are important in type determination. High cognitive ability leads to a high probability of being type I and a relatively low probability of being type III or type IV. A high score of openness to experience implies a high probability of being types I or IV but a low probability of being type II. A person with high conscientiousness is more likely to be type I but less likely to be types II or IV. High agreeableness leads individuals to be type II rather than type I. The last two rows of table 28 show the malleability of types over time and how types become

[23] We compare our estimated costs with the real cost collected in Australia. For example, a 2014 HSBS report lists a per year cost for undergraduate study as AU$42,093, which includes AU$24,081 for fees and AU$18,012 for living costs. Source:http://www.about.hsbc.com.au/news-and-media/australia-the-most-expensive-country-for-education-hsbc-report. Another official website for Australia gives annual tuition fees for Bachelor's degree, Master's degree, and Doctoral degree in the range of AU$15,000-AU$33,000, AU$20,000-AU$37,000 and AU$14,000 to AU$37,000, respectively. Source:http://www.studyinaustralia.gov.au/global/australian-education/education-costs/education-costs-in-australia.

more persistent with age. As shown in figure 19, the probability of changing type starts at around 0.4 at age 15 then diminishes to 0 around age 34. In other words, our estimation results show that the types become relatively fixed in the mid-30s.

Table 29 shows the estimates of the probability that personality traits change, which is assumed to potentially depend on education and age.[24] Education has positive and significant effects only for the conscientiousness and agreeableness traits. One additional year of education at age 15 increases the level of conscientiousness and agreeableness by 4.60% and 3.64%, whereas it increases openness to experience, extraversion and emotional stability by only 0.22%, 0.49% and 0.79%.[25] The negative estimated coefficient on the interaction term between education and age (γ_{3n}) implies that the effect of education diminishes with age. For example, the effect of education on conscientiousness is negligible by age 55. The age effects on conscientiousness, extraversion and emotional stability are significantly larger than those on the other two traits. Conscientiousness increases with age at a diminishing rate. On the other hand, aging has an increasingly positive effect on emotional stability. Lastly, the results show that extraversion decreases with age.

2.7.2. Model Fit

Figures 15 and 16 display the data moments and compare model simulations with the data. The estimated moments pertain to three categories: the proportion choosing different sectors over life cycle (figure 15); the log wage of both white-collar and blue-collar occupations over life-cycle (figure 15); and the personality trait values over life-cycle (figure 16).

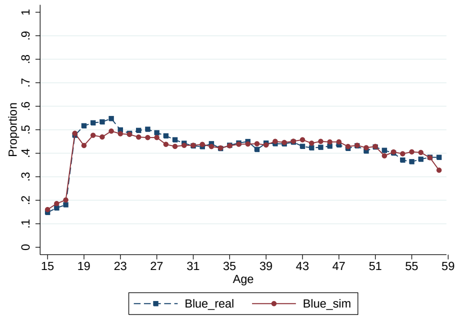
As seen in figure 15, the model captures salient features of data: (1) The fraction of blue-collar occupational choices exhibits an upward jump at age 18 and then declines gradually.

[24] Recall that personality trait changes were found to be strongly associated with schooling and to change with age up until the mid 30s, but were not found to be associated with white collar or blue collar job experience.

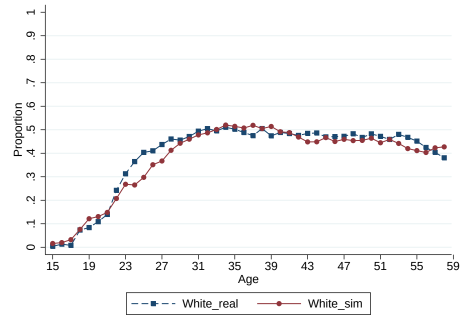
[25] By comparison, Schurer et al. (2015) find that university education increases scores on agreeableness for male students from low socioeconomic backgrounds but has no effect on conscientiousness. Our sample includes individuals with both senior secondary and university education, whereas their sample focuses only on individuals with university education. Li and Powdthavee (2014) studies the effect of a policy change that increased the compulsory minimum leaving school age, using HILDA data, and concludes that the average conscientiousness rises after the reform.

Figure 15: The comparison of choice distribution and earning profile between real data and model simulations

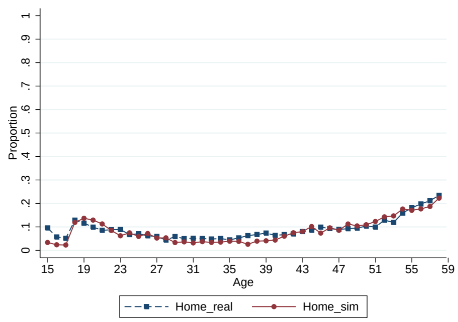
(a) Blue Collar Occupation



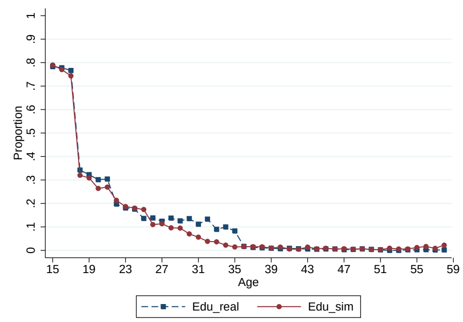
(b) White Collar Occupation



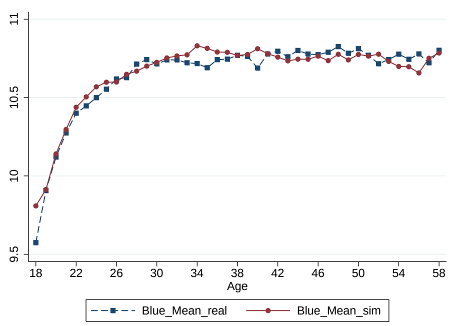
(c) home-staying



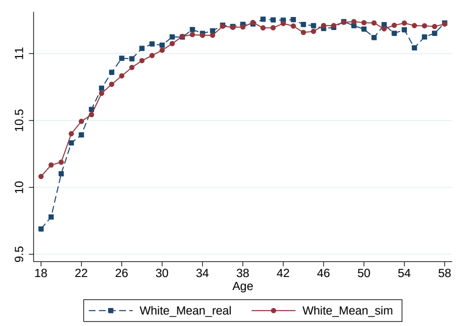
(d) Schooling



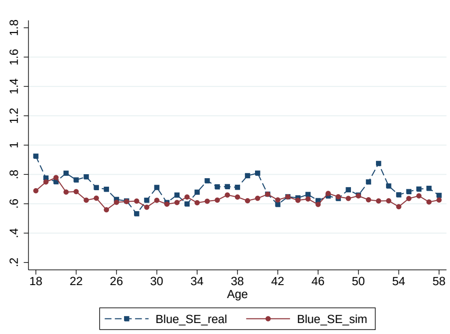
(e) Mean of log earning in blue-collar job



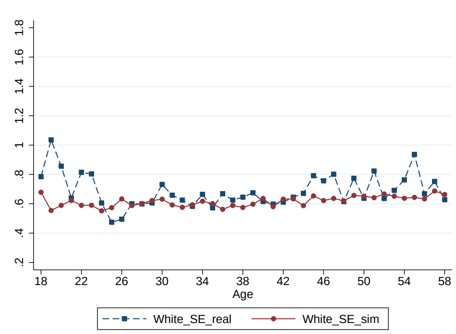
(f) Mean of log earning in white-collar job



(g) Standard error of log earning in blue-collar job



(h) Standard error of log earning in white-collar job



(2) The fraction of white-collar occupation choices grows smoothly from nearly 0 at age 18, reaches its peak in the mid-30s, and then moves downwards slowly. (3) Except for a small hump shape in the early 20s, the fraction that stays home exhibits a slow but persistent increase over the life cycle. (4) The fraction in school rapidly drops at age 18. Subsequently, a moderate decreasing trend takes over until it eventually reaches a stable level before age 39. (5) The concavity and the level of the earning profile are also captured by our simulated sample, both for white-collar occupation and blue-collar occupation. (6) Although the standard errors of log earnings from the data are more volatile, the simulated standard errors fit the observed average level reasonably well.

Figure 16 compares simulated personality traits and with the data measurements. In general, our simulated moments fit the data very well. Only 3 out of 165 moments fall outside of the 95% confidence interval generated by the corresponding data moments. Our simulations capture the following trait-specific patterns: 1. Openness to experience is quite stable over life-cycle. 2. Conscientiousness and emotional stability increase monotonically with age. 3. Extraversion decreases over time. 4. Agreeableness displays a hump shape during younger ages and then achieves a stable level after that.

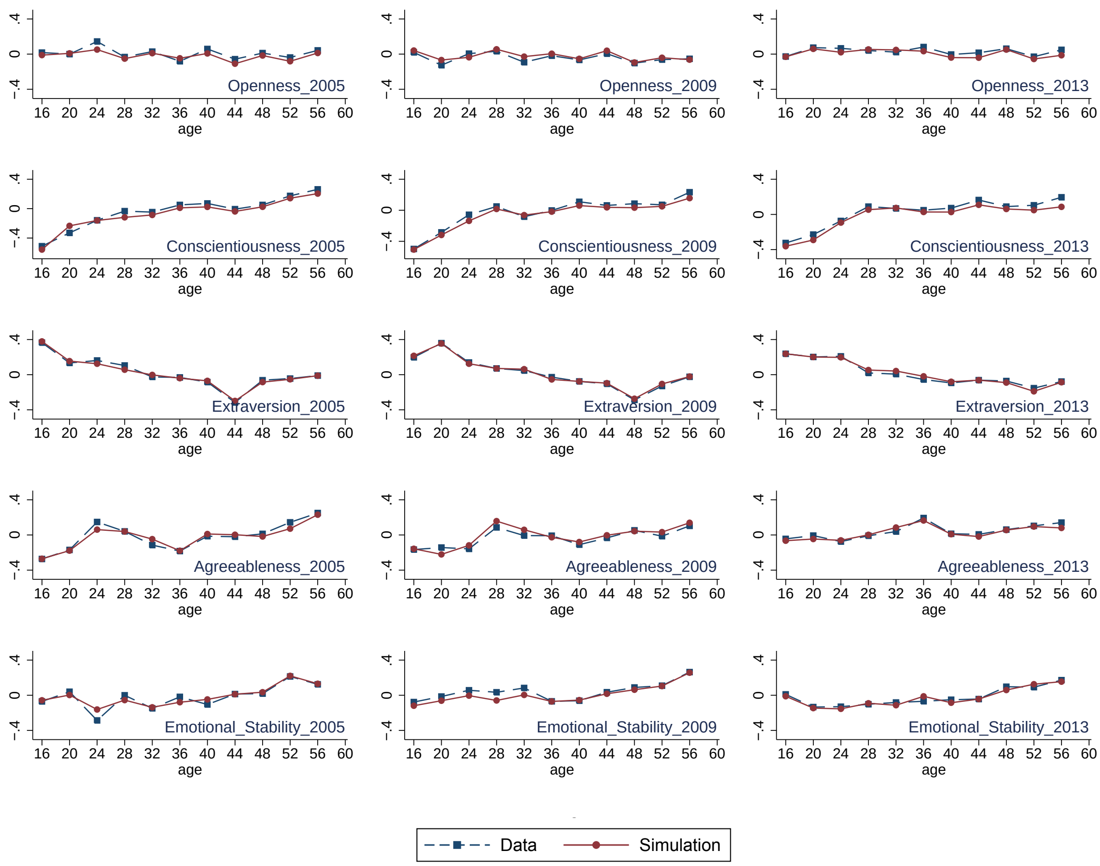
2.8. Model Simulation Results

We next use the estimated model to simulate individuals' choices. First, we explore the link between personality traits, types and choices. Second, we examine the relative importance of personality traits in explaining ex-ante heterogeneity compared with other initial endowments (e.g. cognitive ability, family background). Third, we implement a likelihood ratio test to test the hypothesis that the unobserved types are stable over time, which is assumed in many other studies. We reject this hypothesis.

2.8.1. Understanding the link between personality traits, types and choices

Table 30 examines the type distributions within the different sectors. The row labeled "original" shows the proportions of the four types within each sector. The row labeled

Figure 16: The comparison of personality traits between real data and model simulations



Data Source: "Big-Five" personality traits gathered in wave 2005, 2009 and 2013.

"adjusted" gives the proportions adjusted by the proportion of each type in the population. Our simulation results show that type I has a comparative advantage in schooling and in the white-collar sector. Type II has a comparative advantage in the blue-collar sector. Type III is more likely to be in the blue-collar sector or to stay at home. Type IV is more often at school or at home.

Table 30: Simulated type proportions for different sector choices

Occupation		Type I	Type II	Type III	Type IV
White-collar	Original	47.89	17.58	9.00	25.53
	Adjusted	43.87	15.05	12.92	24.45
Blue-collar	Original	8.61	48.03	26.47	16.89
	Adjusted	7.89	41.12	38.01	16.18
Schooling	Original	37.53	15.28	4.17	43.03
	Adjusted	34.38	13.08	5.99	41.22
Home staying	Original	7.46	9.17	29.37	54.00
	Adjusted	6.83	7.85	42.17	51.72
Total		27.29	29.20	17.41	26.10

Table 31 shows mean personality trait values and cognitive scores for each of the four types. Type I has higher personality scores in all five dimensions and also high cognitive scores. As was seen in Table 7, openness to experience and conscientiousness and a high cognitive score were the most predictive of higher educational attainment. Type II has the lowest score on openness to experience and also lowest cognitive score. Type III has high scores on extraversion and conscientiousness, although not as high as observed for Type I. Lastly, type IV (the home-staying type) has low scores on both emotional stability and conscientiousness.

Figure 17, a radar chart, provides a graphical depiction of the average levels of personality traits and cognition among types. Each equi-angular spokes ("radii") represents one dimension of personality traits. Each star-like hectagon denotes the values of the "Big-five" along with the cognitive score for each type. It is clear that type I has the highest values of all five traits and for cognition, because its hectagon totally covers the other three types' hectagon. It seems that high cognitive ability and high values of personality traits tend to

Table 31: Average personality traits and cognitive ability by type

		Type I	Type II	Type III	Type IV
Openness	Mean	0.466	-0.614	-0.253	0.324
	SE	(0.004)	(0.004)	(0.004)	(0.004)
Conscientiousness	Mean	0.453	-0.274	0.069	-0.406
	SE	(0.004)	(0.004)	(0.005)	(0.004)
Extraversion	Mean	0.289	0.113	0.168	-0.427
	SE	(0.004)	(0.004)	(0.005)	(0.004)
Agreeableness	Mean	0.300	-0.201	0.002	-0.059
	SE	(0.004)	(0.004)	(0.005)	(0.004)
Stability	Mean	0.127	0.022	0.088	-0.318
	SE	(0.004)	(0.004)	(0.005)	(0.004)
Cognition	Mean	0.473	-0.165	0.056	0.011
	SE	(0.004)	(0.004)	(0.005)	(0.004)

be clustered in type I individuals, who are those that tend to acquire more schooling and to work in the white collar sector.

Figure 17: Average personality traits of each type

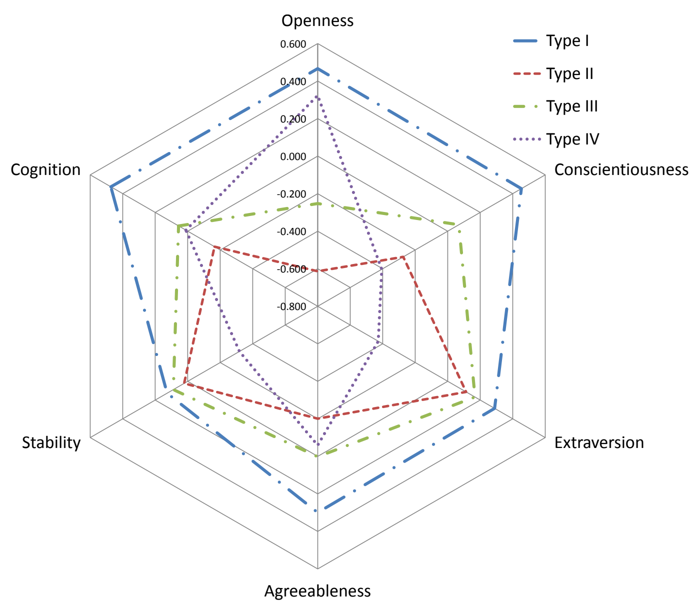
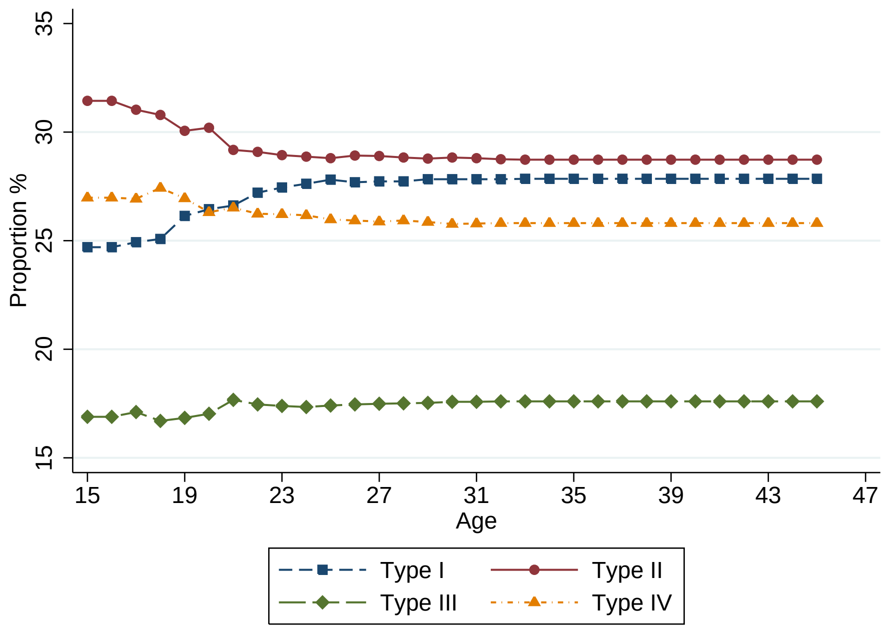


Figure 18 shows how the fraction of types change for different age cohorts. With age, the proportions of type II and IV, decrease while the proportions of type I and III increase. Those changes are driven primarily by increasing levels of conscientiousness. Our estimates in table 29 indicate that the average level of conscientiousness increases over time, both because of increasing levels of education and because of a direct effect of age. A higher conscientiousness score increases the probability of being type I or type III. Figure 19 plots the probability that types change over time. In general, types become more stable with age. The probability of switching types starts at around 0.4 at age 15 and diminishes to 0 by age 36.

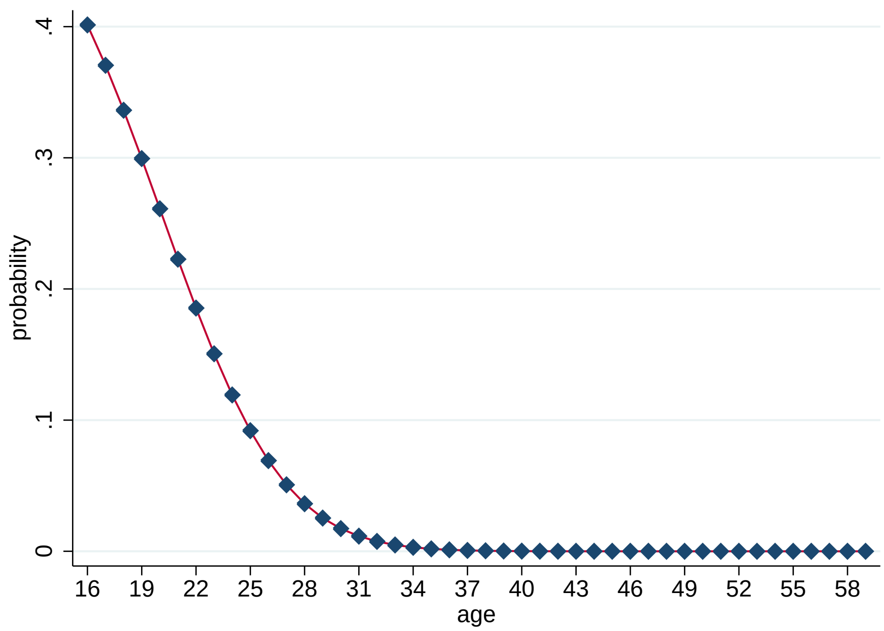
Figure 18: The fraction of types by age cohort



2.8.2. Exploring the importance of personality traits in explaining ex-ante life-time utility heterogeneity

To understand the importance of personality traits in explaining ex-ante utility heterogeneity, we estimate a linear regression where the dependent variable is the expected present value of lifetime utility at the age of 15 and the independent variables are initial personality traits, cognitive ability and family background. Table 32 summarizes the regression results under three different specifications of regressors: (i) only family background (ii) both family background and cognitive ability, (iii) personality traits, cognitive ability and family characteristics. The first two regressions are comparable to model specifications reported

Figure 19: The probability of type change by age



in Keane and Wolpin (1997). They estimate a similar regression and report that adding Armed Forces Qualification test score(AFQT), the "ability" score measure, increases the R^2 from 0.10 to 0.14. In our case, including the cognitive ability measurement increase the R^2 from 0.095 to 0.121. When we estimate regression specification (iii) including, in addition, personality traits, we get a further increase in R^2 to 0.154.

2.8.3. Testing the hypothesis of type stability

A novel feature of our model relative to other models in the literature is that unobserved types evolve over time in a way that may depend on age and the evolution of personality traits. In this section, we test the validity of this assumption by comparing our model with an alternative "fixed types" model. In the fixed type model, the probability of each type does not depend on personality traits but is still determined by other age 15 endowments (e.g. family background). Types are assumed to be fixed after age 15. This assumption is almost the same as that of Keane and Wolpin (1997)[26].

[26] In Keane and Wolpin (1997), they assume the initial type distribution only depends on with initial schooling years(10 years or more V.S. nine years or less.) Then they calculate the conditional probability of

Table 32: Determinants of ex-ante utility variation

	Reg 1	Reg 2	Reg 3
Intactness	-0.814*	-0.612	-0.579
Father Occupation	0.476	0.264	0.300
Parental Education	0.546**	0.284	0.235
Sibling	-0.417**	-0.337*	-0.297*
Cohort	1.454***	1.452***	1.406***
State	0.850***	0.844***	0.836***
Cognitive		2.080***	1.975***
Openness			0.245
Conscientiousness			1.210***
Extraversion			0.905***
Agreeableness			0.347*
Emotional Stability			-0.234
Observation	4215	4215	4215
R square	0.095	0.121	0.154

Table 33 shows the null hypotheses of the alternative model specifications and its corresponding criteria function. The LR-test indicates that the "fixed type" model is rejected with a p-value less than 0.01.

