



# DISSERTATION

*Submitted in partial fulfillment of the requirements  
for the degree of*

**DOCTOR OF PHILOSOPHY  
ECONOMICS**

*Titled*  
**“THREE ESSAYS IN MACROECONOMICS”**

*Presented by*  
**Shutian Zeng**

*Accepted by*

Laurence Ales

4/27/15

---

**Co-Chair: Prof. Laurence Ales**

---

**Date**

Nicolas Petrosky-Nadeau

4/27/15

---

**Co-Chair: Prof. Nicolas Petrosky-Nadeau**

---

**Date**

*Approved by The Dean*

Robert M. Dammon

4/27/15

---

**Dean Robert M. Dammon**

---

**Date**

# Three Essays in Macroeconomics

by  
Shutian Zeng

Submitted in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy  
at  
Carnegie Mellon University  
Department of Economics  
Pittsburgh, Pennsylvania

Advised by:  
Professor Laurence Ales (co-chair),  
Professor Nicolas Petrosky-Nadeau (co-chair),  
Professor Shu Lin Wee,  
Professor Bryan Routledge

April 28, 2015

## **Abstract**

The first chapter argues that the countercyclical workforce occupational mobility could be caused by aggregate economic fluctuations, as a result, the workforce composition shifts over business cycle, which in turn can affect how the economy recovers from recession. Using the Current Population Survey, I find: a) workers were more likely to separate from their occupations during last two recessions, b) the deteriorated economic conditions during last recession significantly reduced the average productivity change upon occupation switch, which suggests the additional occupation switches were triggered by worsened economic conditions. To demonstrate how workforce composition could play a role in business cycle, I calibrate a model with aggregate TFP, occupational productivity, match quality and increasing productivity in occupational tenure. Workers can be separated endogenously from current occupation due to low continuation value for the firm. Aggregate TFP shifts the threshold of endogenous separation, and less productive workers are fired first in a recession, leaving workers who are in expanding occupations, have better match quality or longer occupational tenure in the workforce. The output recovers faster than the labor market because newly hired workers tend to concentrate in expanding occupations and have better match quality than before. The model is able to generate faster output recovery, slow labor market recovery after a recession and countercyclical endogenous occupation mobility rate. The result suggests that other than the total amount of labor input, the workforce composition change could also plays an important role in real business cycle and this channel shall be explored more in the future.

The second chapter, co-authored with Nicolas Petrosky-Nadeau and Etienne Wasmer, studies the shopping time over business cycle. Renewed interest in macroeconomic theories of search frictions in the goods market requires a deeper understanding of the cyclical properties of the intensive margins in this market. Using the American Time Use Survey we construct a shopping time indicator, both searching and purchasing goods, based on 25 time use categories (out of more than 400 time use categories). We find that average time spent shopping declined in the aggregate over the period 2008-2010 compared to 2005-2007. The decline was largest for the unemployed who went from spending more time shopping for goods than the employed to roughly the same, or

even less, time. Cross-state and individual regressions indicate procyclical consumer shopping in the goods market, and refute models in which price comparisons are a driver of business cycles.

The third chapter proposes a new way to model the personal income process with stochastic monthly transitions between occupations and labor force statuses. Log monthly income while employment is decomposed into two components, the first one is determined by personal characteristics and evolves deterministically, and the second components is determined by the occupation which change over time stochastically. The heterogenous transition probabilities between labor force statuses and occupations can generate heterogenous life-cycle income profiles. Transition probabilities are decomposed into: a) categorical probabilities which govern the career development and moving in and out of labor force, and b) conditional probabilities which specifies the new occupation upon promotion, demotion or re-employment. Two types of probabilities are estimated separated and combined to construct the transition probabilities. Using the estimated transition probabilities, I can construct the distribution of income at any time for any initial occupation and personal characteristics. The results highlight that heterogenous people with the same initial income can have very different income growth rate due to different starting occupation and personal characteristics. Lastly, by decomposing the variance of income growth rate into expected individual uncertainty and variance of heterogeneous income profile, the model predicts that 55-57% of the income growth variance is due to heterogeneous income profile before age 55.