Table 33: Model specification test

	Baseline model	"Fixed type" model
Null Hypothesis		H0: $P_a = 0, \gamma_{4kn} = 0$
Distance Measure	2279.976	2406.325
LR test		126.349
The number of restrictions		18
$\chi^2(0.01)$ criteria		34.80

2.9. Two education policy experiments: compulsory senior secondary school and a college subsidy

We next use the estimated dynamic discrete choice models, both the variable types and fixed types versions, to evaluate the effects of two education policies, a tuition subsidy program and a compulsory schooling policy.

becoming each type on individual's family background information. We model the dependence between the initial types distribution and family background characteristic directly

2.9.1. Using the model to simulate the effects of educational policies

Since the late 1980s, the Australian government started providing financial assistance to students through a program called the Higher Education Contribution Scheme (HECS) and, after 2005, the Higher Education Loan Programme (HELP). With the goal of relieving the financial burden of a university education, those eligible for HECS-HELP can either receive no interest student loans or get a 10 % discount on the upfront payment. Some students also receive direct financial help to cover living expenditures through a means-tested programs (such as Austudy or Youth Allowance). Motivated by these financial aid programs, we use the model to simulate the effects of a hypothetical policy that reduces the cost of attending college by 50%.

Our second policy experiment is motivated by the spatial variation in compulsory schooling requirements across different states and territories. The compulsory education policy in Australia is age-based. In 2009, the minimum school leaving age in Queensland, Western Australia, South Australia and Tasmania was 17, whereas the leaving age in other areas was between 15-16. [27] In 2010, areas with lower compulsory school attendance ages came up with plans to increase compulsory schooling.[28] As a result, students in all states and territories are now required to stay in school until age of 17. (National Report on Schooling in Australia 2011) Inspired by these policies, we consider the imposition of a national compulsory secondary school rule that forces individuals to stay in school until at least age 17.[29]

We next use model to simulate to the effects of the two education policies previously de-

[27]Source: National Report on Schooling in Australia 2009.

[28]From 2010, New South Wales, Victoria, Northern Territory and Australian Capital Territory all claim that local students need to complete Year 10 and then participate in education, training or employment until they turn 17.

[29]We aware that the individuals who are younger than age 18 after the year 2009 in HILDA data should be already subject to the compulsory education policy. However, currently, the policy is not strictly enforced. The school enrollment rates for teenagers ages 15-18 are 84.9%(175/206) in year 2010, 90.0%(226/251) in the year 2011, 89.8%(211/235) in the year 2012 and 83%(176/212) in the year 2013. These enrollment rates are stable and do not significantly differ for years prior to 2009. Our baseline model estimation assumes no compulsory schooling law and we simulate the effects of a compulsory schooling law that is strictly enforced.

scribed, considering both mean effects and distributional effects. To understand the importance of allowing for time-varying types, we compare the simulated policy effects of 4215 individuals obtained under the baseline model to that obtained under a restricted "fixed type" model. Table 34 shows the effect of the two policies.[30] Specifically, we examine effects on (1)the percentage high school graduates; (2) the percentage college graduates; (3) the average years of education; (4) the annual earnings for workers; and (5) the expected lifetime utility gain. In each of these categories, we first present the values under baseline model in the row labeled as "benchmark". The two rows labeled "50% college subsidy" and "compulsory senior secondary school" show the deviations from baseline values under two separated policy experiments.

Table 34: The effect of educational policies on schooling and labor market outcomes, by type

Model	Type I	Type II	Type III	Type IV	Total
	Percentage Finishing High school				
Benchmark	98.4	78.4	75.5	97.9	88.2
50% college subsidy	0.2	0.5	0.5	0.2	0.2
Compulsory schooling	1.6	21.6	24.5	2.1	11.8
	Percentage College Graduates				
Benchmark	43.6	19.5	21.8	43.9	32.4
50% college subsidy	32.1	19.1	21.6	28.3	25.2
Compulsory schooling	0.5	1.9	4.2	0.8	1.7
	Years of Education				
Benchmark	14.347	12.132	12.072	14.832	13.431
50% college subsidy	1.003	0.593	0.687	0.924	0.799
Compulsory schooling	0.031	0.477	0.635	0.051	0.279
	Annual Earnings (for workers)				
Benchmark	96852.8	71946.8	34145.8	44211.4	66324.1
50% college subsidy	6672.8	2389.8	2340.0	8208.7	4718.7
Compulsory schooling	606.2	3616.2	2804.3	451.0	2210.4
	Utility Change(Unit: AU$10,000)				
Benchmark	80.132	73.908	68.623	73.786	74.520
50% college subsidy	1.758	0.574	0.687	0.477	1.135
Compulsory schooling	-0.831	-3.434	-5.328	-0.882	-2.423

Comparing the effects of two policies, two features stand out. First, the compulsory school-

[30]Because the types change over time and are potentially influenced by education, we classified agents according to their initial type at age 15.

ing policy has the most direct positive effect on the secondary school completion rate (+11.8pp), whereas the college subsidy has the largest positive impact on the fraction of college graduates (+25.2pp). However, because individuals are forward looking and graduation from senior secondary school is the prerequisite to attending college, the compulsory school policy also stimulates college graduation (+1.7pp) and college subsidy policy also encourages senior secondary school completion (+0.2pp). Second, these two policies affect different types of individuals. The college subsidy increases the average years of completed education by around one year for types I and IV but only by around a half year for types II and III. In contrast, the compulsory school policy increases years of education by about one half year for types II and III but has almost no effect for types I and IV.

We observe a similar pattern for labor market outcomes. Under the college subsidy intervention, types I and IV experience an average increase in annual earnings of AU$ 6,772.8 and AU$ 8,208.7. The increases observed for types II and III are only AU$ 2,389.8 and AU$ 2,340.0. When implementing the compulsory schooling policy, types II and III benefit the most. The annual earnings increases of those two types are AU$ 3,616.2 and AU$ 2,804.3, whereas the increases of other two types are only AU$ 606.2 and AU$ 451.0. The differences between types are caused by the original education level of each type. Secondary school completion is already so prevalent among types I and IV, thus few individuals of those types are affected by the compulsory schooling policy. These individuals are more likely to face the trade-off between finishing college or not and are most strongly influenced by the college subsidy policy.

Table 35 reports the effects of two policies on personality traits at age 30, when most have completed their education. The "benchmark" row shows the average trait score of each type. The rows "50% college subsidy" and "compulsory senior secondary school" report the additional change under these two policies. In general, the effects of both policies on traits are positive. However, the change of conscientiousness and agreeableness are one-order of magnitude larger than the change of the other three traits. Conscientiousness increases by 0.026, equivalent to 0.93 of its standard error, under a 50% college subsidy policy and 0.015 (equivalent to 0.53 of its standard error) under compulsory senior secondary school policy. Agreeableness increases by 0.020, equivalent to 0.71 standard error, under a 50% college subsidy policy and 0.011 (equivalent to 0.39 standard error) under compulsory senior secondary school policy.

Table 35: The effect of educational policies on personality traits, by type

Model	Type I	Type II	Type III	Type IV	Total
			Openness to experience (at age 30)		
Benchmark	0.458	-0.634	-0.262	0.319	-0.018
50% college subsidy	0.002	-0.001	0.003	0.001	0.001
Compulsory schooling	-0.007	0.003	0.005	0.000	0.001
			Conscientiousness (at age 30)		
Benchmark	0.388	-0.357	-0.008	-0.450	-0.113
50% college subsidy	0.031	0.010	0.017	0.034	0.026
Compulsory schooling	0.000	0.027	0.020	0.001	0.015
			Extraversion (at age 30)		
Benchmark	0.338	0.144	0.212	-0.374	0.075
50% college subsidy	0.003	-0.001	-0.002	0.004	0.004
Compulsory schooling	0.007	-0.008	0.004	-0.001	0.002
			Agreeableness (at age 30)		
Benchmark	0.251	-0.279	-0.058	-0.103	-0.048
50% college subsidy	0.024	0.012	0.005	0.030	0.020
Compulsory schooling	0.005	0.016	0.018	0.003	0.011
			Emotional Stability (at age 30)		
Benchmark	0.027	-0.073	0.003	-0.406	-0.118
50% college subsidy	0.001	0.009	-0.008	0.008	0.005
Compulsory schooling	0.002	0.001	0.012	-0.006	0.003

2.9.2. Understanding the importance of changeable types

To understand the empirical importance of allowing for changing types, we evaluate the same two education policies under the restricted "fixed types" model. The results are reported in table 36. Compared with table 34, there are two main differences. First, the policy impacts are now more concentrated among types. The college subsidy policy only affects the college graduation decision of type I and type IV, while compulsory senior secondary school policy essentially only affects the senior secondary school certificate completion rate of types II

Table 36: The effects of educational policies under the restricted model with fixed types

model simulation	Type I	Type II	Type III	Type IV	Total
	Percentage Finishing High school				
Benchmark	100.0	73.6	41.2	100.0	81.8
50% college subsidy	0.0	0.0	0.0	0.0	0.0
Compulsory senior secondary school	0.0	26.4	58.8	0.0	18.2
	Percentage College Graduates				
Benchmark	55.8	0.2	0.0	77.1	34.9
50% college subsidy	35.0	8.3	0.0	17.0	15.9
Compulsory senior secondary school	0.8	0.0	0.0	1.2	0.5
	Years of Education				
Benchmark	14.637	11.813	10.993	15.409	13.354
50% college subsidy	1.053	0.249	0.004	0.547	0.487
Compulsory senior secondary school	0.023	0.484	1.150	0.039	0.361
	Annual Earnings(for workers)				
Benchmark	100481.5	69533.0	29793.3	47273.6	66004.0
50% college subsidy	4656.1	909.5	9.4	7232.8	2943.7
Compulsory senior secondary school	484.9	2390.3	2565.4	760.0	1592.8
	Utility Change(Unit: AU$10,000)				
Benchmark	83.971	77.209	53.354	65.214	71.559
50% college subsidy	2.504	0.064	0.000	2.246	1.251
Compulsory senior secondary school	-0.261	-2.196	-4.386	-0.396	-1.597

and III. Second, the effects on labor market earnings are smaller. In our baseline model, the 50% college subsidy policy and the compulsory senior secondary school policy boost employed workers' average annual earnings by AU$ 4,718.7 and AU$ 2,210.4. In contrast, the earning increase drops to AU$ 2,943.7 and AU$ 1,592.8 in the restricted "fixed type" model.

The reason for these differences is fairly straightforward. When type is changeable, the education investment has both a direct reward in terms of increasing wage offers and an indirect reward through the chance to become a different type. Table 37 shows the type distribution in each age group under the baseline model and under the two policies. When years of education increases, a larger fraction of type II and type IV agents switch to type I. [31] The "fixed types" model shuts down the second indirect channel, which lowers the marginal benefits of education is lower. Thus, individuals in the fixed type model,

[31] Although the proportion of type III also increases, the increment of type I is much larger.

Table 37: The effect of education policies on type proportions at different ages

Age		Type I	Type II	Type III	Type IV
15	Benchmark	24.70	31.44	16.89	26.98
	50% college subsidy	24.70	31.44	16.89	26.98
	Compulsory senior secondary school	24.70	31.44	16.89	26.98
21	Benchmark	26.14	30.06	16.84	26.95
	50% college subsidy	26.19	30.04	16.92	26.86
	Compulsory senior secondary school	26.43	29.94	17.13	26.50
27	Benchmark	27.45	28.94	17.39	26.22
	50% college subsidy	27.69	28.75	17.58	25.98
	Compulsory senior secondary school	27.73	28.78	17.58	25.91
> 33	Benchmark	27.83	28.78	17.53	25.86
	50% college subsidy	28.09	28.61	17.77	25.53
	Compulsory senior secondary school	28.11	28.61	17.67	25.60

especially from disadvantaged groups, are less responsive to educational policies, which also leads to an overestimation of the inequality in policy effects.

2.9.3. Heterogeneous policy effects by family background social-economic status (SES)

Lundberg (2013) emphasizes the importance of family background in understanding the correlation between personality traits and college graduation. Therefore, we investigate the heterogeneous effects of two policies on individuals from different family backgrounds. The social-economic status is defined in terms of parents' educational attainment. In group I, both parents have education equal to high school or less. In group II, one parent has some college, and in group III, both parents have above high school graduation.[32] We find the personality patterns between individuals from different SES are exactly the same as those reported in Lundberg (2013). Individuals from more advantaged family backgrounds tend to have high scores for conscientiousness, openness to experience as well as emotional stability (the opposite of neuroticism).

Tables 38 and 39 summarize the effects of both the college subsidy policy and the compulsory senior secondary school policy. The policy effect of the college subsidy is roughly equally

[32]We did not consider the family intactness as additional dimension, because the majority (82.89%) grew up with both biological parents in our sample.

distributed across the different SES groups in terms of education increase. The increases in conscientiousness are 0.025, 0.028 and 0.025 for Groups I, II and III. [33] However, with regard to annual earnings increase, the benefits of the college subsidy policy accrue more to advantaged families. The earnings increase for Group I is AU$4410.2, while the earnings increase for Group III is AU$4869.6. On the other hand, the compulsory senior secondary school policy has larger effects on individuals from disadvantaged backgrounds. The average education enhancement for Group I is 0.32 year, whereas the average increase of Group III is only 0.21 year. Considering the labor market outcomes, the earnings increase of Group I is AU$ 2,148.4, and the earning increase of Group III is AU$ 2,000.8. Regarding the personality traits, we observe a larger improvement for the least advantaged group in both conscientiousness (0.017 of Group I vs. 0.011 of Group III) and openness to experience (0.013 of Group I vs. 0.009 of Group III).

2.10. Conclusions

This paper develops a dynamic model of schooling and occupational choices that incorporates personality traits. As is common in the discrete choice literature, we introduce unobservable types' to capture agents' heterogeneous comparative advantages in schooling and in particular occupational sectors. One innovative aspect of the model is that we allow the unobserved types to change over time in a way that may depend on age and on evolving personality traits. We perform a likelihood ratio test to examine the assumption that types are fixed, which is strongly rejected. Our estimates indicate that types are malleable when agents are young but become stable after age 36. Another finding is that high levels of cognitive skills and high personality trait scores, in all five dimensions, tend to be clustered in a certain type of individual. This type also acquires more schooling and tends to work in the white collar sector. Much of the prior economics literature emphasizes the role of cognitive skills in determining lifetime outcomes, but our analysis shows that high cognitive skills, on average, go hand-in-hand with high non cognitive skills.[34] For this reason, the relevance of

[33] This increase is equal to about one standard error of the mean value.
[34] See, for example, Neal and Johnson (1996).

Table 38: The effect of educational policies on labor market outcomes by SES background

	Socio Economic Status (SES)			
Model simulation	I	II	III	Total
	Percentage of Finishing High school			
Benchmark	84.7%	87.9%	91.8%	88.1%
50% college subsidy	0.3%	0.2%	0.4%	0.3%
Compulsory senior secondary school	15.3%	12.1%	8.2%	11.9%
	Percentage of College Graduates			
Benchmark	25.4%	30.2%	41.6%	32.4%
50% college subsidy	25.1%	26.9%	23.5%	25.3%
Compulsory senior secondary school	0.9%	2.1%	1.9%	1.7%
	Education Years			
Benchmark	13.081	13.346	13.865	13.431
50% college subsidy	0.778	0.854	0.756	0.799
Compulsory senior secondary school	0.324	0.301	0.210	0.278
	Annual Earning(for workers)			
Benchmark	62861.9	66433.2	69580.6	66324.1
50% college subsidy	4410.2	4860.8	4869.0	4718.7
Compulsory senior secondary school	2148.4	2434.2	2000.8	2210.4
	Utility Gain(Unit: AU$10,000)			
Benchmark	73.135	74.380	76.014	74.500
50% college subsidy	0.878	1.091	1.434	1.155
Compulsory senior secondary school	-2.812	-2.426	-2.044	-2.403

Table 39: The effect of educational policies on personality traits by SES background

	Socio Economic Status (SES)			
Model simulation	I	II	III	Total
	Personality Traits at age 30			
	Openness to experience			
Benchmark	-0.180	0.019	0.138	-0.018
50% college subsidy	0.002	-0.036	0.002	0.001
Compulsory senior secondary school	0.001	-0.037	0.001	0.001
	Conscientiousness			
Benchmark	-0.128	-0.127	-0.084	-0.113
50% college subsidy	0.025	0.028	0.025	0.026
Compulsory senior secondary school	0.017	0.016	0.011	0.014
	Extraversion			
Benchmark	0.037	0.066	0.121	0.075
50% college subsidy	0.003	0.004	0.004	0.004
Compulsory senior secondary school	0.002	0.003	0.002	0.002
	Agreeableness			
Benchmark	-0.098	-0.063	0.019	-0.047
50% college subsidy	0.018	0.021	0.020	0.020
Compulsory senior secondary school	0.013	0.012	0.009	0.011
Emotional Stability				
Benchmark	-0.194	-0.102	-0.064	-0.118
50% college subsidy	0.006	0.006	-0.005	0.005
Compulsory senior secondary school	0.003	0.003	0.001	0.002

the cognitive dimension as a determinant of labor market success may be overemphasized in studies that ignore non-cognitive attributes.

Using the estimated dynamic discrete choice model, we evaluate two education policies: a compulsory senior secondary school policy and a 50% college subsidy policy. Both policies increase educational attainment, but their distributional effects are very different. The compulsory school policy is effective for individuals from more disadvantaged backgrounds, whereas the college subsidy mainly benefits those from more advantaged backgrounds who have a comparative advantage in the schooling sector. We show that a model with fixed types ignores the indirect reward of education in acquiring better personality traits, which is empirically important to consider when evaluating the distributional effect of these policies. The policy response is greater in a model that allows types to change.

Our results highlight the importance of personality traits in explaining ex-ante heterogeneity, which, as was demonstrated in Keane and Wolpin (1997), is a major determinant of ex-ante life-time inequality. We find that one of the benefits of attending school is that it changes personality characteristics, which, along with increased schooling levels, enhances earnings. One caveat to our findings is that personality endowments are measured as of age 15. They likely reflect parental investment and life experience from conception to age 15. As emphasized in Cunha et al. (2010), the most cost effective policies for fostering the accumulation of non-cognitive skills, such as personality traits, may be policies that are targeted during early childhood years rather than high school or post-secondary schooling interventions.

2.11. Appendix: method used to impute initial age 15 personality traits

In many cases, sampled individuals are older than age 15, so we do not directly observe initial personality traits. The data contain up to three measures of personality traits, each measured at a time four years apart. We next describe the method that we use to impute the initial personality traits $\theta_n(15)$ based on these three measures, $\theta_n^{M1}(a_1), \theta_n^{M2}(a_2), \theta_n^{M3}(a_3)$, observed at ages a_1, a_2, a_3 and using the structure of our model. Given the current trial parameter values Ω, personality trait n at age 15 ($\theta_n(15)$) is obtained as follows:

1. From equation (2) in subsection 2.4.1, we solve for

$$\theta_n(a_1 - 1) = \theta_n(a_1) - (\gamma_{0n} + \gamma_{1n}(a_1 - 1 - 15) + \gamma_{2n} d_3(a_1) + \gamma_{3n}(a_1 - 1 - 15) d_3(a_1))$$

   where a_1 is the age when individual is surveyed and $d_3(a_1)$ is the indicator whether the individual is in school (alternative $m = 3$) at age a_1.

2. Substituting $\theta_n(a_1) = \theta_n^{M1}(a_1) - \zeta_n(a_1)$, we get

$$\theta_n(a_1 - 1) + \zeta_n(a_1) = \theta_n^{M}(a_1) - (\gamma_{0n} + \gamma_{1n}(a_1 - 16) + \gamma_{2n} d_3(a_1) + \gamma_{3n}(a_1 - 16) d_3(a_1))$$

3. Given $\theta_n(a_1 - 1) + \zeta_n(a_1)$, recover $\theta_n(a_1 - 2) + \zeta_n(a_1)$ following the same approach.

$$\theta_n(a_1 - 2) + \zeta_n(a_1) = (\theta_n(a_1 - 1) + \zeta_n(a_1)) - (\gamma_{0n} + \gamma_{1n}(a - 17) + \gamma_{2n} d_3(a_1 - 1) + \gamma_{3n}(a_1 - 17) d_3(a_1 - 1))$$

   Continue this way until we get $\theta_n^{M1}(15) \equiv \theta_n(15) + \zeta_n(a_1)$.

4. For the other two personality measurements at age a_2 and age a_3, ($\theta_n^{M2}(a_2)$ and $\theta_n^{M3}(a_3)$), repeat steps (1)-(3) to get

$$\theta_n^{M2}(15) \equiv \theta_n(15) + \zeta_n(a_2), \theta_n^{M3}(15) \equiv \theta_n(15) + \zeta_n(a_3)$$

5. This procedure provides three different imputed values of initial personality traits, each with a measurement error that is assumed to be mean zero. We obtain our measure of the personality trait at age 15 $\theta_n(15)$ as the mean of these three values:

$$\theta_n(15) = \frac{1}{3}(\theta_n^{M1}(15) + \theta_n^{M2}(15) + \theta_n^{M3}(15))$$

CHAPTER 3 : Personality Traits, Intra-household Allocation and the Gender Wage Gap

Christopher Flinn, Petra Todd and Weilong Zhang

3.1. Introduction

Early models of household decision-making specified a unitary model that assumed that a household maximizes a single utility function. (e.g. Becker (1981)) In recent decades, however, researchers have made substantial progress towards modeling the household as a collection of individual agents with clearly delineated preferences, which permits consideration of questions related to the production and distribution of household resources. The agents are united through the sharing of public goods, through joint production technologies for producing public goods, through shared resource constraints, and through preferences. One approach is the cooperative approach that allows differences between spouses to affect household decision-making by specifying a sharing rule or the Pareto weights of what is essentially a household social welfare function. Cooperative models assume that the household reaches Pareto efficient outcomes. Variations in the class of cooperative models specify different ways in which households reach a particular point on the Pareto frontier (e.g. Manser and Brown (1980), McElroy and Horney (1981), and Chiappori (1988)). An alternative approach assumes that household members act noncooperatively. This approach is also based on a model with individual preferences, but assumes that realized outcomes are determined by finding a Nash equilibrium using the reaction functions of the household members. These equilibria are virtually never Pareto efficient (e.g. Lundberg and Pollak (1993), Bourguignon (1984), Del Boca and Flinn (1995)).

In reality, it is likely that different households behave in different ways and even that the same household might behave differently at different points in time. One of the few studies to combine these different modeling approaches into one paradigm is Del Boca and Flinn (2012). Their study estimates a model of household time allocation, allowing for both efficient and inefficient household modes of interaction. In their model, two spouses allocate time to market work and to producing a public good and their decisions are repeated over an indefinitely long time horizon. The model incorporates incentive compatibility constraints that require the utility of each household member to be no lower that it would be in the (non cooperative) Nash equilibrium. Del Boca and Flinn (2012) find that the constraints are binding for many households and that approximately one-fourth of households behave in an inefficient manner.

This paper adopts a cooperative/noncooperative modeling framework similar to that of Del Boca and Flinn (2012), but our focus is on understanding the role of personality traits in affecting household time allocation decisions and labor market outcomes. Personality trait measures aim to capture "patterns of thought, feelings and behavior" that correspond to "individual differences in how people actually think, feel and act" (Borghans et al. (2008)). The most commonly used measures, which are the ones used in this paper, are the so-called Big Five. They measure individual openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism (the opposite of emotional stability).[1] The model we develop and estimate incorporates public and private goods consumption, labor supply at the extensive and intensive margins, and time allocated to home production. Personality traits operate as potential determinants of household bargaining weights and wage offers.

There is an increasing recognition that noncognitive traits play an important role in explaining a variety of outcomes related to education, earnings, and health. Heckman and Raut (2016) and Heckman et al. (2006) argue that personality traits may have both direct effects on an individual's productivity and indirect effects by affecting preferences for schooling or occupation choices. A study by Fletcher (2013) finds a robust relationship between personality traits and wages using sibling samples to control for family-level unobservables. Specifically, conscientiousness, emotional stability, extraversion and openness to

[1]The Big Five traits have the acronym OCEAN.

experience positively affect wages. Cubel et al. (2016) examine whether Big Five personality traits affect productivity using data gathered in a laboratory setting where effort on a task is measured. They find that individuals who exhibit high levels of conscientiousness and emotional stability perform better on the task.

Recent reviews of gender differences in preferences and in personality traits can be found in Croson and Gneezy (2009) and Bertrand (2011). Studies across many different countries find that women are on average more agreeable and more neurotic than men and that gender differences in personality are associated with differences in wages.[2] However, the most crucial traits in affecting wages differ by country. Using Dutch data, Nyhus and Pons (2005) find that emotional stability is positively associated with wages for both genders and agreeableness is associated with lower wages for women. Using data from the British Household Panel Study, Heineck (2011) analyzes correlations between Big Five personality traits and wages and finds a positive relationship between openness to experience and wages and a negative linear relationship between agreeableness and wages for men. He also finds a negative relationship between neuroticism and wages for women. Mueller and Plug (2006), using data from the Wisconsin Longitudinal Study, find that nonagreeableness, openness to experience and emotional stability are positively related to men's earnings, whereas conscientiousness and openness to experience are positively related to women's earnings. They find that the return that men receive for being nonagreeable is the most significant factor explaining the gender wage gap. Applying decomposition methods to data from the NLSY and using different measures of personality, Cattan (2013) finds that gender differences in self-confidence largely explain the gender wage gap, with the strongest effect being at the top of the wage distribution.[3] Braakmann (2009), using German Socioeconomic Panel (GSOEP) data, finds that higher levels of conscientiousness increase the probability of being full-time employed for both genders, while higher levels of neuroticism and agreeableness

[2] Women also exhibit differences in competitive attributes, risk aversion, preferences for altruism, and inequality aversion.

[3] The National Longitudinal Survey of Youth Data do not contain the Big Five personality trait measurements. NLSY measurements include a ten-item scale of self-esteem ((Rosenberg, 1965)) and a four-item scale of locus of control (Rotter (1966)).

have the opposite effect.

It is only recently that survey data have been collected on the personality traits of multiple household members for large random samples, which permits analysis of how personality traits affect marriage and the division of labor/resources within the household.[4] Lundberg (2012) notes that personality traits can shape preferences and capabilities that affect the returns to marriage and that they may also influence the ability of partners to solve problems and to make long-term commitments. Using data from the German Socioeconomic Panel (GSOEP), she finds that Big Five traits significantly affect the probability of marriage, the probability of divorce, and the duration of marriage. Using data from the Netherlands, Dupuy and Galichon (2014) show that Big Five personality traits are significant determinants of marriage matches and that different traits matter for men and women.

In this paper, we use a structural behavioral model to explore the extent to which personality traits of husbands and wives affect household time and resource allocation decisions. In particular, we examine how personality traits affect the mode of interaction the household adopts (cooperative or noncooperative), the amount of labor each spouse supplies to home production and market work, the provision of public goods, wage offers and accepted wages. Our analysis focuses on couples where the head of the household is age 30-50, because education and personality traits have largely stabilized by age 30. The model is static and takes the observed marriage sorting patterns with regard to spouse characteristics as given. In the model, spouses have their own preferences over consumption of a private good and a public good. They choose the amount of time to allocate to market work and to the production of a public good. There is a production technology that specifies how household members' time translates into public good production. The model incorporates household bargaining weights that may depend on the personality characteristics of both spouses, their education levels, ages,and cognitive abilities.[5]

[4]Examples include the British Household Panel Study (BHPS), the German Socioeconomic Panel (GSOEP), and Household Income and Labor Dynamics in Australia (HILDA), from which the data used in this paper are drawn.

[5]This formulation differs from Del Boca and Flinn (2012).

We use data from the Household Income and Labor Dynamics survey in Australia (HILDA). An unusual feature of these data is that they contain the Big Five personality measures at three points in time (over a span of eight years) for multiple household members. In addition to the personality trait measures, we also use information on age, gender, educational attainment, cognitive ability, wages, hours worked, and time spent engaging in home production.

Model parameters are estimated using the Method of Simulated Moments. The moments used in estimation pertain to wages, labor market hours, housework hours and labor force participation of different types of households. Model parameters are chosen to minimize the weighted distance between moments simulated using the model data generating process and moments based on the data.

We use the estimated model to analyze the determinants of male-female earnings differentials. The vast majority of papers in the gender earnings gap literature (e.g. Altonji and Blank (1999), Blau and Kahn (1997, 2006); Autor et al. (2008)) consider male and female earnings without taking into account that most adults are tied to individuals of the opposite sex through marriage or cohabitation and that these ties likely affect their decision-making. There are a few papers, however, that analyze male and female labor supply decisions and wage outcomes within a household framework. For example, Gemici (2011) analyzes household migration decisions in response to wage offers that males and females receive from different locations. Gemici and Laufer (2011) studies household formation, dissolution, labor supply, and fertility decisions. Tartari (2015) studies the relationship between children's achievement and the marital status of their parents within a dynamic framework in which partners decide whether to stay married, how to interact (with or without conflict), on labor supply and on child investments. Joubert and Todd (2016) analyze household labor supply and savings decisions within a collective household model, with a focus on the gender gap in pension receipt.

Within a household modeling framework, we analyze the manner in which households make decisions regarding whether a man or woman works in the labor market, how many hours they work, how many hours they devote to housework and the implications for earnings. Given the model's assumptions concerning male and female preferences, wage offer distributions, and the method of determining household allocations, we are able to assess the impact of individual and household characteristics not only on observed differences in wages but also the utility realizations of household members. Below, we will show that differences in utility levels of males and females inhabiting households together are more important indicators of systematic gender differences than are differences in observed wage rates.

Our analysis yields a number of potentially important findings. First, personality traits are significant determinants of household bargaining weights and of offered wages. Second, men and women have different traits on average and their traits are valued differently in the labor market as reflected in estimated wage offer equations. The combined effect of personality traits on offered wages is comparable in magnitude to the effect of education. Third, decomposition results show that gender differences in market valuations of personality traits explain a significant fraction of observed wage gaps. We find that if women were paid according to the male wage offer equation, the observed wage gap would be eliminated. Fourth, we find that the gender gap in accepted wages is smaller than the gap in offered wages. This difference arises because of the labor market participation decisions of husbands and wives, notably, because women are more selective than men in accepting employment. Fifth, we find that 38.7 percent of households choose to behave cooperatively, which also affects working decisions. Cooperation tends to increase the desired level of household public goods, which require both time and monetary investment, and therefore tends to increase labor supply for both men and women. Sixth, the marriage market exhibits positive assortative matching on personality traits, which tends to increase gender gaps in accepted wages relative to what it would be if spouses were randomly matched.

The paper proceeds as follows. The next section presents our baseline model. Section 3 describes the data. Section 4 discusses the econometric specification and estimation implementation. Section 5 and 6 present the estimation results and counterfactual experiments. Section 7 concludes.

3.2. Model

We begin by describing the preferences of the household members and the household production technology. Next, we describe the cooperative and noncooperative solutions to the model. The section concludes with an examination of the choice of the household members to behave cooperatively or not, and the potential role that personality traits play in this decision.

3.2.1. Preferences and Household Production Technology

A household is formed with a husband and a wife, distinguished by subscripts m and f, respectively. Each individual has a utility function given by

$$\begin{aligned} U_m &= \lambda_m \ln l_m + (1 - \lambda_m) \ln K \\ U_f &= \lambda_f \ln l_f + (1 - \lambda_f) \ln K, \end{aligned}$$

where λ_m and λ_f are both elements within $(0, 1)$, l_j denotes the leisure of spouse j $(j = m, f)$, and K is the quantity of produced public good. The household production technology is given by

$$K = \tau_m^{\delta_m} \tau_f^{\delta_f} M^{1-\delta_m-\delta_f},$$

where τ_j is the housework time of spouse j, δ_j is a Cobb-Douglas productivity parameter specific to spouse j, and M is the total income of the household. Income M depends on the labor income of both spouses as well as nonlabor income:

$$M = w_m h_m + w_f h_f + y_m + y_f,$$

Here, w_j is the wage rate of spouse j, h_j is the amount of time that the supply to the labor market, and y_j is their amount of nonlabor income. The time constraint of each spouse is given by

$$T = \tau_j + h_j + l_j,\ j = m, f.$$

A few comments are in order concerning this model specification. We have assumed that all of the choice variables relate to time allocation decisions, with no explicit consumption choice. This is standard since most data sets used by microeconomists contain fairly detailed information on labor market behavior and some information on housework, with little in the way of consumption data. We have made Cobb-Douglas assumptions regarding individual preferences and the household production technology. Because we assume that there exists heterogeneity in the preference parameters, λ_m and λ_f, and the production function parameters, δ_m and δ_f, we are able to fit patterns of household behavior very well, even under these restrictive functional forms.[6]

To this point, we have largely followed Del Boca and Flinn (2012); Del Boca et al. (2014). Our points of departure are the addition of personality traits to their formulation, the addition of working decisions, and the addition of wage offer equations to the model. Del Boca and Flinn (2012) restricted their sample to include only households in which both spouses work and they simply conditioned on husbands' and wives' observed wages. Because one of the main focuses of our analysis is to examine the impact of personality traits on household behavior and on a woman's labor market participation decision, it is necessary for us to estimate wage equations for both husbands and wives. Let x_j denote observable characteristics of spouse j and θ_j the personality characteristics of spouse j. Then a household is

[6] Del Boca and Flinn (2012) actually estimate the distribution of the individual characteristics nonparametrically, and show that by doing so the model is "saturated." That is, there are the same number of free parameters as there are data points. Model fit is perfect in such a case. For the purposes of this exercise, we assume that these characteristics follow a parametric distribution, but we utilize one that is flexible and capable of fitting patterns in the data quite accurately.

characterized by the state vector

$$S_{m,f} = (\lambda_{m,}\delta_m, w_m, y_m, \theta_m, x_m) \bigcup (\lambda_f, \delta_f, w_f, y_f, \theta_f, x_f).$$

Given $S_{m,f}$, either mode of behavior is simply a mapping

$$(\tau_m, h_m, l_m, \tau_f, h_f, l_f) = \Psi_E(S_{m,f}), \;\; E = NE, PW$$

where $E = NE$ is the (noncooperative) Nash equilibrium case and $E = PW$ is the (cooperative) Pareto weight case. We note that each spouses' wage offer w_j is observed by the household, but the analyst will not observe w_j if $h_j = 0$. Certain elements of $S_{m,f}$ may not play roles in the determination of equilibrium outcomes in certain behavioral regimes.

In our static model, couples do not have an option to get divorced. With an additional divorce option, couples might choose to cooperate when their utility from cooperation exceeds the utility from the inefficient Nash equilibrium and the utility from divorce. Of course there are many additional considerations other than current period utility in modeling divorce decisions, such as the division of assets upon divorce, the presence of children, child support, alimony and the state of the marriage market. For the sake of simplicity, our model focuses on married couples without considering divorce, which may to some extent limit external validity.

We now turn to a detailed description of the non-cooperative and cooperative solutions.

3.2.2. Non-Cooperative Behavior

In the noncooperative regime, the nature of interaction between the spouses is limited and personality characteristics only play a role through their effects on wage offers. Under modeling assumptions that are the same as ours, Del Boca and Flinn (2012) show that there exists a unique equilibrium solution in reaction functions, at least in the cases in which spouses are both in or both out of the labor market.[7] Because ours is a model of complete information, each spouse is fully aware of the other's preferences, productivity characteristics, wage offer, and non-labor income. The decisions made by each spouse are best responses to the other spouse's choices, and are (most often) unique and stable. In this environment, little interaction between the spouses is required.

Each spouse makes three time allocation choices. Because they must sum to T, it is enough to describe the equilibrium in terms of each spouse's choices of labor supply and housework time. The reaction functions given the state vector $S_{m,f}$ are

$$\begin{aligned} \{h_m(NE), \tau_m(NE)\}(h_f, \tau_f; S_{m,f}) &= \arg\max_{h_m,\tau_m} \lambda_m \ln l_m + (1-\lambda_m) \ln K \\ \{h_f(NE), \tau_f(NE)\}(h_m, \tau_m; S_{m,f}) &= \arg\max_{h_f,\tau_f} \lambda_f \ln l_f + (1-\lambda_f) \ln K, \end{aligned}$$

where

$$K = \tau_m^{\delta_m} \tau_f^{\delta_f} (w_m h_m + w_f h_f + y_m + y_f)^{1-\delta_m-\delta_f}$$

For $\lambda_j \in (0,1)$, $j = m, f$, and $0 < \delta_m$, $0 < \delta_f$, and $\delta_m + \delta_f < 1$, Del Boca and Flinn (2012) show that there is a unique equilibrium for their case in which both spouses in the households supply labor to the market. However, if we remove the constraint that the Nash equilibrium always results in both spouses choosing to supply a positive amount of time to the labor market, the possibility of multiple equilibria arises. The multiple equilibria occur due to the constraint that working hours are nonnegative for both spouses. There can be at most two Nash equilibria, with each having only one of the spouses supplying a positive amount of time to the market, and the other in which the spouses switch roles in terms of who is supplying time to the market and who is not. When both supply time to the market, the equilibrium is unique, as it is when neither supplies time to the market. Furthermore, it is the case that when one supplies time to the market and the other does not, the equilibrium may either be unique or not. Given the structure of the model and the

[7] BecauseDel Boca and Flinn (2012); Del Boca et al. (2014) conditioned their analysis on the fact that both spouses were in the labor market, the noncooperative solution was unique.

estimated parameters, the frequency of multiple equilibria is small. However, when they do occur, a position must be taken as to which of the two equilibria are selected. We will follow convention and assume that the equilibrium in which the male participates and the female does not is the one selected.[8] A detailed description of how the non-cooperative equilibrium is computed and selected is provided in Appendix 3.8.2.

The utility value of this equilibrium to spouse j is given by

$$V_j(NE) = \lambda_j \ln(T - h_j(NE) - \tau_j(NE)) + (1 - \lambda_j) \ln K(NE),\ j = m, f,$$

with

$$K(NE) = \tau_m(NE)^{\delta_m} \tau_f(NE)^{\delta_f} (w_m h_m(NE) + w_f h_f(NE) + Y_m + Y_f)^{1-\delta_m-\delta_f},$$

where we have suppressed the dependence of the equilibrium outcomes on the state vector $S_{m,f}$ to avoid notational clutter.

3.2.3. Cooperative Behavior

The Benthamite social welfare function for the household with the Pareto weight α is given by

$$\begin{aligned} W(h_m, h_f, \tau_m, \tau_f; S_{m,f}) = {} & \alpha(S_{m,f}) U_m(h_m, h_f, \tau_m, \tau_f; S_{m,f}) \\ & +(1 - \alpha(S_{m,f})) U_f(h_m, h_f, \tau_m, \tau_f; S_{m,f}), \end{aligned}$$

where we have eliminated the leisure choice variable $l_j, j = f, m$ by imposing the time constraint. The Pareto weight $\alpha(S_{m,f}) \in (0, 1)$, and, as the notation suggests, will be allowed to be a function of a subset of elements of $S_{m,f}$. In the cooperative (efficient)

[8] Alternatively, one could allow the selection mechanism to depend on personality characteristics. However, our estimation results indicate multiple equilibria rarely occur (9 out of 1443 households). Thus, the selection mechanism is unlikely to play a major role in the estimation.

regime, the household selects the time allocations that maximize W, or

$$(h_m, h_f, \tau_m, \tau_f)(S_{m,f}) = \arg \max_{h_j, \tau_j, j=m,f} W(h_m, h_f, \tau_m, \tau_f; S_{m,f}).$$

Because this is simply an optimization problem involving a weighted average of two concave utility functions, the solution to the problem is unique. Then the utility levels of the spouses under cooperative behavior is

$$V_j(PW) = \lambda_j \log(T - h_j(PW) - \tau_j(PW)) + (1 - \lambda_j) \log K(PW),\ j = m, f,$$

with

$$K(PW) = \tau_m(PW)^{\delta_m} \tau_f(PW)^{\delta_f} (w_m h_m(PW) + w_f h_f(PW) + y_m + y_f)^{1-\delta_m-\delta_f}.$$

Once again, we have suppressed the dependence of solutions on the state variable vector $S_{m,f}$. In the cooperative model, there is no danger of multiple equilibria, since it is not really an equilibrium specification at all, but simply a household utility-maximization problem.

3.2.4. Selection Between the Two Allocations

Del Boca and Flinn (2012) constructed a model in which the Pareto weight, α, was "adjustable" so as to satisfy a participation constraint for each spouse that enforced

$$V_j(PW) \geq V_j(NE),\ j = m, f.$$

With no restriction on the Pareto weight parameter α, the $V_j(PW)$ could be less than $V_j(NE)$ for one of the spouses (it always must exceed the noncooperative value for at least one of the spouses). For example, if $V_m(PW) < V_m(NE)$, the husband has no incentive to participate in the "efficient" outcome, because he is worse off under it. To give him enough incentive to participate, the value of α, which is his weight in the social welfare function, is increased to the level at which he is indifferent between the two regimes. Meanwhile, his spouse with the "excess" portion of the household surplus from cooperation has to cede some of her surplus by reducing her share parameter, $(1-\alpha)$, in this case, to the point at which the husband is indifferent between the two regimes.

In such a world, and in a static context, an efficient outcome could always be achieved through adjustment of the Pareto weight, α. As a result, all households would behave cooperatively. To generate the possibility that some households would behave noncooperatively, even when able to adjust α, Del Boca and Flinn assumed a pseudo-dynamic environment, in which the spouses played the same (static) stage game an infinite number of times. They assumed a grim-trigger punishment strategy, so that any deviation from the agreed upon cooperative outcome by either spouse in any period results in a punishment state in which the Nash equilibrium is played in perpetuity. In such a case, the value of the discount factor, $\beta \in [0,1)$, used to weight future rewards, is critical in determining whether a cooperative arrangement is implementable. The value to individual j of playing the cooperative outcome forever is simply $V_j(PW)(1+\beta+\beta^2+...) = V_j(PW)/(1-\beta)$. The present value of the noncooperative outcome is $V_j(NE)/(1-\beta)$. If individual j cheats on the agreement in any period and the spouse does not, the value of cheating in the period is denoted by $V_j(C)$, and it is straightforward to show that $V_j(C) > V_j(PW)$. By cheating in any period, the individual knows that the spouse will not cooperate in all future periods, so that the gain to cheating (assuming the spouse does not) is

$$V_j(C) + \beta\frac{V_j(NE)}{1-\beta},$$

whereas the gain from playing cooperatively throughout (assuming that the spouse does as well) is

$$\frac{V_j(PW)}{1-\beta}$$

Any implementable agreement will have $V_j(PW) > V_j(NE)$ for each j. We can define a critical discount factor β_j^* as one that equates the value of cheating with not cheating in any period for individual j, and this critical value is given by

$$\beta_j^* = \frac{V_j(C) - V_j(PW)}{V_j(C) - V_j(NE)},$$

where it follows that $\beta_j^* \in [0, 1)$. There will be a reallocation (characterized by a value of α) of the cooperative surplus for which the two critical discount factors are equal, and we define this common value as $\tilde{\beta}$, where $\tilde{\beta} = \beta_1^* = \beta_2^*$. Del Boca and Flinn show that if all individuals in the population share the same discount factor, β, then a given household will be able to implement the cooperative outcome if and only if $\beta \geq \tilde{\beta}$. The intuition is fairly straightforward. If individuals are myopic, they give excessive weight to the potential gains from cheating now and far less weight to the costs they will incur in the future by being in the noncooperative regime forever. Both spouses have to be sufficiently forward-looking to be able to implement the cooperative agreement in this simple dynamic setting. Del Boca and Flinn (2012) estimate a common discount factor β, and find that approximately 25 percent of households in their sample behaved in a noncooperative manner given their parameter estimates.

In our model, which focuses on the role of personality traits in explaining wage and welfare differences between husbands and wives, we think of the Pareto weight α as being determined, in part, by the personality characteristics of the husband and wife. For example, someone who is very agreeable and who is married to a nonagreeable person might receive a lower Pareto weight. In this case, it is somewhat problematic to assume that the α can be freely adjusted to satisfy the participation constraint of one of the spouses and more reasonable to assume that α is fixed for each household. Fixed pareto weights simplifies the cooperative versus noncooperative decision of the household, as well as the computation of the model. In this set-up, personality characteristics of both spouses are potentially key factors in how they settle on a particular mode of behavior.[9]

[9] In a more elaborate model, we could imagine a situation in which the Pareto weight could be adjusted, but with a cost depending on the personality characteristics of the spouses. From this perspective, we are assuming that the costs of adjusting the Pareto weight are indefinitely large for one or both of the spouses.

A household will behave cooperatively if and only if both of the following weak inequalities hold:

$$\begin{aligned} V_m(PW) &\geq V_m(NE) \\ V_f(PW) &\geq V_f(NE). \end{aligned}$$

Thus, there is no scope for "renegotiation" in this model. There is a positive probability that any household behaves cooperatively that is strictly less than one given our preference heterogeneity specification. The simplest way to characterize the cooperation decision in our framework is as follows. We begin by explicitly including the value of α in the cooperative payoff function for household j, so that

$$V_j(PW|S_{m,f}, \alpha), \ j = m, f.$$

Given that the function $V_m(PW|S_{m.f}, \alpha)$ is monotonically increasing in α and given that $V_f(PW|S_{m,f}, \alpha)$ is monotonically decreasing in α, we can define two critical values, $\underline{\alpha}^*(S_{m,f})$ and $\overline{\alpha}^*(S_{m,f})$ such that

$$\begin{aligned} V_m(PW|S_{m,f}, \underline{\alpha}^*(S_{m,f})) &= V_m(NE|S_{m,f}) \\ V_f(PW|S_{m,f}, \overline{\alpha}^*(S_{m,f})) &= V_f(NE|S_{m,f}). \end{aligned}$$

The set of α values that produce cooperative behavior in the household is connected, so that the household will behave cooperatively if and only if

$$\alpha(S_{m,f}) \in [\underline{\alpha}^*(S_{m,f}), \overline{\alpha}^*(S_{m,f})].$$

For a given value of the state variables, $S_{m,f}$, the household will either behave cooperatively or not; there is no further stochastic element in this choice after we have conditioned on $S_{m,f}$. The probabilistic nature of the choice is due to the randomness of $S_{m,f}$. Although some elements of $S_{m,f}$ are observable (and do not include measurement error under our assumptions), others are not. There are a subset of elements that are not observed for any household, which include the preference and household production parameters. We denote the set of unobserved household characteristics by $S^u_{m,f} = \{\lambda_m, \delta_m, \lambda_f, \delta_f\}$, with the set of (potentially) observed characteristics given by $S^o_{m,f} = \{w_m, y_m, \theta_m, x_m, w_f, y_f, \theta_f, x_f\}$. We say that these elements are all potentially observable because the wage offers, w_j, $j = m, f$, are only observed if spouse j supplies a positive amount of time to the labor market. The state variable vector $S^o_{m,f}(i)$ that is observed for household i will have a degenerate marginal distribution. The unobserved vector $S^u_{m,f}(i)$ will always have nondegenerate marginal distributions. Let the distribution of $S^u_{m,f}(i)$ be given by G_i, and assume that $G_i = G$ for all i. Then the probability that household i is cooperative is simply the measure of the set of $S^u_{m,f}(i)$ such that the cooperation condition is satisfied, or

$$P(PW|S^o_{m,f}(i)) = \int \chi[\underline{\alpha}^*(S_{m,f}(i)) \leq \alpha(S_{m,f}(i)) \leq \overline{\alpha}^*(S_{m,f}(i))]dG(S^u_{m,f}).$$

For any household i, $0 < P(PW|S^u_{m,f}(i)) < 1$, due to what is essentially a full support condition. The preference weight on leisure for spouse j lies in the interval $(0, 1)$. As $\lambda_j \to 1$, spouse j only cares about leisure and gives no weight to the public good. In the Nash equilibrium, their contribution to household production through time and money will converge to 0, and the cooperative solution, which results in greater production of the public good, will be of no value to them. As $\lambda_j \to 0$, the individual will demand little leisure and will spend all of their time in the labor market and household production. For cases in which λ_m and λ_f are both arbitrarily close to 1, the household will be noncooperative. For cases, in which λ_m and λ_f are close to 0, the household will be cooperative. Thus, independently of the other values in the state vector, variability in the preference parameters on the full support of their (potential) distribution is enough to guarantee that no household can be deterministically classified as cooperative *a priori*.