# Contents

<b>1</b>	<b>Occupational Mobility, Workforce Composition and the Business Cycle</b>	<b>1</b>
1.1	Introduction . . . . .	1
1.2	Empirical Evidences . . . . .	3
1.2.1	Workforce Occupational Mobility . . . . .	4
1.2.2	Additional Occupation Switches During Last Recession . . . . .	9
1.2.3	Employment at Occupation Level . . . . .	13
1.3	The Economy . . . . .	14
1.3.1	Household . . . . .	15
1.3.2	Representative firm in occupation $o$ . . . . .	15
1.3.3	Search and Matching . . . . .	17
1.3.4	Separation and Hiring Decisions . . . . .	18
1.3.5	The Evolution of Economy . . . . .	19
1.3.6	Aggregation . . . . .	20
1.3.7	Competitive equilibrium . . . . .	21
1.4	Model Calibration . . . . .	21

1.5	Model Results . . . . .	22
1.5.1	Impulse Response Analysis . . . . .	24
1.6	Conclusion . . . . .	27
1.A	Occupation Mobility Estimation . . . . .	31
1.A.1	Current Population Survey Brief Introduction . . . . .	31
1.A.2	Matching respondents from consecutive months . . . . .	31
1.A.3	Estimate labor force and occupational transition rates . . . . .	34
1.A.4	EE employer mobility rate . . . . .	36
1.A.5	Data moments Robustness Check . . . . .	36
1.B	Equivalent Economy . . . . .	38
1.C	Solve the Model Numerically . . . . .	39
1.C.1	Calibrating $H(\tau)$ . . . . .	39
1.C.2	Model Algorithm . . . . .	39
<b>2</b>	<b>Shopping Time</b>	<b>41</b>
2.1	Introduction . . . . .	41
2.2	The cyclical pattern of shopping time . . . . .	44
2.2.1	A setup for thinking about shopping time . . . . .	44
2.2.2	Surplus sharing with a fixed supply of the search good $x$ . . . . .	47
2.3	Searching for goods and services in the ATUS . . . . .	48
2.4	Aggregate trends in shopping time . . . . .	49

2.5	Shopping time and the business cycle . . . . .	54
2.5.1	The cycle and goods market search . . . . .	54
2.5.2	Cross-state variations in shopping time . . . . .	58
2.6	Individual regressions . . . . .	60
2.6.1	Household income . . . . .	60
2.6.2	Individual income . . . . .	62
2.6.3	Market work and shopping time . . . . .	63
2.7	Conclusion . . . . .	64
2.A	Theory . . . . .	67
2.A.1	Individual consumption probability: a more general specification . . . . .	67
2.A.2	Non-separable utility . . . . .	67
2.A.3	Price setting - Nash bargaining: . . . . .	68
2.A.4	Endogenous supply of the search good $x$ . . . . .	69
2.A.5	Price posting with homogeneous consumers and a given distribution of prices	71
2.A.6	Correlation between income, shopping time, and working time . . . . .	72
2.A.7	Supplemental Appendix to Section 2.A.4: Optimal consumption choice when $x$ is endogenous . . . . .	73
2.B	Data . . . . .	75
2.B.1	Shopping Time in the American Time Use Survey . . . . .	75
2.B.2	Labor force status in the ATUS . . . . .	77
<b>3</b>	<b>Individual Occupation Switch: the Determinants and the Impact on Income</b>	<b>80</b>

3.1	Introduction . . . . .	80
3.2	Personal Income Process . . . . .	83
3.3	The CPS Data . . . . .	84
3.3.1	Occupation Rank . . . . .	85
3.4	Occupation and Labor Force Status Transition . . . . .	86
3.4.1	Employed Transitions . . . . .	89
3.4.2	Nonemployed Transitions . . . . .	101
3.5	Heterogeneous Income Profile . . . . .	108
3.5.1	Different initial conditions . . . . .	109
3.5.2	Heterogeneity vs Inequality . . . . .	111
3.6	Conclusion . . . . .	113
3.A	Occupation Choice . . . . .	116
3.A.1	Occupation Choice upon EE Promotion/Demotion . . . . .	116
3.A.2	Occupation Choice upon Re-employment . . . . .	116



# Chapter 1

## Occupational Mobility, Workforce Composition and the Business Cycle

### 1.1 Introduction

Occupational specific human capital has been argued to be an important determinant of productivity by recent literature. Early literature focuses on employer tenure and suggest employer tenure is a key determinant of earning. Neal (1995) and Parent (2000) argue the correlation is in fact due to the industry experience which is correlated with employer tenure. More recently, Kambourov and Manovskii (2009) demonstrated that once occupational tenure is taken into account, both employer and industry tenures have much less explanatory power in wage growth. Sullivan (2010) finds that both occupation and industry tenure are key determinants of wages. Such literature argue a large part of post-education human capital are occupational specific based on the found correlation between wage growth and occupational tenure while controlling for other observables. A 5-year occupational experience is associated with about 12% wage growth, which suggests one's productivity increases in occupational tenure.

On the other hand, workers leave their current occupations at a time varying rate. Whenever an experienced worker is replaced by a new worker in the same occupation, the new worker needs to accumulate occupational specific human capital from scratch. There are many

reasons that workers leave their current occupation at different aggregation levels. At individual level, one can leave her current occupation due to promotion, laying off, retirement or other personal issues. Different reasons have different effect on one's productivity upon switching occupation. At occupation level, each occupation faces its own shocks, therefore the occupational employment expands and contracts over time, which leads to worker flows between various occupations. At aggregate level, common shock to all occupations drives the aggregate employment, output and productivity.