3.3. Data Description

3.3.1. Selection of the Estimation Sample

We use sample information from the Household Income and Labour Dynamics in Australia (HILDA) longitudinal data set. HILDA is a representative one in one thousand sample of the Australian population. It is an ongoing longitudinal annual panel starting in the year 2001 with 19,914 initial individuals from 7,682 households. (Summerfield et al. (2015)) Our paper makes use of the following variables: (1) labor market outcomes including annual labor earnings and working hours; (2) housework split information; (3) self-completion life style questions including a question about perceived fairness of the housework arrangement; (4) education levels; (5) cognitive test scores on three tests and (6) the "Big Five" personality traits assessment (collected three times, in waves 5, 9, and 13).

To the best of our knowledge, HILDA has the highest quality information on personality traits among all nationwide data sets.[10] For the majority of respondents, we observe three repeated measurements of personality traits over an eight-year time window. As described in Section 1, the personality trait measurements are based on the Five Factor ("Big Five") Personality Inventory, which classifies personality traits along five dimensions: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability (John and Srivastava (1999)). "Big Five" information in HILDA is constructed by using responses to 36 personality questions, which are shown in table 40.[11] Respondents were asked to pick a number from 1 to 7 to assess how well each personality adjective describes them. The lowest number, 1, denotes a totally opposite description and the highest number, 7, denotes a perfect description. According to Losoncz (2009), only 28 of 36 items load well into their corresponding components when performing factor analysis. The other 8 items

[10] The only other two nation-wide data sets providing personality traits inventory assessments are the German Socio-Economic Panel (GSOEP) study and the British Household Panel (BHPS) study. Both of them also collected "Big Five" measures.

[11] The source of these 36 adjectives come from two parts. Thirty of them are extracted from Trait Descriptive Adjectives - 40 proposed by Saucier (1994), which is a selected version of Traits Descriptive Adjective - 100 (Goldberg (1992b)) to balance the time use and accuracy. And the other additional six items come from various sources.

are discarded due to either their low loading values or their ambiguity in defining several traits.[12] Our construction of the "Big Five" follows the procedure provided by Losoncz (2009). We include all individuals who have at least one personality trait measurement. For the individuals whose personality traits are surveyed in multiple waves, we use the average value.

In addition to the information on personality traits, HILDA also collected information on cognitive ability once in wave 12.[13] We construct a one-dimensional measure from three different measures: (i) Backward Digits Span, (ii) Symbol Digits Modalities and (iii) a 25-item version of the National Adult Reading Test. We construct a single measure by first standardizing each of the three measures and then taking the mean.

The repeated measures of personality traits for the same person during eight-year window allow us to explore how the personality traits evolve over the life-cycle. Following Cobb-Clark and Schurer (2012), we define the mid-term change as the change in reported traits between 2005 and 2009 and the long-term change as the change between 2005 and 2013. The changes range from -6 to 6. Table 41 reports summary statistics for mid-term and long-term changes. Personality trait changes are approximately normally distributed with a mean of 0 and a standard deviation of around 0.80. The majority of individuals (more than 70%) experience changes in their personality traits within one standard deviation. Figure 20 shows the mean midterm changes in personality traits by age. The figures show that traits are more malleable at younger ages. For example, the average change in conscientiousness is above 0 before age 30 and close to 0 after that. We perform an F-test of whether changes in personality traits are independent of age for individuals age 30-50 and do not reject the null. However, the null is rejected with p-value less than 0.001 when the age group is expanded to ages 15-50. The observed pattern is consistent with other evidence from the psychology literature that personality traits stabilize with age. For example, Terracciano et al. (2006)

[12]The way to check each item's loading performance is to calculate the loading value after doing oblimin rotation. The loading values of 8 abandoned items were either lower than 0.45, or did not load more than 1.25 times higher on the expected factor than any other factor.

[13]According to the report of Wooden (2013), the response rate is high, approximately 93%.

Table 40: Personality traits questionnaire

B19 How well do the following words describe you? For each word, cross one box to indicate how well that word describes you. There are no right or wrong answers.

*(Cross ☒ **one** box for **each** word.)*

	Does not describe me at all 1	2	3	4	5	6	Describes me very well 7
talkative	1	2	3	4	5	6	7
sympathetic	1	2	3	4	5	6	7
orderly	1	2	3	4	5	6	7
envious	1	2	3	4	5	6	7
deep	1	2	3	4	5	6	7
withdrawn	1	2	3	4	5	6	7
harsh	1	2	3	4	5	6	7
systematic	1	2	3	4	5	6	7
moody	1	2	3	4	5	6	7
philosophical	1	2	3	4	5	6	7
bashful	1	2	3	4	5	6	7
kind	1	2	3	4	5	6	7
inefficient	1	2	3	4	5	6	7
touchy	1	2	3	4	5	6	7
creative	1	2	3	4	5	6	7
quiet	1	2	3	4	5	6	7
cooperative	1	2	3	4	5	6	7
sloppy	1	2	3	4	5	6	7

	Does not describe me at all 1	2	3	4	5	6	Describes me very well 7
jealous	1	2	3	4	5	6	7
intellectual	1	2	3	4	5	6	7
extroverted	1	2	3	4	5	6	7
cold	1	2	3	4	5	6	7
disorganised	1	2	3	4	5	6	7
temperamental	1	2	3	4	5	6	7
complex	1	2	3	4	5	6	7
shy	1	2	3	4	5	6	7
warm	1	2	3	4	5	6	7
efficient	1	2	3	4	5	6	7
fretful	1	2	3	4	5	6	7
imaginative	1	2	3	4	5	6	7
enthusiastic	1	2	3	4	5	6	7
selfish	1	2	3	4	5	6	7
careless	1	2	3	4	5	6	7
calm	1	2	3	4	5	6	7
traditional	1	2	3	4	5	6	7
lively	1	2	3	4	5	6	7

Table 41: Summary statistics of personality traits and their changes over time

	Level		Mid-term change between 05-09					Long-term change between 05-13				
	Mean	S.D.	Mean	S.D.	10th	50th	90th	Mean	S.D.	10th	50th	90th
Openness	4.27	1.03	-0.074	0.776	-1.000	0	0.833	-0.004	0.817	-1.000	0	1.000
Conscientiousness	5.15	1.02	0.028	0.772	-0.833	0	1.000	0.106	0.824	-0.833	0	1.167
Extraversion	4.42	1.08	-0.036	0.744	-1.000	0	0.833	-0.049	0.786	-1.000	0	0.833
Agreeableness	5.40	0.88	-0.010	0.761	-1.000	0	1.000	0.087	0.794	-1.000	0	1.000
Emotional Stability	5.19	1.07	0.092	0.870	-1.000	0	1.167	0.100	0.919	-1.000	0	1.167

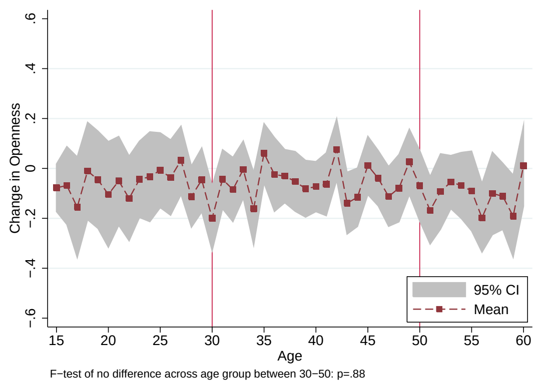
Data come from wave 2005, 2009 and 2013. The sample consists of 6,330 individual observations (2,913 males and 3,417 females) with three completed repeated measures. S.D.= standard deviation.

and Terracciano et al. (2010) report that intra-individual consistency increases up to age 30 and thereafter stabilizes.

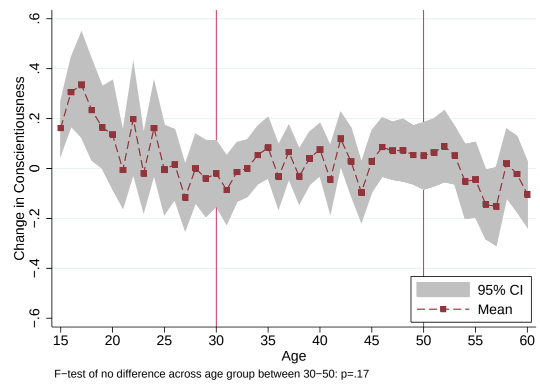
We focus our attention on households whose heads are between the ages 30 and 50 for two reasons. First, household structure may change during earlier ages due to marriage and fertility. Second, as noted, personality traits stabilize after age 30. Thus we can reasonably treat a spouse's personality traits as being fixed after age 30. We drop households for which housework information, labor market information or personality traits are missing. Among 3151 intact households with complete information and with the husband and wife present, 1881 of them have at least one period in which the household head is age 30-50 when surveyed. When a household has multiple qualifying periods, we randomly select one observation period. The hourly wage is calculated by dividing annual earnings by annual working hours. We truncated the top five percent of hourly wage rates to eliminate unrealistically high values. We set the total time available for leisure, housework, and labor supply in a week, T, to 116. Working time has an upper bound of 60 hours while housework has an upper bound of 56 hours.

In general, housework time can be divided into two components: time spent with children and other activities, such as cleaning house, cooking or running errands. Women with younger children are most likely to have their labor supply choices influenced by children. Because our model does not explicitly account for time spent child-rearing, we restrict our estimation sample to only include families that do not have very young children. In table 42, we examine the effects of this restriction by comparing three alternative samples with

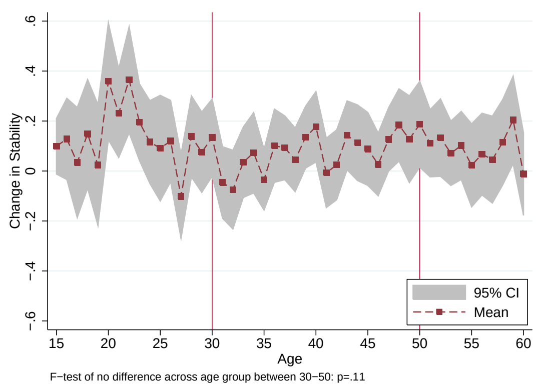
Figure 20: Changes in Big-Five personality over the life-cycle (age 15-60)



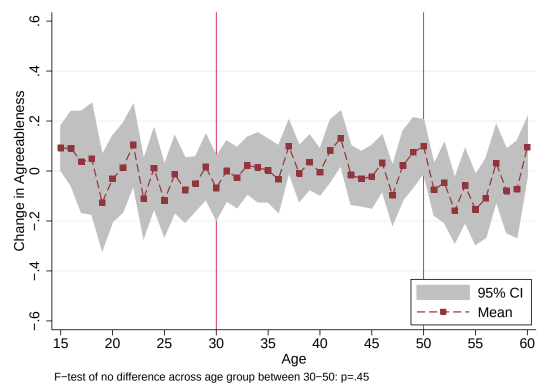
(a) Openness to experience



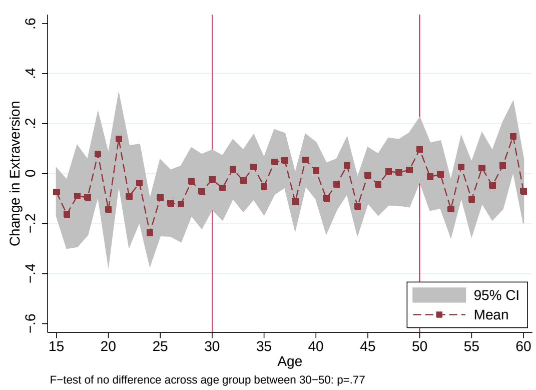
(b) Conscientiousness



(c) Emotional stability



(d) Agreeableness



(e) Extraversion

different age selection criteria. The first sample does not impose any age restriction, the second sample excludes households with any child below age 8 and the third sample excludes families with a child below age 14. The sample size shrinks from 1,881 to 1,443 and 973 with the more stringent age restrictions. However, the average age of husbands is around 40 and the age of wives around 38 in all three samples. The labor market participation of husbands is also similar across the three samples. As is typically found, husbands spend more time in the labor market than do their wives. The employment rate for males is 94% and the average number of working hours (conditional on working) is around 44 hours per week.

The key differences across the three samples are observed in female labor market participation and reported housework time. The average housework hours of husbands decreases from 23.11 hours in sample 1 to 18.19 hours in sample 2 and 14.92 hours in sample 3. The average housework hours of wives decreases from 43.27 hours in sample 1 to 27.87 hours in sample 2 and 20.27 hours in sample 3. These decreases are mainly caused by the reduction of the time spent with children. As shown in table 42, the average time spent with children is 9.12 hours for husbands and 20.89 hours for wives. The time spent with children shrinks to 4.10 and 7.15 hours in sample 1 to 1.47 and 1.89 hours in sample 2. Wives with older children spend fewer hours caring for children and have a higher labor force participation rate, both at the intensive and extensive margins. The employment rate increases from 72% in sample 1 to 85% in sample 1 and 88% in sample 3. The average working hours increases from 30.16 hours in sample 1 to 33.48 hours in sample 2 and 36.29 in sample 3.

In all three samples, the distributions of wives' working hours and housework hours are more dispersed than that of husbands'. Their accepted wages are also lower. In general, the time allocations described in our paper using the HILDA dataset are consistent with patterns described in Del Boca and Flinn (2012); Del Boca et al. (2014) for US Panel Study of Income Dynamics (PSID) data (the 2005 wave).

We use the sample of households without children below age 8 (sample 2) as the primary estimation sample. However, for comparison purposes, we also provide in the appendix estimation results based on the more restricted sample (sample 3).

With regard to personality traits and cognitive ability, Table 3 shows significant gender differences but similar patterns across the three samples. On average, men have lower scores on agreeableness, extraversion and conscientiousness compared with women. Gender differences in openness to experience and emotional stability are less significant. In our sample, wives have higher cognitive scores than husbands.

We estimate a preliminary OLS regression to examine the relationship between measured personality traits and log wages (for those who were working) and their relationship with labor participation decisions. The regression results with log wage as the dependent variable are shown in the first two columns in Table 43. We find for both that men and women that education and cognitive ability increase earnings. In addition, conscientiousness increases wages for men. In the last two columns in Table 43, the dependent variable is labor force participation. Higher education and cognitive scores are associated with higher rates of labor force participation for both men and women. Openness to experience tends to decrease labor force participation for both men and women. Conscientiousness is associated with higher rates of participation but only for women.

3.3.2. Assortative Matching of Personality Traits

Although our paper does not explicitly model the marriage market, we are able to examine marital sorting on personality traits and cognitive scores in our sample. Figure 21 displays the scatter plots of spousal personality traits as well as cognitive abilities. We observe a strong positive assortative matching in the cognitive ability dimension with a correlation equal to 0.34. Among the "Big Five" personality traits, emotional stability and openness to experience are the traits that exhibit the most significant pattern of positive sorting (correlation larger than 0.1), whereas agreeableness has a less strong positive sorting pattern. There is no significant correlation in the extraversion and conscientiousness traits of

Table 42: Key variables in HILDA, means and (standard errors)

	Sample 1		Sample 2		Sample 3	
	Full sample		Excluded households with dependents < 8		Excluded households with dependents < 14	
Variable	Male	Female	Male	Female	Male	Female
Age	40.00	37.86	41.17	39.20	40.04	38.12
	(6.41)	(7.21)	(6.56)	(7.65)	(7.30)	(8.72)
Employment	0.94	0.72	0.94	0.85	0.94	0.88
	(0.24)	(0.45)	(0.24)	(0.35)	(0.24)	(0.32)
Hourly Wage	30.10	25.49	29.52	24.85	29.06	25.33
	(15.02)	(13.91)	(14.60)	(12.46)	(15.99)	(14.63)
Working hours	44.08	30.16	44.10	33.48	44.03	36.29
	(9.84)	(13.25)	(9.54)	(12.31)	(9.70)	(13.14)
Housework	23.11	43.27	18.19	27.87	14.92	20.27
	(15.83)	(25.54)	(12.86)	(17.59)	(10.62)	(13.48)
Time with Children	9.12	20.89	4.10	7.15	1.47	1.89
	(9.89)	(23.64)	(6.49)	(11.16)	(3.64)	(5.83)
Education	13.22	13.31	13.16	13.22	13.17	13.37
	(2.38)	(2.38)	(2.38)	(2.40)	(2.37)	(2.44)
Other Income	335.0	-	333.7	-	311.9	-
	(251.4)	-	(256.9)	-	(251.0)	-
Obs.	1,881	1,881	1,443	1,443	973	973
Average values of personality traits and cognitive ability						
Cognition	0.13	0.23	0.12	0.23	0.13	0.22
	(0.70)	(0.66)	(0.70)	(0.65)	(0.70)	(0.66)
Openness	4.37	4.24	4.39	4.23	4.41	4.27
	(0.96)	(0.98)	(0.96)	(0.99)	(0.95)	(1.00)
Conscientiousness	5.09	5.29	5.11	5.34	5.12	5.38
	(0.94)	(1.00)	(0.93)	(0.98)	(0.94)	(0.98)
Extraversion	4.32	4.65	4.29	4.63	4.31	4.61
	(0.99)	(1.13)	(1.00)	(1.15)	(0.99)	(1.14)
Agreeableness	5.22	5.74	5.20	5.73	5.20	5.70
	(0.85)	(0.76)	(0.85)	(0.76)	(0.85)	(0.77)
Emotional Stability	5.14	5.18	5.15	5.19	5.14	5.15
	(1.00)	(0.99)	(0.97)	(0.98)	(0.96)	(0.99)

(1) Both employment and hourly wage are conditional on being employed. (2) The other income is the pooled household income other than labor earnings.

Table 43: The Effects of Personality on Earnings and Labor Market Participation

	Log Hourly Earning		Labor Market Participation	
	Males	Females	Males	Females
Openness	-0.027	-0.013	-0.023***	-0.040***
	(0.015)	(0.015)	(0.008)	(0.010)
Conscientiousness	0.050***	0.027	0.013	0.034***
	(0.015)	(0.015)	(0.007)	(0.010)
Extraversion	0.009	0.023	0.005	0.019
	(0.013)	(0.012)	(0.006)	(0.008)
Agreeableness	-0.040*	-0.041*	0.006	0.001
	(0.017)	(0.020)	(0.008)	(0.013)
Stability	0.011	-0.006	-0.002	-0.014
	(0.015)	(0.016)	(0.007)	(0.010)
Education	0.053***	0.065***	0.007*	0.023***
	(0.006)	(0.006)	(0.003)	(0.004)
Cognition	0.124***	0.064**	0.057***	0.067***
	(0.020)	(0.022)	(0.010)	(0.015)
Exp	0.031**	0.016*	0.004	0.007
	(0.010)	(0.006)	(0.006)	(0.004)
$Exp^2/100$	-0.080***	-0.045**	-0.010	-0.020
	(0.024)	(0.023)	(0.016)	(0.012)

*Notes:*Robust standard errors are reported in parentheses. The potential experience is calculated by "Age - 6 - education years" as a standard approach in the literature.

* Statistically significant at the .05 level; ** at the .01 level; *** at the .001 level.

Table 44: Responses to the “fair share” question

Percentage(%)	Male	Female
Much More(1)	1.80	19.89
A bit More(2)	12.54	37.98
Fair Share(3)	58.21	36.87
A bit Less(4)	25.09	4.78
Much Less(5)	2.36	0.49
Correlation	-0.379	

husbands and wives. There is also strong positive marital sorting on cognitive ability.

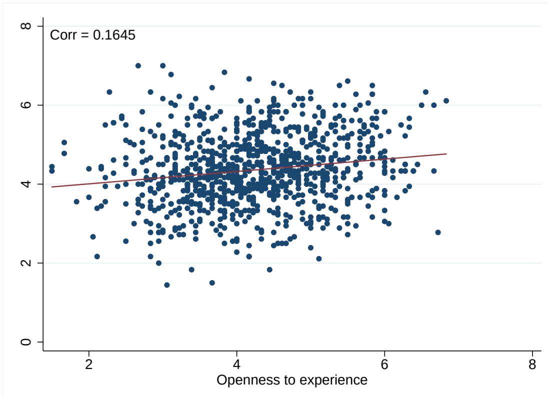
3.3.3. Other Variables: Fair Share

Despite the important role played by the by the Pareto weight in cooperative models of the household, there is no direct measurement proposed in the literature. The HILDA data provide a fairly high quality record of household activities. We consider the following question, completed by the respondent in the self-completion portion of the questionnaire, to be potentially related to the household allocation rule: “Do you think you do your fair share around the house?” The respondent has the option of choosing: (1) I do *much more* than my fair share. (2) I do *a bit more* than my fair share. (3) I do my fair share. (4) I do *a bit less* than my fair share. (5) I do *much less* than my fair share.

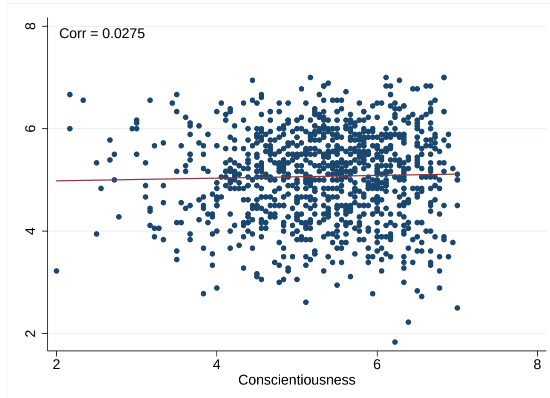
The distribution of fair share choices for both men and women is shown in table 44. The majority of husbands report that they do a fair share of housework, while the majority of wives report doing more than their fair share. The significantly negative correlation between men and women’s report indicates that a better condition for the husband implies a worse condition for the wife, consistent with a Pareto weight interpretation.

We do not make direct use of the fair share variable in estimation. Rather, as described below, we will examine how simulations based on the estimated model relate to the fair share variable as a way of examining support for the model.

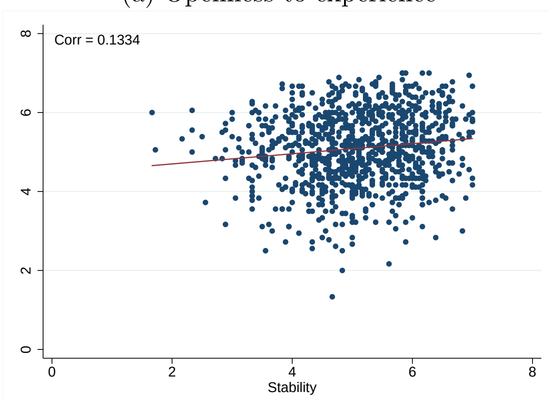
Figure 21: Assortative matching of personality traits and cognitive ability



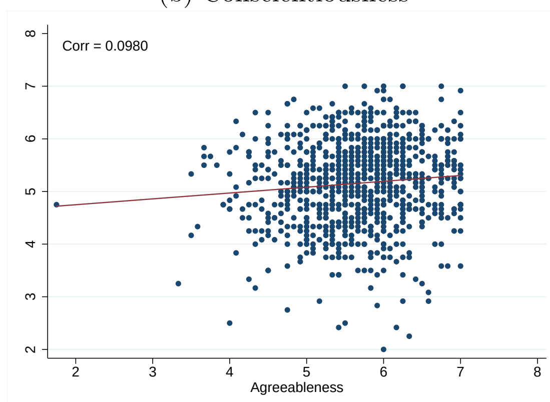
(a) Openness to experience



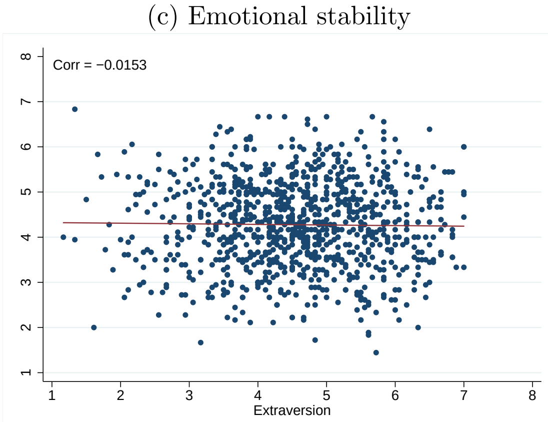
(b) Conscientiousness



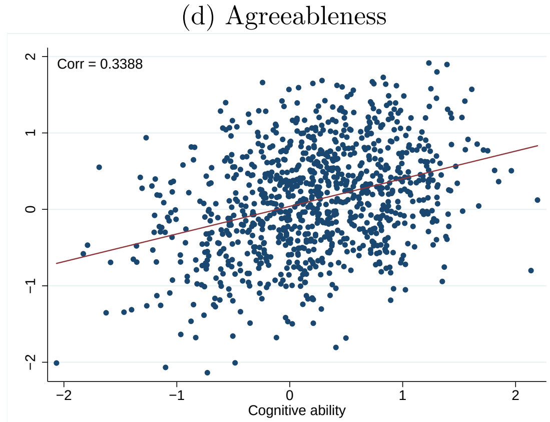
(c) Emotional stability



(d) Agreeableness



(e) Extraversion



(f) Cognitive ability

3.4. Econometric Implementation

As previously noted, a household i is uniquely characterized by the vector $S_{m,f}(i) = (\lambda_{im}, \delta_{im}, w_{im}, y_{im}, \theta_{im}, exp_{im}, edu_{im}, c_{im}, a_{im}) \bigcup (\lambda_{if}, \delta_{if}, w_{if}, y_{if}, \theta_{if}, exp_{if}, edu_{if}, c_{if}, a_{if})$.[14] Given the vector $S_{m,f}(i)$, the equilibrium of the game that characterizes the time allocations of the household is uniquely determined.[15] The log-wage equation for males and females comprising household i is specified:

$$\begin{aligned} \ln w_{im} &= \gamma_{0m} + \gamma_{1m}\theta_{im} + \gamma_{2m}edu_{im} + \gamma_{3m}c_{im} + \gamma_{4m}exp_{im} + \gamma_{5m}exp_{im}^2 + \epsilon_{im} \\ \ln w_{if} &= \gamma_{0f} + \gamma_{1f}\theta_{if} + \gamma_{2f}edu_{if} + \gamma_{3f}c_{if} + \gamma_{4f}exp_{if} + \gamma_{5f}exp_{if}^2 + \epsilon_{if} \end{aligned}$$

This specification is treated as a standard Mincer equation with additional personality trait θ_{if} and cognitive ability c_{if} components. The potential working experience term $\exp_{ij}$ is defined as $age - 6 - edu$. The disturbances $(\epsilon_{im}, \epsilon_{if})$ are assumed to follow a joint normal distribution:

$$\begin{bmatrix} \epsilon_{im} \\ \epsilon_{if} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\epsilon m}^2 & \rho\sigma_{\epsilon m}\sigma_{\epsilon f} \\ \rho\sigma_{\epsilon m}\sigma_{\epsilon f} & \sigma_{\epsilon f}^2 \end{bmatrix} \right)$$

where $\sigma_{\epsilon m}$ denotes the standard deviation of the male's wage, $\sigma_{\epsilon f}$ denotes the standard deviation of the wife's wage, and ρ denotes the correlation of the wage disturbances.

The model incorporates household heterogeneity in preferences and in the production technology by assuming the parameters, $(\lambda_m, \lambda_f, \delta_m, \delta_f)$ are drawn from a joint distribution $G_u(S_{m,f}^u)$, where the u subscript denotes the fact that these parameters are unobserved to the analyst, although they are assumed known by both spouses. The distribution G_u is para-

[14] $\{\lambda_{im,}\delta_{im}, \lambda_{if}, \delta_{if}\}$ are the unobserved preferences and production technology of household i drawn from distribution $G_u(S_{m,f}^u)$. $\{w_{im,}y_{im}, w_{if,}y_{if}\}$ are wages and other incomes in the household. Finally, $\{\theta_{im}, exp_{im}, edu_{im}, c_{im}, a_{im}, \theta_{if}, exp_{if}, edu_{if}, c_{if}, a_{if}\}$ are personality traits θ, potential working experience exp, education attainment edu, cognitive ability c and age a for both spouses m and f in household i.

[15] As noted above, in the noncooperative case, there is the possibility of two equilibria existing, one with the husband supplying time to the market and the wife not, and the other in which the wife works in the market and the husband does not. We use the convention that the one in which the male supplies time to the market is the one that is implemented.

metric, although it is "flexible" in the sense that it is characterized by a high-dimensional parameter vector. The distribution is created by mapping a four-dimensional normal distribution into the appropriate parameter space using known functions. Define the random vector $x_{4\times 1} \sim N(\mu_1, \Sigma_1)$, where μ_1 is 4×1 vector of means and $\sum_1$ is a 4×4 symmetric, positive-definite covariance matrix. The random variables $(\lambda_m, \lambda_f, \delta_m, \delta_f)$ are then defined using the link functions

$$\begin{aligned} \lambda_m &= \tfrac{\exp(x_1)}{1+\exp(x_1)} \\ \lambda_f &= \tfrac{\exp(x_2)}{1+\exp(x_2)} \\ \delta_m &= \tfrac{\exp(x_3)}{1+\exp(x_3)+\exp(x_4)} \\ \delta_f &= \tfrac{\exp(x_4)}{1+\exp(x_3)+\exp(x_4)} \end{aligned} \tag{3.1}$$

The joint distribution of preference and production technology parameters, $(\lambda_m, \lambda_f, \delta_m, \delta_f)$, is fully characterized by 14 parameters.

We assume that the household Pareto weights may depend on education, cognitive scores and personality traits as well as the ages of both spouses through the following parametric specification:

$$\alpha(i) = \frac{Q_m(i)}{Q_m(i) + Q_f(i)}, \tag{3.2}$$

where

$$Q_j(i) = \exp(\gamma_{6j} + \gamma_{7j}\theta_j(i) + \gamma_{8j} edu_j(i) + \gamma_{9j} c_j(i) + \gamma_{10j} a_j(i)),\ j = m, f.$$

The coefficients of γ_{7j},γ_{8j},γ_{9j},γ_{10j} capture the effects of personality traits, education, cognitive ability and age on the Pareto weight of the husband in household i. The Pareto weight of the wife is simply $1 - \alpha(i)$, the weights are both positive and normalized so as to sum to 1.

Dividing both the numerator and denominator of (3.2) by $Q_f(i)$, we have

$$\alpha(i) = \frac{\tilde{Q}(i)}{1 + \tilde{Q}(i)},$$

where

$$
\begin{aligned}
\tilde{Q}(i) &= Q_m(i)/Q_f(i) \\
&= \exp(\sum_{k=6}^{10}[\gamma_{km} z_{km}(i) - \gamma_{kf} z_{kf}(i)]),
\end{aligned}
$$

where the index k runs over all of the characteristics included in the $\alpha(i)$ function, and where $z_{6j}(i) = 1$, $z_{7j}(i) = \theta_j(i)$, $z_{8j}(i) = e_j(i)$, $z_{9j}(i) = c_j(i)$, and $z_{10j}(i) = a_j(i)$. We note that as long as the values of $z_{kj}(i)$ differ for husbands and wives in a sufficiently large number of households, the parameters γ_{km} and γ_{kf} are separately identified. With regard to the constant terms, only the difference $\gamma_{6m} - \gamma_{6f}$ is identified.

We compute the elasticity of husband's Pareto weight $\alpha(i)$ with respect to his personality traits $\theta_m(i)$ as

$$
\eta_m(i) = \frac{\partial \alpha(i)}{\partial \theta_m(i)} \frac{\theta_m(i)}{\alpha(i)}, \tag{3.3}
$$

and the elasticity of wife's Pareto weight $1 - \alpha(i)$ with respect to her personality traits $\theta_f(i)$ as

$$
\eta_f(i) = \frac{\partial (1 - \alpha(i))}{\partial \theta_f(i)} \frac{\theta_f(i)}{(1 - \alpha(i))}, \tag{3.4}
$$

for each household. In section 5 below, we will present the distribution of these elasticities for the five-dimensional personality traits included in the Pareto weight function.

3.4.1. Identification

The model described in section 2 is not nonparametrically identified, for reasons related to those given in Del Boca and Flinn (2012). It is useful to discuss identification in that model to see what the complicating factors are here. Del Boca and Flinn (2012) condition their analysis on both spouses being employed, which means that wages are observable. Del Boca and Flinn (2012) show that it is possible to nonparametrically identify the joint distribution of $(\lambda_m, \lambda_f, \delta_m, \delta_f, w_m, w_f, y_m + y_f)$ given $(h_m, \tau_m, h_f, \tau_f, w_m, w_f, y_m, y_f)$ under the assumption that all households behave noncooperatively (Nash equilibrium). They show that such a model is saturated, i.e., there are the same number of parameters to estimate as there are data points. A cooperative version of this model adds an additional parameter, either a scalar or a function, to select a point on the Pareto frontier that corresponds to the household's allocation. With this addition, the model is under-identified; to remedy this, the authors impose the assumption that $\alpha = 0.5$. If one is willing to assume either that all households use the value $\alpha = 0.5$ or that spouses reach a cooperative outcome where each obtains no less utility than they would in the Nash equilibrium, then cooperative models are also nonparametrically identified.[16]

The most important difference between the problem in Del Boca and Flinn (2012) and ours is the introduction of wage equations that are both independent objects of interest and that are required to correctly account for nonrandom selection into the labor force for husbands and wives.[17] These wage equations take the generic form

$$\begin{aligned} \ln w_{im} &= \gamma_{0m} + \gamma_{1m}\theta_{im} + \gamma_{2m} edu_{im} + \gamma_{3m} c_{im} + \gamma_{4m} exp_{im} + \gamma_{5m} exp_{im}^2 + \epsilon_{im} \\ \ln w_{if} &= \gamma_{0f} + \gamma_{1f}\theta_{if} + \gamma_{2f} edu_{if} + \gamma_{3f} c_{if} + \gamma_{4f} exp_{if} + \gamma_{5f} exp_{if}^2 + \epsilon_{if} \end{aligned}$$

We know that the sample selection problem leads to inconsistent OLS estimates of γ_{kj}, $k = 0, 1, 2, 3, 4, 5, j = f, m$ if only using observations with observed wages, because in general $E(\epsilon_{ij}|\theta_{ij}, edu_{ij}, c_{ij}, exp_{ij}) \neq 0$ for $j = m, f$. The selection mechanism that operates jointly on $(\epsilon_{im}, \epsilon_{if})$ is a rather complex one in our decision-making framework. Moreover, the parameters of the wage function that we attempt to estimate are just a subset of the parameters characterizing the household's decision problem and choices.

Under our model specification, the data generating process (DGP) depends on the observed

[16] In the latter case, individual utility is first computed assuming the weight $\alpha = 0.5$. At this outcome, if one spouse has a lower utility than in the Nash equilibrium, their weight is increased until that spouse is made indifferent between the cooperative and Nash equilibrium outcome. This means that in the population, a mass of households will have a weight of $\alpha = 0.5$, namely those households for which the participation constraint does not bind, with a distribution of ex post values of α not equal to 0.5 for households in which the participation constraint is binding.

[17] Allowing for spouses being out of the labor market in the Del Boca and Flinn (2012) model would have required the introduction of wage equations and would have also resulted in the model being no longer nonparametrically identified (even given an assumed value of the Pareto weight α).

characteristics of the husband and wife, and G_u, the distribution of preference and production technology parameters, and F_ϵ, the bivariate normal distribution of (ϵ_m, ϵ_f). Draws from G_u and F_ϵ along with the observed values of state variables $\{\theta_m, \theta_f, a_m, a_f, edu_m, edu_f, c_m, c_f, exp_m, exp_f\}$, determine the household's preference and production technology $\{\lambda_m, \lambda_f, \delta_m, \delta_f\}$, the wage offers $\{w_m, w_f\}$, and the household's value of α. Based on these state variable realizations, each of the household's four choices are determined (which are the labor market $\{h_m, h_f\}$ and housework time allocations $\{\tau_m, \tau_f\}$ of each spouse given the endogenous equilibrium choice). We denote all of the unknown parameters to be estimated by Ω. The model generates a joint conditional distribution over the endogenous variables $A_E \equiv (w_m, w_f, h_m, h_f, \tau_m, \tau_f)'$ given the vector of observed exogenous covariates A_C and the parameter vector Ω,

$$Q(A_E|A_C, \Omega).$$

This distribution cannot be expressed in a closed-form, but it is straightforward to simulate it by taking a large number of draws from G_u and F_ϵ. The conditional distribution $Q(A_E|A_C, \Omega)$ is the basis of the estimator described in more detail in the following subsection.

If the selection criteria sorting households into cooperation and spouses into labor market participation were less complex, it would be possible to employ a maximum likelihood estimator directly constructed from the conditional distribution $Q(A_E|A_C, \Omega)$ for each household. Parameter identification is relatively straightforward to analyze in such a case simply by determining whether the first order conditions are linearly independent. When using moment-based estimators, as we do, typically it is not possible to explicitly demonstrate the identification of all of the model parameters. The hope is that by including enough sample statistics, all of the model parameters will be identified and precisely estimated. Moments of the data are chosen, $m_h(A_E, A_C)$, $h = 1, ..., H$ where H is at least as large as the dimensionality of the parameter space, given by $\#\Omega$.[18] For example, one of the moments used in

[18]In our case, we use 85 moments to identify 54 parameters $\#\Omega = 54$. Appendix 3.8.3 provides a list of moments used in estimation.

forming the estimator is the proportion of wives in the sample with characteristics $a_C \in A_C$ who are in the labor market, in which case $m(A_E, A_C) = N^{-1}\sum_{i=1}^{N}\chi[h_f(i) > 0; a_C \in A_C, \Omega]$, where χ is the indicator function and N is the number of households in the sample. In general, there does not exist a unique deterministic solution $\bar{\Omega}$ such that $M = \tilde{M}(\bar{\Omega})$. Instead, we define a distance function, $D(\tilde{M}(\Omega), M)$, and the minimum distance estimator of Ω is given by

$$\hat{\Omega} = \arg\min_{\Omega} D(\tilde{M}(\Omega), M).$$

Whether or not the model is "well-identified" using a particular vector of sample moments is often determined after estimation has been attempted. Different sets of moments can yield different point estimates and associated standard errors in small samples, but it is seldom possible to determine an "optimal" vector of moments to use in a reasonably complex estimation problem. A specific parameter is said to be precisely estimated if the ratio of its point estimate to its estimated standard error is large in absolute value. In our case, it is almost never the case that this ratio of the parameter to its standard error is close to zero.

3.4.2. Model Estimation

We estimate the model using a relatively standard Method of Simulated Moments approach. Given a set of parameters, we repeatedly draw from the distributions of household preference parameters, production function parameters, and potential wage offers, $(\delta_m^r, \lambda_m^r, w_m^r, \delta_f^r, \lambda_f^r, w_f^r)$, R times for each household. Combined with other observed variables

$(y_m, \theta_m, c_m, a_m, edu_m, exp_m, y_f, \theta_f, c_f, a_f, edu_f, exp_f)$, we solve for the time allocation of the household $(h_f^r, \tau_f^r, h_f^r, \tau_f^r)$ within the selected equilibrium. Model parameters are estimated by choosing the parameters that minimize the quadratic distance function,

$$\hat{\Omega} = \arg\min_{\Omega}(\tilde{M}_{NR}(\Omega) - M_N)' W_N(\tilde{M}_{NR}(\Omega) - M_N),$$

where W_N is a positive definite weighting matrix. The NR subscript on $\tilde{M}$ signifies that these population analogs are computed from R simulations for each of the N households in the sample. Under standard conditions used to obtain consistency of GMM estimators, $plim_{N,R\to\infty}\hat{\Omega} = \Omega$ for any positive definite W. We compute the weighting matrix W_N following Del Boca et al. (2014) using a resampling method.[19] The weight matrix is the inverse of the diagonal of the bootstrapped covariance matrix of M_N:

$$W = P^{-1}\left(\sum_{p=1}^{P}(M_N^p - M_N)(M_N^p - M_N)\right)^{-1}.$$

Standard errors associated with the parameters $\hat{\Omega}$ are obtained using the standard asymptotic formula for generalized method of moments estimators.

3.4.3. Principal Component Analysis

Because many of the model parameters are associated with personality traits, the moments used in estimation need to capture the relationship between choices, outcomes and personality traits. There are five traits, each of which can take on values ranging from 1 to 7. To specify the moments in a parsimonious way, we first apply principal-components analysis (PCA) to the five personality trait variables to obtain linear combinations of traits that are used in estimation.[20]

We do the PCA separately for husbands and wives and, for each, retain the first two principal components, which have eigenvalues greater than 1. They are shown in Table 45. For the first component, the most crucial loadings are conscientiousness, agreeableness and emotional stability (.517, .543 and .493) in the male case. For women, all traits except openness

[19] We resample the original N observations a total of P times (where $P = 100$), and compute the vector of sample characteristics at each simulation s, which is given by M_N^p.

[20] Principal components is a statistical procedure that converts a set of possibly correlated variables into a set of linearly uncorrelated variables called principal components. The transformation is defined so that the first principal component has the largest possible variance (accounts for as much of the variability in the data as possible), and each succeeding component has the highest variance possible under the constraint that it is orthogonal to the preceding components.

Table 45: Principal-components analysis for five dimensional personality traits

	Male		Female	
	1	2	1	2
Eigenvalues	1.47	1.22	1.59	1.18
Variance	29.4%	24.5%	31.8%	23.5%
Openness to experience	0.144	0.771	0.087	0.820
Conscientiousness	0.556	-0.086	0.511	-0.066
Extraversion	0.363	-0.168	0.466	-0.009
Agreeableness	0.543	0.407	0.456	0.408
Emotional stability	0.493	-0.452	0.553	-0.397

to experience contribute almost equally to the first component. For the second component, loadings are concentrated on openness to experience for both males and females (.788 and .789). We then discretize the first two principal components into three levels (low, middle and high) and construct moments conditioning on these components and categories.[21]

3.4.4. Selection of Sample Characteristics

We estimated the above parameters by matching the following six groups of moments: (1) proportion employed; (2) average working hours ; (3) average housework hours; (4) average wage for the employed workers; (5) standard error of male's log wage and female's log wage; (6) fraction of working hours in certain intervals; (7) fraction of housework hours in certain intervals; (8) Covariance between men and women's time allocations; (9) Correlation between men and women's accepted wages. We calculated moments 1-7 for husbands and wives separately. For moments 1-4 we use marginal moments conditional on education level (college, no college), principal component 1 range (low, middle and high) and principal component 2 range (low, middle and high). The remaining moments are unconditional. In total, there are 85 moments. A detailed description can be found in Appendix 3.8.3.

[21]The cut-offs to assign observations to the low, middle and high categories correspond to the 33rd and 66th percentiles.

3.5. Estimation Results

3.5.1. Model Estimates

Table 46 reports the estimated model parameters. Part 1 (the upper panel) displays the coefficient estimates associated with personality traits in the wage offer function and their impact on the Pareto weight for husbands and wives. Given our specification, we obtain a *certeris paribus* effect of personality traits on log wages, conditional on education and cognitive ability. Personality traits are also likely to be important at younger ages in shaping education choices and cognitive skills. There is evidence that attending college also influences the evolution of personality traits. (e.g.Todd and Zhang (2017)) We cannot use our static model to fully examine the influence of personality traits over an individual's lifetime as they operate through various channels at different ages. Rather, we use the model to examine the role played by personality traits in determining household interaction, wages and labor supply, above and beyond that of other characteristics, such as education and cognitive ability.

The parameter estimates show that the education coefficient (conditional on the other included variables) is somewhat larger for women (0.0673) than for men (0.0538). Personality traits are important determinants of wage offers for men but individually are not statistically significant determinants for women. Of the personality traits, conscientiousness significantly increases male wage offers and agreeableness significantly lowers wage offers. Cognitive ability increases wage offers for both men and women. The wage return for cognitive ability is twice as high for men (0.12) as for women (0.06).

Table 47 reports F-statistics and associated p-values from Wald tests for the joint statistical significance of the five personality traits in the wage equations and in the pareto weights for both men and women. We reject the null that personality traits are not significant in all cases.

The estimated parameters that determine the Pareto weights are shown in Table 46. The

Table 46: Parameter estimates

Part 1	Male			Female	
	Estimates	S.E.		Estimates	S.E.
		Log wage equation			
γ_{0m}	2.2467	(0.1448)	γ_{0f}	2.1326	(0.2008)
$\gamma_{1m}(Opn)$	-0.0256	(0.0217)	$\gamma_{1f}(Opn)$	-0.0141	(0.0152)
$\gamma_{1m}(Cos)$	0.0504	(0.0132)	$\gamma_{1f}(Cos)$	0.0259	(0.0336)
$\gamma_{1m}(Ext)$	0.0087	(0.0089)	$\gamma_{1f}(Ext)$	0.0146	(0.0199)
$\gamma_{1m}(Agr)$	-0.0446	(0.0215)	$\gamma_{1f}(Agr)$	-0.0171	(0.0212)
$\gamma_{1m}(Stb)$	0.0091	(0.0133)	$\gamma_{1f}(Stb)$	-0.0392	(0.0243)
$\gamma_{2m}(Edu)$	0.0538	(0.0108)	$\gamma_{2f}(Edu)$	0.0673	(0.0143)
$\gamma_{3m}(Cog)$	0.1230	(0.1073)	$\gamma_{3f}(Cog)$	0.0610	(0.0620)
$\gamma_{4m}(Exp)$	0.0294	(0.0085)	$\gamma_{4f}(Exp)$	0.0163	(0.0066)
$\gamma_{5m}(\frac{Exp^2}{100})$	-0.0786	(0.0450)	$\gamma_{5f}(\frac{Exp^2}{100})$	-0.0447	(0.0290)
$\sigma(\epsilon_m)$	0.6366	(0.0713)	$\sigma(\epsilon_f)$	0.8571	(0.1000)
ρ	0.7548	(0.1542)			
		Pareto Weight			
$\gamma_{6m} - \gamma_{6f}$	-0.2616	(0.0685)			
$\gamma_{7m}(Opn)$	-0.1406	(0.0396)	$\gamma_{7f}(Opn)$	-0.2390	(0.0429)
$\gamma_{7m}(Cos)$	-0.0776	(0.0205)	$\gamma_{7f}(Cos)$	0.1393	(0.0275)
$\gamma_{7m}(Ext)$	0.0651	(0.0407)	$\gamma_{7f}(Ext)$	0.0412	(0.0394)
$\gamma_{7m}(Agr)$	-0.3630	(0.1099)	$\gamma_{7f}(Agr)$	-0.2739	(0.0755)
$\gamma_{7m}(Stb)$	0.1322	(0.0480)	$\gamma_{7f}(Stb)$	0.0052	(0.0060)
$\gamma_{8m}(Edu)$	0.1391	(0.0288)	$\gamma_{8f}(Edu)$	0.0631	(0.0215)
$\gamma_{9m}(Cog)$	-0.6623	(0.1064)	$\gamma_{9f}(Cog)$	-0.2805	(0.2935)
$\gamma_{10m}(Age)$	0.0401	(0.0187)	$\gamma_{10f}(Age)$	0.0945	(0.0317)

Part 2	*Preference and Production Parameters*				
	Mean (μ)	Co-variance (σ^2)			
λ_m	0.3621	0.0123	0.0092	-0.0008	0.0016
S.E.	(0.0145)	(0.0024)	(0.0028)	(0.0007)	(0.0006)
λ_f	0.3372	0.0092	0.0099	0.0019	-0.0009
S.E.	(0.0132)	(0.0038)	(0.0051)	(0.0012)	(0.0006)
δ_m	0.1395	-0.0008	0.0019	0.0072	-0.0002
S.E.	(0.0221)	(0.0007)	(0.0012)	(0.0058)	(0.0016)
δ_f	0.1953	0.0016	-0.0009	-0.0002	0.0170
S.E.	(0.0190)	(0.0006)	(0.0006)	(0.0016)	(0.0100)

Table 47: Joint Wald test for the explanatory power of "Big-five" personality traits in the wage equation and in the bargaining equation

	Null Hypothesis	$F-value$	$p-value$
Wage (Male)	$\gamma_{1m}=0$	137.8	0.000
Wage (Female)	$\gamma_{1f}=0$	35.3	0.000
Bargaining (Male)	$\gamma_{5m}=0$	169.6	0.000
Bargaining (Female)	$\gamma_{5f}=0$	872.1	0.000

Pareto weights for both men and women are significantly influenced by personality traits, education and cognitive ability. Extraversion and emotional stability have a ceteris paribus positive effect on the Pareto weight, whereas openness to experience and agreeableness have a negative effect. Conscientiousness increases the Pareto weight for women but decreases it for men. Age increases the Pareto weight for both men and women and cognitive ability decreases it.

When particular personality traits have negative ceteris paribus effects on the Pareto weight, it does not mean that individuals with higher values of these traits necessarily have a lower household Pareto weight. This is because it is the the relative difference between spouses rather than the absolute value that determines the overall Pareto weight. With assortative matching on traits, individuals with high scores on certain traits are likely to have spouses with high scores on the same traits. We will further explore the importance of this assortative matching in affecting the degree of household cooperation below.

To better understand how personality traits of both spouses affect the Pareto weight, we calculate $\{\eta_m(i), \eta_f(i)\}$, the elasticity of Pareto weights with respect to their personality traits separately, following equation 3.3 and 3.4 described in the last section. Figure 22 displays the distribution of $\eta_m(i)$ and $\eta_f(i)$. In general, personality traits demonstrate significant asymmetric effects on Pareto weights. Among all personality traits, agreeableness is the most important trait in determining the Pareto weights. The average elasticities of agreeableness are -0.861 for men and -0.880 for women. That is, a percent increase of the husband's agreeableness decreases his Pareto weight (α) by 0.861%. A percent increase in wives' agreeableness decreases her Pareto weight ($1-\alpha$) by 0.880%.

Figure 22: The distributions of both $\eta_m(i)$ and $\eta_f(i)$

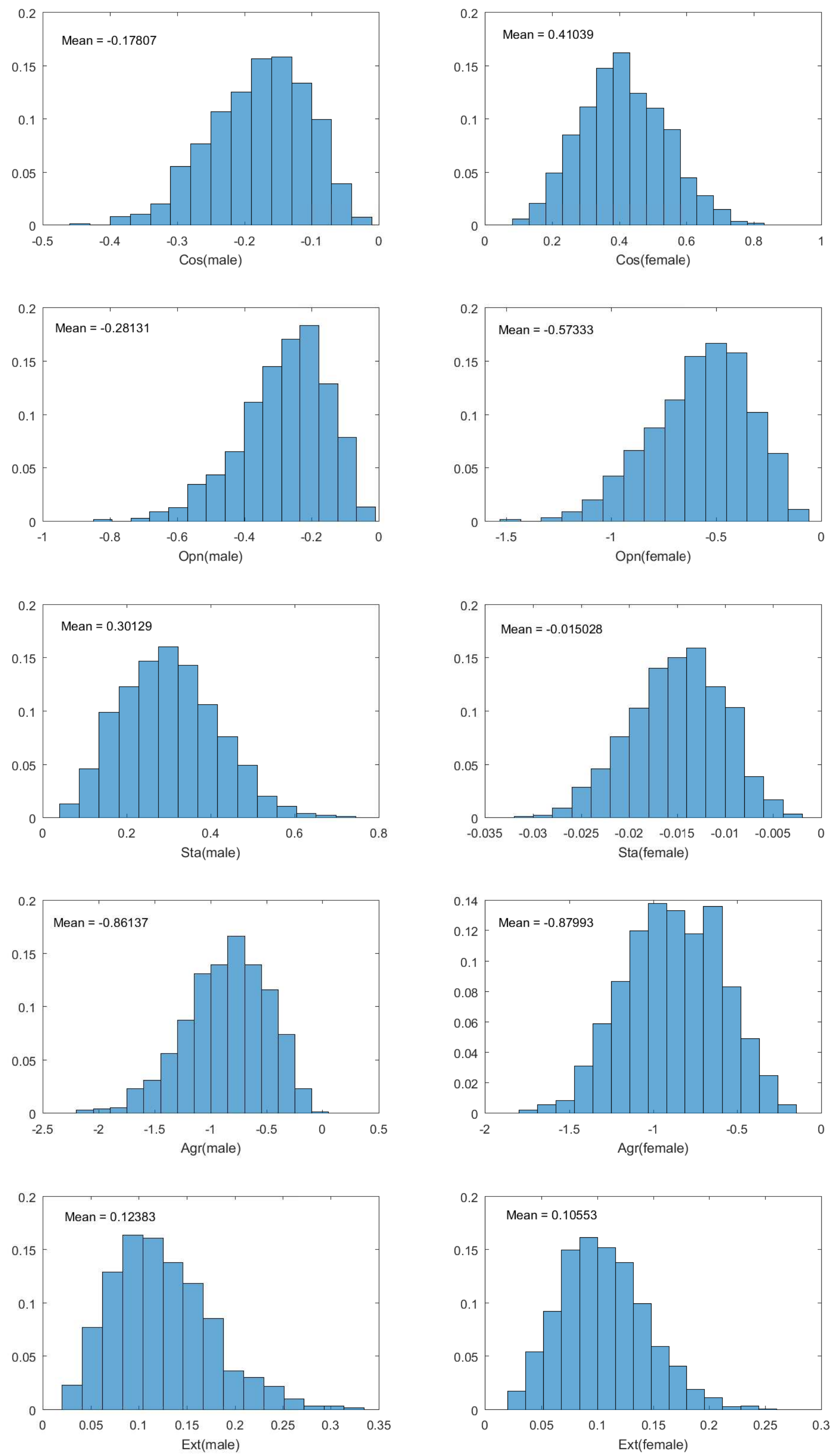


Table 48: The fraction of households playing cooperatively along with Pareto weight α

α	Fraction	Obs.
[0.10,0.20)	0	11
[0.20,0.30)	0	59
[0.30,0.40)	0.243	169
[0.40,0.50)	0.618	280
[0.50,0.60)	0.694	359
[0.60,0.70)	0.266	293
[0.70,0.80)	0.068	205
[0.80,0.90)	0	65
[0.90,1.00]	0	2
Total	0.387	1,443

Note: all Pareto wights α are within range [0.1,1]. The mean and S.D. of α are 0.555 and 0.151.

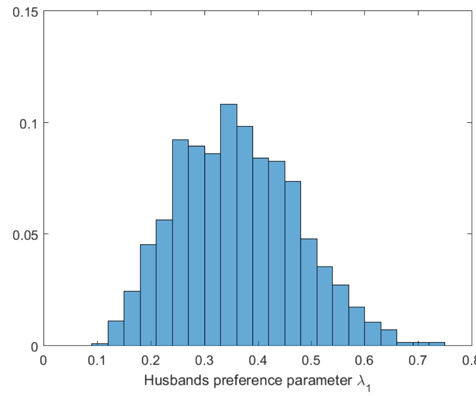
Next, we explore how the Pareto weight affects the possibility of a household adopting a cooperative allocation. Table 48 displays the fraction of cooperative households for different values of α. Although α assumes values from 0 to 1, the Pareto weights of most households lie in the range [0.20,0.80], indicating that spouses in most households share fairly equal weights. Also, we observe that households are more likely to adopt a cooperative interaction mode when the Pareto weight is close to 0.5. Around 61.8% and 69.4% of households choose to play cooperatively when $\alpha \in [0.40, 0.50)$ and $\alpha \in [0.50, 0.60)$, whereas none of households play cooperatively when the value of α is extreme ($\alpha > 0.80$ or $\alpha < 0.20$).

We next describe the estimated distribution of the spousal preference and production parameters $\{\delta_m, \delta_f, \lambda_m, \lambda_f\}$. Figure 24 (a) shows the marginal distributions of the preference and production parameters. The distribution of husbands' leisure preferences and the distribution of wives' leisure preference are similar. The husbands' production parameter distribution is more left-skewed than wives'; males are estimated to be less efficient in producing public goods than females.

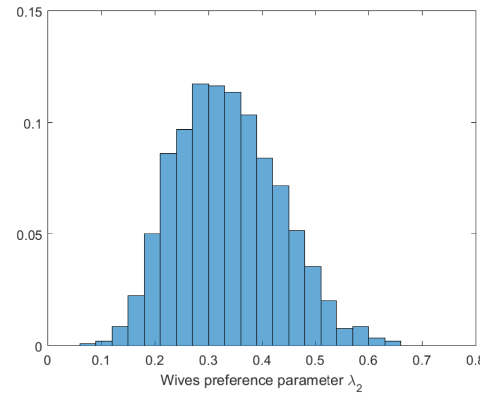
Figure 24 (b) plots the bivariate distribution of spousal preference and production parameters. There is a strong relationship between both the preference (λ_1 and λ_2) and the production parameters (δ_1 and δ_2) over most of the support of the distribution. That is, husbands and wives exhibit a substantial degree of positive assortative matching with respect

Figure 23: Histograms of spousal production and preference parameters

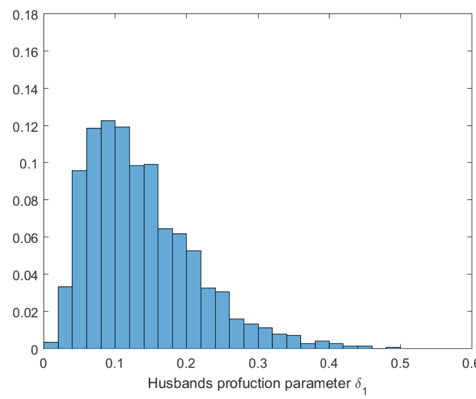
(a) Husbands' preference parameters λ_m



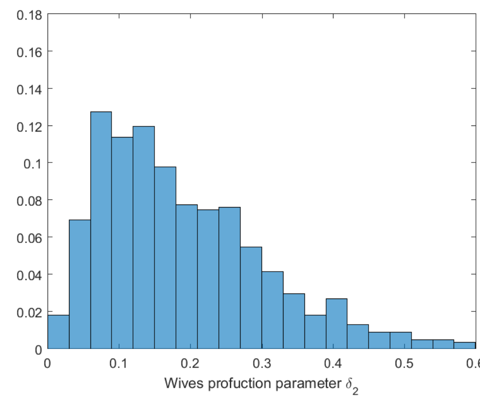
(b) Wives' preference parameters λ_f



(c) Husbands' production parameters δ_m



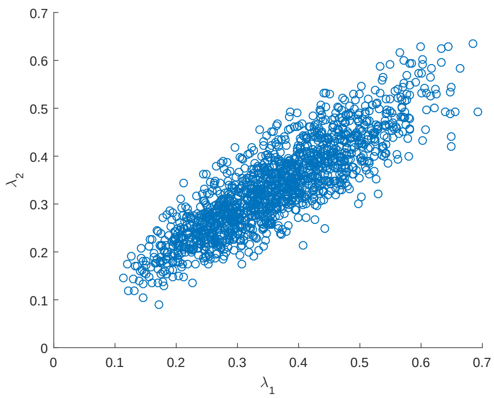
(d) Wives' production parameters δ_f



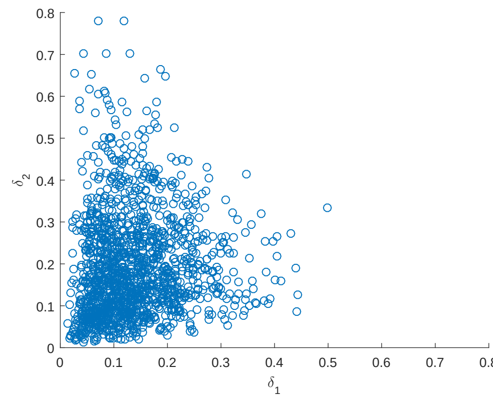
to both preference and production characteristics, although the sorting on the production parameter is less pronounced. For husbands, there is a weak positive correlation between the preference and production parameters as seen in figure 24(b) - subfigure (c). For wives (figure 24(b) - subfigure (d)), there is little evidence of a systematic relationship between productivity and preference parameters. Although the sample used for our estimation differs from that used in Del Boca and Flinn (2012), the estimated unobserved preference and production distributions are similar.

Figure 24: Bivariate relationships between production and preference parameters

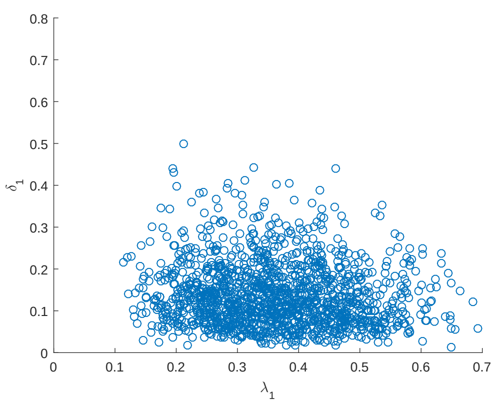
(a) Spousal preference parameters



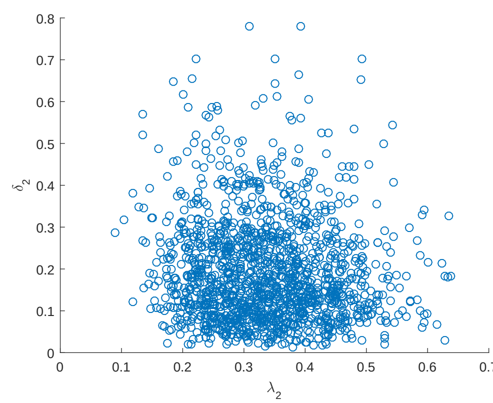
(b) Spousal productivity parameters



(c) Husbands' primitive parameters



(d) Wives' primitive parameters



3.5.2. Goodness of Model Fit

The goodness of model fit is shown in tables 49, 50 and 51. Table 49 shows the mean labor participation rate, accepted wage, working hours and housework hours. All moments are conditional on education levels and on ranges of values of the first and second principal components. The model captures well the proportion working, the average wages for workers, the average hours for workers and the hours of housework by education level (low or high) for the different categories.

Table 50 shows the fit of the time allocation distribution, where time is divided into three intervals, chosen so that the number of observations in each interval is roughly equal. Although the model reproduces the distribution of household hours fairly well (both for men and women), it under-predicts the number of individuals with working hours in the middle interval. This underestimation is caused by non-smoothness of working hour distribution, as many people report working hours equal to 40. The fraction of men's working hour in (40,46) is only 0.157, which is much close to our simulation.

3.5.3. External Validation

As previously noted, we use the "fair share" question as a way of examining the validity of the model's implications. To determine a fair share reference point, we use the estimated model to simulate housework time allocations under the case where husbands and wives have equal pareto weights. We compare the housework hours implied by the model with $\alpha = 0.5$ to that reported by the household. If the "fair share" question is informative, then individuals who report "I do *much more* than my fair share" should be observed to do more housework than their "fair share " and vice versa.[22]

Table 52 reports the housework hours and Pareto weight for both spouses categorized by

[22] A household's optimal time allocation may change over time with changes in job opportunities as well as the number of children, and respondents' interpretation of fair share might be one that views allocation over a span of time. Given that our model is static, though, we interpreted the response to the "fair share" as relating to the present time.

Table 49: Sample fit of husbands' and wives' wages and time allocations (mean level)

	Probability work > 0 hours				*Wages if work (avg.)*			
	Male		Female		Male		Female	
	Simulated	Data	Simulated	Data	Simulated	Data	Simulated	Data
Education								
Low	0.937	0.917	0.821	0.790	24.003	21.688	15.505	14.709
High	0.969	0.966	0.887	0.909	30.869	31.058	27.414	27.912
First Principal Component								
Low	0.927	0.925	0.859	0.790	26.473	26.304	20.667	19.587
Middle	0.961	0.943	0.859	0.882	26.174	25.064	20.638	21.634
High	0.966	0.949	0.844	0.878	28.832	26.56	21.524	20.936
Second Principal Component								
Low	0.950	0.950	0.828	0.858	27.321	25.675	19.615	20.817
Middle	0.950	0.933	0.857	0.868	27.617	26.331	20.945	20.645
High	0.951	0.934	0.877	0.825	26.516	25.887	22.306	20.752
	Hours worked if work (avg.)				*Hours of housework (avg.)*			
	Male		Female		Male		Female	
	Simulated	Data	Simulated	Data	Simulated	Data	Simulated	Data
Education								
Low	40.511	42.765	25.416	23.379	18.545	18.651	25.479	26.747
High	43.197	44.58	38.610	37.847	17.713	17.637	30.096	29.266
First Principal Component								
Low	40.163	43.82	31.322	27.811	18.151	18.052	28.207	27.785
Middle	40.814	43.829	31.959	30.389	18.639	18.180	26.792	27.342
High	44.207	43.135	31.028	31.675	17.713	18.358	27.662	28.518
Second Principal Component								
Low	42.464	44.421	29.291	29.243	17.511	17.526	28.544	27.650
Middle	42.289	43.2	32.337	30.867	18.654	19.102	26.184	27.871
High	40.415	43.148	32.714	29.831	18.349	17.940	27.950	28.115

Table 50: Sample fit of husbands' and wives' wages and time allocations (distribution)

	Male			Female	
	Simulated	Data		Simulated	Data
Working Hours					
[0,40]	0.420	0.501	[0,24]	0.400	0.356
()40,46)	0.161	0.157	(24,38)	0.219	0.238
[46,60]	0.420	0.342	[38,60]	0.381	0.406
Housework Hours					
[0,11)	0.301	0.322	[0,16)	0.295	0.311
[11,20)	0.337	0.321	[16,34)	0.382	0.349
[20,56]	0.360	0.357	[34,56]	0.322	0.340
S.D. of log wages					
S.D.	0.334	0.364	S.D.	0.360	0.351
Corr	0.686	0.410		0.686	0.410

Table 51: Sample fit of covariance matrix of time allocation

Correlation of Hours	Working (M) Sim	Working (M) Data	Housework (M) Sim	Housework (M) Data	Working (F) Sim	Working (F) Data	Housework (F) Sim	Housework (F) Data
Working (M)	1.000	1.000	-0.092	-0.123	-0.092	-0.123	0.077	-0.029
Housework (M)	-0.092	-0.123	1.000	1.000	0.093	0.105	0.127	0.106
Working (F)	0.204	0.243	0.093	0.105	1.000	1.000	-0.421	-0.374
Housework (F)	0.077	-0.029	0.127	0.106	-0.421	-0.374	1.000	1.000

Table 52: The average housework hours categorized by "fair share" question

	Male			Female		
Fair share	Housework hours	Relative difference	Number of obs.	Housework hours	Relative difference	Number of obs.
Much More(1)	31.38	10.86	26	35.00	3.72	287
A bit More(2)	21.34	0.97	181	28.30	-1.15	548
Fair Share(3)	18.64	-2.21	840	25.15	-4.75	532
A bit Less(4)	15.55	-5.55	362	17.78	-14.01	69
Much Less(5)	8.68	-13.80	34	9.29	-15.36	7

Note: "Relative difference" column displays the relative difference between the actual household hours and the simulated housework hours when setting Pareto weight $\alpha = 0.500$.

"fair share" question. The column "Relative Difference" displays the difference between the actual housework hours and the simulated housework hours under the Pareto weight set equal to 0.5. When individuals report doing more than their "fair share", the actual housework hours are larger than the simulated hours, and when the opposite is true the individual tends to work less time in the household than the simulated hours.

3.5.4. Wage Decomposition

We next do a series of wage decompositions to understand the importance of education and personality traits in explaining gender gaps in accepted and offered wages. In our first decomposition, we decompose the mean log wage gap into five sources:

$$\underbrace{log\,\bar{w}_m - log\,\bar{w}_f}_{\text{mean wage gap}} = \underbrace{\gamma_{0m} - \gamma_{0f}}_{\text{unexplained part}} + \underbrace{(\gamma_{1m}\bar{\theta}_m - \gamma_{1f}\bar{\theta}_f)}_{\text{explained by traits}} + \underbrace{(\gamma_{2m}\bar{e}_m - \gamma_{2f}\bar{e}_f)}_{\text{explained by education}} + \underbrace{(\gamma_{3m}\bar{c}_m - \gamma_{3f}\bar{c}_f)}_{\text{explained by cognition}}$$
$$+ \underbrace{(\gamma_{4m}\bar{exp}_m - \gamma_{4f}\bar{exp}_f)}_{\text{explained by experience}} + \underbrace{(\gamma_{5m}\bar{exp}^2_m - \gamma_{5f}\bar{exp}^2_f)}_{\text{explained by experience}^2} + \underbrace{(\bar{\epsilon}_m - \bar{\epsilon}_f)}_{\text{selection bias}}$$

Table 53: The decomposition of gender wage gap

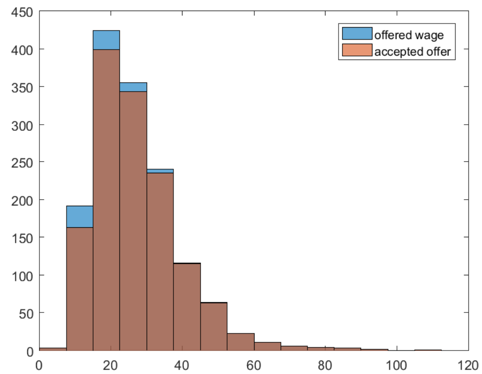
	Offered Wage	Accepted Wage
$\Delta\gamma_0$	0.1142	0.1142
Education	-0.1819	-0.1869
Conscientiousness	0.1195	0.1200
Openness	-0.0529	-0.0528
Stability	0.2502	0.2495
Agreeableness	-0.1337	-0.1338
Extraversion	-0.0306	-0.0308
Cognitive	0.0006	0.0012
Experience	0.3220	0.3271
Experience^2	-0.2109	-0.2165
$\Delta\epsilon$	0	-0.0303
Total	0.1931	0.1610

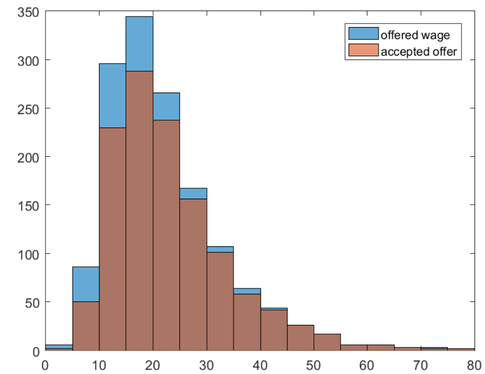
Table 53 shows the gender wage gap attributable to these different sources. The gap in accepted wages is 16.10%, while the gap in offered wages is 19.31%. As was seen in Table 46, females receive a slightly higher return for their educational attainment than males. Education narrows the offered-wage gap by 18.19 percentage points and the accepted-wage gap by 18.69 percentage points. However, the female advantage in the return to education is largely offset by a relative disadvantage in the return to potential work experience. Work experience increases the offered-wage gap by 11.11 percentage points and the accepted-wage gap by 11.06 percentage points. Among the "Big Five" personality traits, conscientiousness and emotional stability are the most important two traits contributing to a widening of the wage gap (11.95 percentage points and 25.02 percentage points for the offered-wage gap). Agreeableness, on the other hand, narrows the gender offered wage gap by 13.37 percentage points. The impact of openness to experience and extraversion in explaining the log wage gap is not significant. In total, personality traits explain 15.25 percentage points of the offered log wage gap. Their combined contribution to explaining the gender gap is the same magnitude as the contribution of education and working experience. Cognitive ability explains only a small fraction of the wage gap.

Figure 25 plots the distributions of both offered wages and accepted wages. Female workers are on average more selective than male workers; that is, a lower fraction of females (85.2

Figure 25: Distributions of accepted wages and offered wages

(a) Male's offered wage and accepted wage (b) Female's offered wage and accepted wage





percent) accepts the offered wage and works in the labor market. Male workers' accepted wages are on average 1.63 percent higher than offered wages, whereas female workers accepted wages are on average 4.45 percent higher than offered wages. For this reason, the gender gap in accepted wages is smaller than the gap in offered wages.

Table 53 shows that personality traits and education levels are both important to explaining gender wage gaps. Wage gaps can arise either because women have on average different traits and/or because women receive different payoffs in the labor market for their traits (as was evident in Table 46). We next explore whether and to what extent the gender gap is explained by differences in observed traits or differences in the market valuation of those traits. Following Oaxaca (1973) and Blinder (1973), we perform the following decomposition:

$$\gamma_{1m}\bar{\theta}_m - \gamma_{1f}\bar{\theta}_f = \underbrace{\gamma_{1m}(\bar{\theta}_m - \bar{\theta}_f)}_{\text{personality difference}} + \underbrace{(\gamma_{1m} - \gamma_{1f})\bar{\theta}_f}_{\text{coefficient difference}}$$

The first term is interpreted as the part of the log wage differential due to differences in traits, and the second term is the difference arising from gender differences in the estimated coefficients associated with those traits. The decomposition results are reported in table

Table 54: The Oaxaca-Blinder decomposition for personality traits, education and working experience

	Total difference	Due to characteristic	Due to coefficients
Constant	0.1142	0.1142	0.0000
Education	-0.1819	-0.0032	-0.1787
Conscientiousness	0.1195	-0.0117	0.1313
Openness	-0.0529	-0.0040	-0.0489
Stability	0.2502	-0.0003	0.2506
Agreeableness	-0.1337	0.0236	-0.1573
Extraversion	-0.0306	-0.0029	-0.0277
Cognitive	0.0006	-0.0138	0.0144
Experience	0.3220	0.0596	0.2624
Experience^2	-0.2109	-0.0516	-0.1593
Total	0.1965	0.1098	0.0867

54.

In general, gender wage gaps are largely explained by gender differences in labor market evaluations of characteristics (education, personality traits and potential working experience). For example, the differences in personality traits mean values explain 0.44 percentage points of the offered wage gap, but the differences in trait premia/penalties explain 8.67 percentage points. The gender difference in the valuation of emotional stability widens the offered wage difference by 25.06 percentage points and is the most important single factor to explain the gender wage gap. Another important factor is the male-female difference in the premium for conscientiousness, which widens the offered wage gap by 13.13 percentage points. In contrast, the gender difference in the valuation of agreeableness shrinks the gender wage gap by 15.73 percentage points. The contributions of other two traits - openness to experience and extraversion - in explaining gender differences in wage offers are minor.

3.6. Counterfactual Experiments

3.6.1. Comparing Different Modes of Interaction between Spouses

We next examine how household behaviors differ in the cooperative and noncooperative regimes by using the estimated model to simulate behaviors that would result if all house-

Table 55: The new allocations under different forms of interactions

	Gender	Participation rate	Housework hours	Working hours	Accepted wages	Wage Gap	Utility
Baseline	Males	95.2%	18.2	41.7	27.2	15.2%	5.45
	Females	85.0%	27.5	35.0	23.3		5.50
Cooperative	Males	95.9%	22.0	46.7	27.1	21.5%	5.48
	Females	92.7%	32.5	43.5	22.8		5.49
Non-cooperative	Males	93.6%	16.0	37.5	27.2	13.6%	5.42
	Females	80.2%	24.6	31.9	23.7		5.47

Note: The average working hours To be consistent with the previous wage decomposition, the wage gap here is defined as $\log W_m - \log W_f$

holds interacted in a cooperative or noncooperative manner. We compare the time allocations and outcomes to our baseline model, where we found that 38.4 percent of sample households choose to cooperate. As seen in Table 55, under the cooperative regime, both men and women supply more hours to market work and to household work than in the baseline case. The working hours for men and women increase on average by 5.0 and 8.5 hours, respectively, while housework hours increase by 3.8 hours and 5.0 hours. The gap in accepted wages increases from 15.2 percent in the baseline model to 21.5 percent in the cooperative regime. In the noncooperative regime, both men and women supply fewer hours to the labor market and to household work and devote more hours to leisure. The accepted wage gap is largest under the cooperative regimes, and the average utility levels for men under this regime is also the highest. The accepted wage gap is lowest under the noncooperative regime and the average utility values are also lowest. This indicates that reducing the observed gender wage gap is not necessarily welfare improving.

The explanation for the different time allocations under cooperative and noncooperative regimes is intuitive. For any set of state variables characterizing the household, when both the public good K and the private good l are valued by husbands and wives, the household will produce more of the public good in the cooperative equilibrium. The Nash equilibrium is inefficient in that neither spouse takes account of the fact that by spending more time in the market and housework they will increase the welfare of their spouse, and so leisure is over-consumed relative to its efficient level. In any efficient equilibrium (i.e., whatever the value of α), more K will be produced than in the Nash equilibrium. Because both labor supply, that generates income, and housework time are inputs in the production of K, both will generally increase. The reservation wage of either spouse will be lower for each spouse in the cooperative equilibrium. There is an increase gender wage gap in part because women are now willing to work at lower wages. Although the cooperative allocation is always an efficient equilibrium, it does not guarantee that both husbands and wives are able to attain a higher welfare level simultaneously compared with their baseline levels. In our case, the cooperative allocation improves men's utility but hurts women's utility on average. This also explains why the cooperative equilibrium is not chosen by all households.

3.6.2. The Effect of Positive Assortative Matching

According to Figure 21 and Figure 24(b), married couples display positive assortative matching on both unobserved preference and production parameters as well as on observed personality traits and cognitive abilities. We perform two counterfactual experiments to measure the effect of positive assortative matching on the gender wage gap. In the first, we randomly assign males and females from the original sample to form new households, and we refer to this environment as one of "pure random matching." In the second experiment, we preserve the education and the age of the matched spouses but reshuffle the personality traits of households. The comparison of these two experiments to the baseline scenario allows examination of the effects of positive sorting on cooperation and on household time allocation.

Table 56 suggests that eliminating positive assortative matching between spouses generates significant effects on labor force participation rates and wage gaps. In the pure random matching experiment, the labor participation rate of males and females decreases by 10.5 and 10.1 percentage points, and the accepted wage increases by \$1.0 for men and by \$1.3 for women. As a result, the gender gap in accepted wages shrinks from 15.2 percent in the baseline case to 13.4 percent in the pure random matching case. Meanwhile, the "limited" random matching experiment displays a very similar effect. The labor participation rate

Table 56: The counterfactual experiment of random match

	Gender	Participation rate	Housework hours	Working hours	Accepted wages	Wage Gap	Cooperative Fraction
Baseline	Males	95.2%	18.2	41.7	27.2	15.2%	38.7%
	Females	85.0%	27.5	35.0	23.3		
Pure random matching	Males	84.7%	18.5	42.5	28.2	13.4%	32.6%
	Females	74.9%	26.9	38.7	24.6		
Limited random matching	Males	85.2%	18.6	42.3	28.2	13.1%	33.1%
	Females	75.2%	26.9	38.7	24.7		

of males and females decreases by 10.0 and 9.8 percentage points, and the gender wage gap decreases to 13.1%. The driving force behind these results in both experiments is the decrease numbers of cooperative households. The reason is that extreme Pareto weights are easier to generate when matching is random, resulting in a lower fraction of cooperative households. When households adopt noncooperative behaviors, they obtain less utility from the public good. Consequently, they choose to work fewer hours and gender a lower level of public goods. Thus, labor force participation rates also fall (especially for females). By comparing the two counterfactual experiments (random match and limited random match), we can ascertain that the effect of age and education sorting on cooperation is relatively small.

3.6.3. Equalizing pay opportunities

Lastly, we use our model to simulate household time allocations and wage outcomes that would result if women were paid according to the male wage offer equation. That is, women may still receive different wage offers from men because their personal attributes differ, but their education, personality traits, and cognitive skill are valued in the same way in the labor market as they are for men.

As seen in Table 57, when men and women have the same wage offer equation, then women have better opportunities in the labor market on average. Women choose to spend more hours and men fewer hours doing market work. In terms of housework, women decrease their housework hours relative to the baseline and men increase their housework hours. In-

Table 57: The new allocations under equal pay experiment

	Gender	Participation rate	Housework hours	Working hours	Accepted wages	Wage Gap	Cooperative Fraction
Baseline	Males	95.2%	18.2	41.7	27.2	15.2%	38.7%
	Females	85.0%	27.5	35.0	23.3		
Equal pay	Males	90.9%	19.3	38.3	27.4	-2.45%	38.8%
	Females	92.0%	25.9	39.2	28.1		

terestingly, when women and men have the same wage offer functions, the accepted wages for women are on average similar to the accepted wages for men. The gender wage gap in accepted wages changes from 15.2 percentage points in the baseline model to -2.45 percentage points in this "equal pay opportunities" simulation.

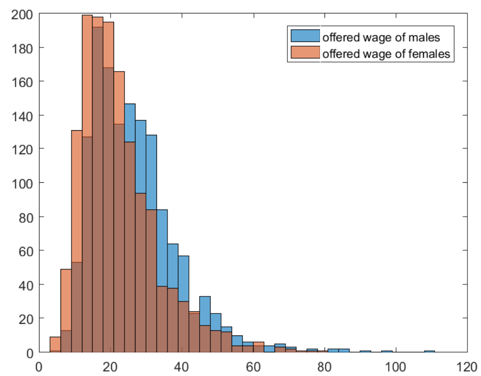
Figure 26 displays the distributions of offered wages and accepted wages in both the baseline and the counterfactual models. In the baseline model, the female's wage distribution is more left-skewed than male's wage distribution, indicating that the offered wages and accepted wages are lower for women than for men. However, this gap is totally eliminated and under the equal pay opportunities simulation.

3.7. Conclusions

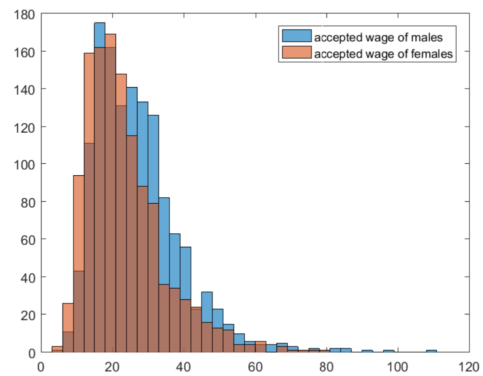
In this paper, we study the role of personality traits in household decision-making, specifically with regard to decisions about time allocation to housework and market work and the implications for gender wage disparities. First, we find that personality traits are significant determinants of household Pareto weights for both men and women. Second, we find that personality traits are also statistically significant determinants of offered wages for both men and women. Males receive a positive return for being conscientious and a negative return for being agreeable. For women, the individual personality traits are not statistically significant but they are jointly significant. Overall, the effect of personality traits on the wage equation is comparable to the effect of education and potential work experience. Third, an Oaxaca-Blinder type decomposition analysis shows that the gender wage gap largely is attributable to gender differences in market valuations of traits rather

Figure 26: Distributions of accepted wages and offered wages

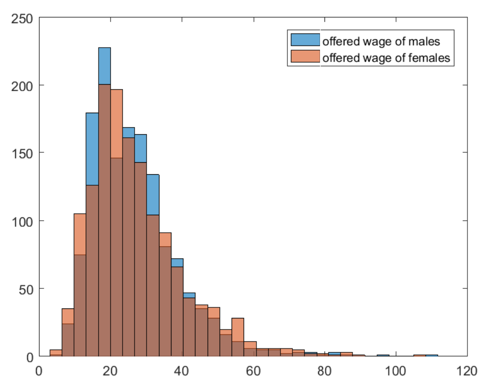
(a) Offered wages in baseline model



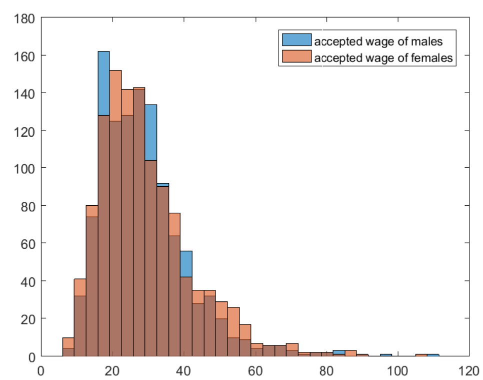
(b) Accepted wages in baseline model



(c) Offered wages in equal pay experiment



(d) Accepted wages in equal pay experiment



than to differences in the levels of those traits. Male-female differences in the return to conscientiousness and emotional stability emerge as the most important factors contributing to a widening of the wage gap. Differences in the return to agreeableness and to education contribute to a narrowing of the wage gap.

The model we estimated allows households to choose to behave cooperatively or noncooperatively, with personality attributes potentially affecting household Pareto weights. We find that 38.7 percent of households behave cooperatively. Cooperation leads a household to assign a higher value to public goods that require both monetary and time investments to produce. This leads both men and women to supply a greater number of hours to the labor market and to housework than they would under a noncooperative regime. Observed wage gaps are higher under a cooperative regime but utility is also higher.

We also document positive assortative matching of men and women with regard to education and personality traits. The assortative matching tends to lead to higher levels of cooperation than would be observed under random matching of personality types. Simulation results show that eliminating the positive sorting decreases the gender gap of accepted wage from 15.2 percent to 13.5 percent, largely because of the reduction in the proportion of households behaving cooperatively.

We use the model to simulate time allocations that would result if women would receive the same wage offer equation as men. Simulation results show that women would work about 4.2 hours more per week and the accepted wage gap would be eliminated, with women having 2.45 percent higher wages than men.

There are several ways that our analysis could be extended in future work. First, we focused on individuals age 30-50 who do not have children under the age of 8 living in their home. Children and their effects on housework and labor supply decisions could be explicitly incorporated into the model. Second, our model viewed the household decision as a static decision at each age, but it could be extended to an explicit dynamic life-cycle framework.

3.8. Appendix

3.8.1. Comparison of estimated parameters based on restricted sample 1 and restricted sample 2

In this appendix, we compare the estimation results obtained using the sample that excludes the households with any child age less than 8 (sample 1, the one that was used in the estimation for the paper) and the results obtained using a more restrictive sample that excludes households with any child age less than 14 (sample 2). The left panel in table 58 displays the key differences between the two samples. When restricting the sample to households without dependents below age 14, the housework time of both spouses is reduced. The average housework hours of husbands decreases from 18.19 to 14.29, and the average hours of wives decreases from 27.87 to 20.27. For females, the reduction in housework time is associated with an increase in hours supplied to the labor market. The working hours in sample 2 are 3 hours more than that in sample 1.

The differences in the parameter estimates reasonably reflect different features of these two samples. Sample 1 also has different home production technology (δ_1,δ_2) that generates a greater male-female division of labor. Woman appear to be relatively more efficient in the home sector when children are young compared to when children are older. Therefore, we observe more hours of home production but fewer hours for market labor supply for females in sample 1. The greater division of labor also is associated with a higher degree of household cooperation. Our estimation shows that 38.7% households choose to cooperate in sample 1, whereas only 19.3% households cooperate in sample 2.

Table 58: Estimated parameters under two alternative estimation samples

	Data	Sample 1	Sample 2	Parameters	Sample 1	Sample 2
Male	Working Hours	44.10	44.03	λ_m	0.3625	0.3584
	S.D.	(9.54)	(9.70)		(0.1109)	(0.1220)
	Housework	18.19	14.29	δ_m	0.1367	0.1201
	S.D.	(12.86)	(10.62)		(0.0755)	(0.0744)
Female	Working Hours	33.48	36.29	λ_f	0.3365	0.3289
	S.D.	(12.31)	(13.14)		(0.0960)	(0.1123)
	Housework	27.87	20.27	δ_f	0.1936	0.1416
	S.D.	(17.59)	(13.48)		(0.1237)	(0.1103)

Note: while we only show selected estimates for sample 2 in this table, the full list of parameters is available upon request.

3.8.2. Algorithm used to solve and select equilibrium under non-cooperative regime

1. We solve the optimal allocation $(h_m^*, h_f^*, \tau_m^*, \tau_f^*)$ without constraint using the best reaction arrays (R_m, R_f):

$$\begin{aligned} R_m(h_m^*, \tau_m^*)(h_f, \tau_f) &= \arg\max_{h_m, \tau_m} \lambda_m \ln l_m + (1-\lambda_m) \ln K \\ R_f(h_f^*, \tau_f^*)(h_m, \tau_m) &= \arg\max_{h_f, \tau_f} \lambda_f \ln l_f + (1-\lambda_f) \ln K \end{aligned}$$

2. If h_m^* and h_f^* are both non-negative, this is the equilibrium with inner allocation.

3. If either $h_m^* < 0$ or $h_f^* < 0$, then the equilibrium is a boundary case.

   (a) Guess and verify the case when female is out if labor force $h_f = 0$

      i. Guess: the allocation $(\hat{h}_m, \hat{\tau}_m, 0, \hat{\tau}_f)$ with the constraint female is out of labor force

      ii. Verify: whether this allocation $(\hat{h}_m, \hat{\tau}_m, 0, \hat{\tau}_f)$ is an equilibrium by inserting $\left(\hat{h}_m, \hat{\tau}_m\right)$ into $R_f(\widetilde{h}_f, \widetilde{\tau}_f)(\hat{h}_m, \hat{\tau}_m)$. It is truly an equilibrium if and only if $\widetilde{h}_f \leq 0$

   (b) Guess and verify the case when male is out if labor force $h_m = 0$

      i. Guess: the allocation $(0, \hat{\tau}_m, \hat{h}_f, \hat{\tau}_f)$ with the constraint female is out of labor force

ii. Verify: whether this allocation $(0, \hat{\tau}_m, \hat{h}_f, \hat{\tau}_f)$ is an equilibrium by inserting $\left(\hat{h}_f, \hat{\tau}_f\right)$ into $R_m(\widetilde{h}_m, \widetilde{\tau}_m)(\hat{h}_f, \hat{\tau}_f)$. It is truly an equilibrium if and only if $\widetilde{h}_m \leq 0$

(c) If only one in (a) and (b) is an equilibrium, choose this one.

(d) if both (a) and (b) are equilibria, we choose the one female is out of labor force. $(\hat{h}_m, \hat{\tau}_m, 0, \hat{\tau}_f)$

3.8.3. List of Moments used in Estimation

This appendix provides a list of the moments used in estimation. Each moment is based on all the households in the estimation sample, so that the N is the same value for each of the moments (N is same for males and females as they are husbands and wives). In each case, we use indicator functions to select particular subsets of the sample. For example, the average wage for the males in the sample is the fraction of individuals who are male and who are working times the average wage for those individuals plus the fraction of individuals who are not male or not working times zero.

Husband's employment rate for sub-groups

1.I(Husband's Work Hours > 0)$I(Education = High)$ 2.I(Husband's Work Hours > 0)$I(Education = Low)$ 3.I(Husband's Work Hours > 0)I(First principal component $\leq$ $rank(33\%)$) 4.I(Husband's Work Hours > 0)I($rank(33\%)$ <First principal component$\leq$ $rank(66\%)$) 5.I(Husband's Work Hours > 0)I($rank(66\%)$ <First principal component) 6.I(Husband's Work Hours > 0)I(Second principal component $\leq$ $rank(33\%)$) 7.I(Husband's Work Hours > 0)I($rank(33\%)$ <Second principal component $\leq$ $rank(66\%)$) 8.I(Husband's Work Hours > 0)I($rank(66\%)$ <Second principal component) Wife's employment rate for sub-groups 9.I(Wife's Work Hours > 0)$I(Education = High)$ 10.I(Wife's Work Hours > 0)$I(Education = Low)$ 11.I(Wife's Work Hours > 0)I(First principal component$\leq$ $rank(33\%)$) 12.I(Wife's Work Hours > 0)I($rank(33\%)$ <First principal component$\leq$ $rank(66\%)$)

13.I(Wife's Work Hours $>$ 0)$I(rank(66\%)$ <First principal component) 14.I(Wife's Work Hours $>$ 0)I(Second principal component$\leq$ $rank(33\%)$) 15.I(Wife's Work Hours $>$ 0)$I(rank(33\%)$ <Second principal component$\leq$ $rank(66\%)$) 16.I(Wife's Work Hours $>$ 0)$I(rank(66\%)$ <Second principal component)

Husband's average wage for sub-groups

17. Husband's average wage I(Husband's Work Hours $>$ 0)$I(Education = High)$ 18. Husband's average wage I(Husband's Work Hours $>$ 0)$I(Education = Low)$ 19. Husband's average wage I(Husband's Work Hours $>$ 0)I(First principal component$\leq$ $rank(33\%)$) 20. Husband's average wage I(Husband's Work Hours $>$ 0)$I(rank(33\%)$ <First principal component$\leq$ $rank(66\%)$) 21. Husband's average wage I(Husband's Work Hours $>$ 0)$I(rank(66\%)$ <First principal component) 22. Husband's average wage I(Husband's Work Hours $>$ 0)I(Second principal component$\leq$ $rank(33\%)$) 23. Husband's average wage I(Husband's Work Hours $>$ 0)$I(rank(33\%)$ <Second principal component$\leq$ $rank(66\%)$) 24. Husband's average wage I(Husband's Work Hours $>$ 0)$I(rank(66\%)$ <Second principal component)

Wife's average wage for sub-groups

25. Wife's average wage I(Wife's Work Hours $>$ 0)$I(Education = High)$ 26. Wife's average wage I(Wife's Work Hours $>$ 0)$I(Education = Low)$ 27. Wife's average wage I(Wife's Work Hours $>$ 0)I(First principal component$\leq$ $rank(33\%)$) 28. Wife's average wage I(Wife's Work Hours $>$ 0)$I(rank(33\%)$ <First principal component$\leq$ $rank(66\%)$) 29. Wife's average wage I(Wife's Work Hours $>$ 0)$I(rank(66\%)$ <First principal component) 30. Wife's average wage I(Wife's Work Hours $>$ 0)I(Second principal component$\leq$ $rank(33\%)$) 31. Wife's average wage I(Wife's Work Hours $>$ 0)$I(rank(33\%)$ <Second principal component$\leq$ $rank(66\%)$) 32. Wife's average wage I(Wife's Work Hours $>$ 0)$I(rank(66\%)$ <Second principal component)

Husband's average working hours for sub-groups

33. Husband's average working hours$I$(Husband's Work Hours $>$ 0)$I(Education = High)$ 34. Husband's average working hours$I$(Husband's Work Hours $>$ 0)$I(Education = Low)$ 35. Husband's av-

erage working hours$I$(Husband's Work Hours $> 0)I$(First principal component$\leq$ $rank(33\%)$) 36. Husband's average working hours$I$(Husband's Work Hours $> 0)I(rank(33\%)$ <First principal component$\leq$ $rank(66\%)$) 37. Husband's average working hours$I$(Husband's Work Hours $> 0)I(rank(66\%)$ <First principal component) 38. Husband's average working hours$I$(Husband's Work Hours $> 0)I$(Second principal component$\leq$ $rank(33\%)$) 39. Husband's average working hours$I$(Husband's Work Hours $>$ $0)I(rank(33\%)$ <Second principal component$\leq$ $rank(66\%)$) 40. Husband's average working hours$I$(Husband's Work Hours $> 0)I(rank(66\%)$ <Second principal component)

Wife's average working hours for sub-groups

41. Wife's average working hours I(Wife's Work Hours $> 0)I(Education = High)$ 42. Wife's average working hours I(Wife's Work Hours $> 0)I(Education = Low)$ 43. Wife's average working hours I(Wife's Work Hours $> 0)I$(First principal component $\leq$ $rank(33\%)$) 44. Wife's average working hours I(Wife's Work Hours $> 0)I(rank(33\%)$ <First principal component $\leq$ $rank(66\%)$) 45. Wife's average working hours I(Wife's Work Hours $> 0)I(rank(66\%)$ <First principal component) 46. Wife's average working hours I(Wife's Work Hours $> 0)I$(Second principal component $\leq$ $rank(33$) 47. Wife's average working hours I(Wife's Work Hours $> 0)I(rank(33\%)$ <Second principal component $\leq$ $rank(66\%)$) 48. Wife's average working hours I(Wife's Work Hours $> 0)I(rank(66\%)$ <Second principal component)

Husband's average housework hours for sub-groups

49. Husband's average housework hours $I(Education = High)$ 50. Husband's average housework hours $I(Education = Low)$ 51. Husband's average housework hours I(First principal component$\leq$ $rank(33\%)$) 52. Husband's average housework hours $I(rank(33\%)$ <First principal component$\leq$ $rank(66\%)$) 53. Husband's average housework hours $I(rank(66\%)$ <First principal component) 54. Husband's average housework hours I(Second principal component$\leq$ $rank(33\%)$) 55. Husband's average housework hours $I(rank(33\%)$ <Second principal component$\leq$ $rank(66\%)$) 56. Husband's average housework hours $I(rank(66\%)$ <Second principal component)

Wife's average housework hours for sub-groups

57. Wife's average housework hours $I(Education = High)$ 58. Wife's average housework hours $I(Education = Low)$ 59. Wife's average housework hours I(First principal component $\leq rank(33\%)$) 60. Wife's average housework hours $I(rank(33\%)$ <First principal component $\leq rank(66\%)$) 61. Wife's average housework hours $I(rank(66\%)$ <First principal component) 62. Wife's average housework hours I(Second principal component $\leq rank(33\%)$) 63. Wife's average housework hours $I(rank(33\%)$ <Second principal component $\leq rank(66\%)$) 64. Wife's average housework hours $I(rank(66\%)$ <Second principal component)

Moments for the distribution of time allocation

65. Prob. of husband's working hours in [0,40]

66. Prob. of husband's working hours in (40,46]

67. Prob. of husband's working hours in (46,60]

68. Prob. of husband's housework hours in [0,11]

69. Prob. of husband's housework hours in (11,20]

70. Prob. of husband's housework hours in (20,56]

71. Prob. of wife's working hours in [0,24]

72. Prob. of wife's working hours in (24,38]

73. Prob. of wife's working hours in (38,60]

74. Prob. of wife's housework hours in [0,16]

75. Prob. of wife's housework hours in (16,34]

76. Prob. of wife's housework hours in (34,56]

77. Corr. between husband's working hours and wife's working hours

78. Corr. between husband's working hours and wife's housework hours

79. Corr. between husband's working hours and husband's housework hours

80. Corr. between husband's housework hours and wife's working hours

81. Corr. between husband's housework hours and wife's housework hours

82. Corr. between wife's working hours and wife's housework hours

83. S.D. of husband's log wage$|$(Husband's working hours > 0)

84. S.D. of wife's log wage$|$(Wife's working hours > 0)

85. Corr. between wife's wage and husband's wage

BIBLIOGRAPHY

M. Almlund, A. L. Duckworth, J. J. Heckman, and T. D. Kautz. Personality psychology and economics. Technical report, National Bureau of Economic Research, 2011.

J. G. Altonji and R. M. Blank. Race and gender in the labor market. *Handbook of labor economics*, 3:3143–3259, 1999.

J. G. Altonji and L. M. Segal. Small-sample bias in gmm estimation of covariance structures. *Journal of Business & Economic Statistics*, 14(3):353–366, 1996.

D. H. Autor, L. F. Katz, and M. S. Kearney. Trends in us wage inequality: Revising the revisionists. *The Review of economics and statistics*, 90(2):300–323, 2008.

N. Baum-Snow and R. Pavan. Understanding the city size wage gap. *The Review of Economic Studies*, 79(1):88–127, 2012.

G. S. Becker. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, by Gary S. Becker,...* London, 1964.

G. S. Becker. *A Treatise on the Family.* Harvard university press, 1981.

M. Bertrand. New perspectives on gender. *Handbook of labor economics*, 4:1543–1590, 2011.

F. D. Blau and L. M. Kahn. Swimming upstream: Trends in the gender wage differential in the 1980s. *Journal of labor Economics*, pages 1–42, 1997.

F. D. Blau and L. M. Kahn. The us gender pay gap in the 1990s: Slowing convergence. *Industrial & Labor Relations Review*, 60(1):45–66, 2006.

A. S. Blinder. Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, pages 436–455, 1973.

L. Borghans, A. L. Duckworth, J. J. Heckman, and B. Ter Weel. The economics and psychology of personality traits. *Journal of human Resources*, 43(4):972–1059, 2008.

F. P. Boscoe, K. A. Henry, and M. S. Zdeb. A nationwide comparison of driving distance versus straight-line distance to hospitals. *The Professional Geographer*, 64(2):188–196, 2012.

F. Bourguignon. Rationalité individuelle ou rationalité stratégique: le cas de l'offre familiale de travail. *Revue économique*, pages 147–162, 1984.

N. Braakmann. The role of psychological traits for the gender gap in full-time employment and wages: Evidence from germany. *SOEP Paper*, (162), 2009.

R. V. Burkhauser, K. A. Couch, and D. C. Wittenburg. Who minimum wage increases bite:

An analysis using monthly data from the sipp and the cps. *Southern Economic Journal*, pages 16–40, 2000.

M. Busso, J. Gregory, and P. Kline. Assessing the incidence and efficiency of a prominent place based policy. *The American Economic Review*, 103(2):897–947, 2013.

P. Cahuc, F. Postel-Vinay, and J.-M. Robin. Wage bargaining with on-the-job search: Theory and evidence. *Econometrica*, 74(2):323–364, 2006.

D. Card and A. B. Krueger. Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania. *The American Economic Review*, 84(4): 772–793, 1994.

D. Card and A. B. Krueger. Minimum wages and employment: a case study of the fast-food industry in new jersey and pennsylvania: reply. *The American Economic Review*, 90(5): 1397–1420, 2000.

S. Cattan. Psychological traits and the gender wage gap. *The Institute for Fiscal Studies*, 2013.

M. K. Chan. A dynamic model of welfare reform. *Econometrica*, 81(3):941–1001, 2013.

P.-A. Chiappori. Rational household labor supply. *Econometrica: Journal of the Econometric Society*, pages 63–90, 1988.

D. A. Cobb-Clark and S. Schurer. The stability of big-five personality traits. *Economics Letters*, 115(1):11–15, 2012.

D. Coen-Pirani. Understanding gross worker flows across us states. *Journal of Monetary Economics*, 57(7):769–784, 2010.

R. Cohen, A. Lai, and C. Steindel. The effects of marginal tax rates on interstate migration in the us. *New Jersey Department of the Treasury*, page 10, 2011.

R. Croson and U. Gneezy. Gender differences in preferences. *Journal of Economic literature*, 47(2):448–474, 2009.

M. Cubel, A. Nuevo-Chiquero, S. Sanchez-Pages, and M. Vidal-Fernandez. Do personality traits affect productivity? evidence from the laboratory. *The Economic Journal*, 126 (592):654–681, 2016.

F. Cunha and J. Heckman. The technology of skill formation. *American Economic Review*, 97(2):31–47, 2007.

F. Cunha and J. J. Heckman. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of human resources*, 43(4):738–782, 2008.

F. Cunha, J. J. Heckman, and S. M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

S. Dahmann and S. Anger. The impact of education on personality: Evidence from a german high school reform. 2014.

H. David, A. Manning, and C. L. Smith. The contribution of the minimum wage to us wage inequality over three decades: a reassessment. *American Economic Journal: Applied Economics*, 8(1):58–99, 2016.

D. Deere, K. M. Murphy, and F. Welch. Employment and the 1990-1991 minimum-wage hike. *The American Economic Review*, 85(2):232–237, 1995.

D. Del Boca and C. Flinn. Endogenous household interaction. *Journal of Econometrics*, 166(1):49–65, 2012.

D. Del Boca and C. J. Flinn. Rationalizing child-support decisions. *The American Economic Review*, pages 1241–1262, 1995.

D. Del Boca, C. Flinn, and M. Wiswall. Household choices and child development. *The Review of Economic Studies*, 81(1):137–185, 2014.

A. Dube, S. Naidu, and M. Reich. The economic effects of a citywide minimum wage. *ILR Review*, 60(4):522–543, 2007.

A. Dube, T. W. Lester, and M. Reich. Minimum wage effects across state borders: Estimates using contiguous counties. *The Review of Economics and Statistics*, 92(4):945–964, 2010.

A. Dube, T. W. Lester, and M. Reich. Minimum wage shocks, employment flows, and labor market frictions. *Journal of Labor Economics*, 34(3):663–704, 2016.

A. Dupuy and A. Galichon. Personality traits and the marriage market. *Journal of Political Economy*, 122(6):1271–1319, 2014.

Z. Eckstein and K. I. Wolpin. Estimating a market equilibrium search model from panel data on individuals. *Econometrica: Journal of the Econometric Society*, pages 783–808, 1990.

Z. Eckstein, S. Ge, and B. Petrongolo. Job and wage mobility with minimum wages and imperfect compliance. *Journal of Applied Econometrics*, 26(4):580–612, 2011.

M. Enrico. Local labor markets. *Handbook of Labor Economics*, 4:1237–1313, 2011.

J. M. Fletcher. The effects of personality traits on adult labor market outcomes: Evidence from siblings. *Journal of Economic Behavior & Organization*, 89:122–135, 2013.

C. Flinn and J. Heckman. New methods for analyzing structural models of labor force dynamics. *Journal of Econometrics*, 18(1):115–168, 1982.

C. Flinn and J. Mullins. Labor market search and schooling investment. *International Economic Review*, 56(2):359–398, 2015.

C. Flinn, A. Gemici, and S. Laufer. Search, matching and training. *Review of Economic Dynamics*, 25:260–297, 2017.

C. J. Flinn. Minimum wage effects on labor market outcomes under search, matching, and endogenous contact rates. *Econometrica*, 74(4):1013–1062, 2006.

A. Gemici. Family migration and labor market outcomes. *Manuscript, New York University*, 2011.

A. Gemici and S. Laufer. Marriage and cohabitation. *New York University, mimeo*, 2011.

E. L. Glaeser, J. D. Gottlieb, et al. The economics of place-making policies. *Brookings Papers on Economic Activity*, 39(1 (Spring)):155–253, 2008.

L. R. Goldberg. The development of markers for the big-five factor structure. *Psychological assessment*, 4(1):26, 1992a.

L. R. Goldberg. The development of markers for the big-five factor structure. *Psychological assessment*, 4(1):26, 1992b.

S. D. Gosling, P. J. Rentfrow, and W. B. Swann. A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528, 2003.

C. Gourieroux and A. Monfort. *Simulation-based econometric methods*. Oxford university press, 1996.

M. J. Greenwood. Research on internal migration in the united states: a survey. *Journal of Economic Literature*, pages 397–433, 1975.

Z. Griliches. Estimating the returns to schooling: Some econometric problems. *Econometrica: Journal of the Econometric Society*, pages 1–22, 1977.

J. Heckman and B. Singer. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica: Journal of the Econometric Society*, pages 271–320, 1984.

J. J. Heckman. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. 1981.

J. J. Heckman and L. K. Raut. Intergenerational long-term effects of preschool-structural estimates from a discrete dynamic programming model. *Journal of econometrics*, 191(1): 164–175, 2016.

J. J. Heckman, J. Stixrud, and S. Urzua. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482, 2006.

J. J. Heckman, S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz. The rate of return to the highscope perry preschool program. *Journal of public Economics*, 94(1):114–128, 2010.

G. Heineck. Does it pay to be nice? personality and earnings in the united kingdom. *Industrial & Labor Relations Review*, 64(5):1020–1038, 2011.

J. Horowitz. Identification and diagnosis of specification errors in the multinomial logit model. *Transportation Research Part B: Methodological*, 15(5):345–360, 1981.

J. J. Horton. Price floors and employer preferences: Evidence from a minimum wage experiment. *Unpublished paper*, 2017.

Y. Hu, M. Shum, W. Tan, and R. Xiao. A simple estimator for dynamic models with serially correlated unobservables. *Journal of Econometric Methods*, 2015.

M. Huggett, G. Ventura, and A. Yaron. Sources of lifetime inequality. *The American Economic Review*, pages 2923–2954, 2011.

E. Jardim, M. C. Long, R. Plotnick, E. van Inwegen, J. Vigdor, and H. Wething. Minimum wage increases, wages, and low-wage employment: Evidence from seattle. Technical report, National Bureau of Economic Research, 2017.

O. P. John and S. Srivastava. The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2(1999):102–138, 1999.

C. Joubert and P. Todd. Expanding pension coverage: A dynamic analysis of chile's pension reform. Technical report, unpublished manuscript, 2016.

M. P. Keane and R. M. Sauer. A computationally practical simulation estimation algorithm for dynamic panel data models with unobserved endogenous state variables. *International Economic Review*, 51(4):925–958, 2010.

M. P. Keane and K. I. Wolpin. The career decisions of young men. *Journal of political Economy*, 105(3):473–522, 1997.

M. P. Keane and K. I. Wolpin. The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review*, 42(4):1051–1103, 2001.

J. Kennan and J. R. Walker. The effect of expected income on individual migration decisions. *Econometrica*, 79(1):211–251, 2011.

P. Kline. Place based policies, heterogeneity, and agglomeration. *The American Economic Review*, 100(2):383–387, 2010.

P. Kline and E. Moretti. Local economic development, agglomeration economies, and the big push: 100 years of evidence from the tennessee valley authority. *The Quarterly Journal of Economics*, 129(1):275–331, 2013.

J. Li and N. Powdthavee. Does increasing schooling improve later health habits? evidence from the school reforms in australia. Technical report, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne, 2014.

I. Losoncz. Personality traits in hilda1. *Australian Social Policy No. 8*, page 169, 2009.

S. Lundberg. Personality and marital surplus. *IZA Journal of Labor Economics*, 1(1):1, 2012.

S. Lundberg. The college type: Personality and educational inequality. *Journal of Labor Economics*, 31(3):421–441, 2013.

S. Lundberg and R. A. Pollak. Separate spheres bargaining and the marriage market. *Journal of political Economy*, pages 988–1010, 1993.

J. Mabli and C. Flinn. On-the-job search, minimum wages, and labor market outcomes in an equilibrium bargaining framework. In *2007 Meeting Papers*, number 791. Society for Economic Dynamics, 2007.

A. Manning and B. Petrongolo. How local are labor markets? evidence from a spatial job search model. *The American Economic Review*, forcecoming, 2017.

M. Manser and M. Brown. Marriage and household decision-making: A bargaining analysis. *International economic review*, pages 31–44, 1980.

M. B. McElroy and M. J. Horney. Nash-bargained household decisions: Toward a generalization of the theory of demand. *International economic review*, pages 333–349, 1981.

T. McKinnish. Cross-state differences in the minimum wage and out-of-state commuting by low-wage workers. *Regional Science and Urban Economics*, 64:137–147, 2017.

C. Meghir, R. Narita, and J.-M. Robin. Wages and informality in developing countries. *The American Economic Review*, 105(4):1509–1546, 2015.

R. Molloy, C. L. Smith, and A. Wozniak. Internal migration in the united states. *The Journal of Economic Perspectives*, 25(3):173–196, 2011.

J. Monras. Minimum wages and spatial equilibrium: Theory and evidence. *IZA Discussion Paper No. 9460.*, 2015.

E. Moretti and D. Wilson. The effect of state taxes on the geographical location of top earners: evidence from star scientists. *The American Economic Review*, 107(7):1858–1903, 2017.

D. T. Mortensen and C. A. Pissarides. Job creation and job destruction in the theory of unemployment. *The Review of Economic Studies*, 61(3):397–415, 1994.

G. Mueller and E. Plug. Estimating the effect of personality on male and female earnings. *Industrial & Labor Relations Review*, 60(1):3–22, 2006.

D. A. Neal and W. R. Johnson. The role of premarket factors in black-white wage differences. *Journal of political Economy*, 104(5):869–895, 1996.

D. Neumark. The employment effects of minimum wages: Evidence from a prespecified research design the employment effects of minimumwages. *Industrial Relations: A Journal of Economy and Society*, 40(1):121–144, 2001.

D. Neumark, J. I. Salas, and W. Wascher. More on recent evidence on the effects of minimum wages in the united states. *IZA Journal of Labor Policy*, 3(1):24, 2014a.

D. Neumark, J. I. Salas, and W. Wascher. Revisiting the minimum wage - employment debate: Throwing out the baby with the bathwater? *ILR Review*, 67(3_suppl):608–648, 2014b.

E. K. Nyhus and E. Pons. The effects of personality on earnings. *Journal of Economic Psychology*, 26(3):363–384, 2005.

R. Oaxaca. Male-female wage differentials in urban labor markets. *International economic review*, pages 693–709, 1973.

B. Petrongolo and C. A. Pissarides. Looking into the black box: A survey of the matching function. *Journal of Economic Literature*, 39(2):390–431, 2001.

C. S. Phibbs and H. S. Luft. Correlation of travel time on roads versus straight line distance. *Medical Care Research and Review*, 52(4):532–542, 1995.

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

M. Rosenberg. *Society and the adolescent self-image*, volume 11. 1965.

J. B. Rotter. Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied*, 80(1):1, 1966.

A. Şahin, J. Song, G. Topa, and G. L. Violante. Mismatch unemployment. *The American Economic Review*, 104(11):3529–3564, 2014.

J. F. Salgado. The five factor model of personality and job performance in the european community. *Journal of Applied psychology*, 82(1):30, 1997.

G. Saucier. Mini-markers: A brief version of goldberg's unipolar big-five markers. *Journal of personality assessment*, 63(3):506–516, 1994.

B. Schmutz and M. Sidibe. Frictional labor mobility. 2016.

S. Schurer, S. C. Kassenboehmer, and F. Leung. Testing the human capital model of education: Do universities shape their students' character traits? 2015.

A. Schwartz. Interpreting the effect of distance on migration. *Journal of Political Economy*, 81(5):1153–1169, 1973.

J. C. S. Serrato and O. Zidar. Who benefits from state corporate tax cuts? a local labor markets approach with heterogeneous firms. *The American Economic Review*, 106(9): 2582–2624, 2016.

I. Sorkin. Are there long-run effects of the minimum wage? *Review of Economic Dynamics*, 18(2):306–333, 2015.

P. Sullivan. A dynamic analysis of educational attainment, occupational choices, and job search*. *International Economic Review*, 51(1):289–317, 2010.

M. Summerfield, S. Freidin, M. Hahn, N. Li, N. Macalalad, L. Mundy, N. Watson, R. Wilkins, and M. Wooden. Hilda user manual–release 13, 2014.

M. Summerfield, S. Freidin, M. Hahn, N. Li, N. Macalalad, L. Mundy, N. Watson, R. Wilkins, and M. Wooden. Hilda user manual–release 14, 2015.

M. Tartari. Divorce and the cognitive achievement of children. *International Economic Review*, 56(2):597–645, 2015.

A. Terracciano, P. T. Costa, and R. R. McCrae. Personality plasticity after age 30. *Personality and Social Psychology Bulletin*, 32(8):999–1009, 2006.

A. Terracciano, R. R. McCrae, and P. T. Costa. Intra-individual change in personality stability and age. *Journal of Research in Personality*, 44(1):31–37, 2010.

P. Todd and W. Zhang. A dynamic model of personality, schooling, and occupational choice. 2017.

G. J. Van den Berg and G. Ridder. An empirical equilibrium search model of the labor market. *Econometrica*, pages 1183–1221, 1998.

M. Wooden. The measurement of cognitive ability in wave 12 of the hilda survey. 2013.

S. Yamaguchi. Tasks and heterogeneous human capital. *Journal of Labor Economics*, 30 (1):1–53, 2012.

C. Young and C. Varner. Millionaire migration and state taxation of top incomes: Evidence from a natural experiment. *National Tax Journal*, 64(2):255, 2011.