By using the monthly Current Population Survey (CPS) between 1996 and 2013, I measure the rate at which workers leave their current occupation at monthly frequency. The results indicate that during the 2008-09 recession, workers were more likely to switch occupations, either through nonemployment or continuous employment. To further study if the average worker experienced the same productivity change upon occupation switch during recession, I compare the effect of an occupation switch on one's hourly wage before and during the recession, the result shows that deteriorated economic conditions were negatively associated with the average productivity change upon occupation switch. This suggests the additional occupation switches during recession may be due to worsened economic conditions.

To demonstrate how the aggregate shocks can affect the workforce composition over business cycle, I calibrate a model with aggregate TFP, occupational productivity, match quality and increasing productivity in occupational tenure. The labor market is segregated by occupation, and the output of a worker in a particular occupation is the product of all four levels of productivity: aggregate TFP, occupational productivity, match quality and tenure productivity. All productivities except tenure productivity follow Markov processes, and the tenure productivity is monotonic in tenure. Occupations endogenously break less experienced matches, and more experienced and productive workers are retained in the occupation. TFP shifts the overall threshold at which matches are endogenously separated. During a recession when the TFP is hit by negative shocks, a greater proportion of occupations lay off less productive and less experienced workers. More productive occupations face a less tight labor market and are able to hire more workers at lower cost. The flow of workers from low productive occupation to higher productive occupation, as well as from bad match quality to the possibly good match quality, offsets the drop in TFP and help output to recover faster than the labor market.

Previous literature has incorporated human capital or general intangible capital in growth analysis or business cycle analysis. Corrado, Hulten and Sichel (2005, 2009), Corrado and Hulten (2010) estimate the size of on-the-job accumulated human capital is a big part of the overall intangible capital that hadn't been accounted for by then. McGrattan and Prescott (2010) incorporates an intangible production sector into a neoclassical growth model to help explain the boom in the U.S. economy and low factor incomes in the 1990s. Danthine and Jin (2007) incorporates an intangible capital production sector with stochastic conversion from investment to capital. Conventional macroeconomic models treat physical capital, intangible capital and labor input as separate entities such that each can be adjusted independently at the margin according to inter-temporal optimality conditions. In my model the accumulation of occupational specific human capital is passive and it is embedded in the workforce. The composition of workforce, which gives the distribution of workers over different occupation, match quality and occupational tenure, matters as much as the total labor input itself.

The paper is organized as the following. Section 2 presents the empirical results constructed from CPS. Section 3 describes the theoretical model. Section 4 calibrates the model to match the data. Section 5 presents the model results and impulse response analysis, and section 6 concludes the paper.

## 1.2 Empirical Evidences

In this section I use monthly CPS to show that: a) workers are more likely to separate from their current occupation during recessions, b) switching occupation during recession are associated with lower productivity, c) the employment at occupation level fluctuates over time and the distribution of monthly change resembles a bell shape with heavier tails than a normal distribution. Both CPS and Panel Study of Income Dynamics (PSID) are the leading sources to study worker mobility in terms of occupations, industries and firms. Kambourov and Manovskii (2013) compares (Match) CPS and PSID and studies their applicability in different context. For my purpose, the monthly CPS is the more suitable choice due to its monthly frequency and better representativeness of the workforce. It is widely used to estimate various aggregate variables in the labor market. About 60,000 household, or 150,000 individuals are interviewed each month. Each individual stays in the

survey for 4 consecutive months, then leaves for 8 months, and returns for another 4 months before permanently leaving the sample. This longitudinal aspect of CPS allows me to identify the overlap between consecutive months. However, the CPS has its own shortcomings. First of all, CPS is only suitable to estimate monthly mobility rate due to its short longitudinal aspect. Secondly, only respondents in their 4th and 8th months in the survey are asked about their weekly earning. Thirdly, the CPS is not designed to be a longitudinal dataset, there does not exist a single identifier to link the same respondents from different months. I follow Madrian and Lefgren (1999) to match respondents based on household ID, household number and individual line number. Anyone with different sex, different race or excessive change of age between consecutive months is dropped. After 1994, the monthly CPS adapted the dependent interviewing procedure which greatly reduces the coding errors in occupation. Furthermore, BLS deliberately broke the linking between certain months in 1995 to protect privacy. Therefore, data from 1996 to 2013 are used for the following analysis. This sample period includes the last two recessions.

### 1.2.1 Workforce Occupational Mobility

An occupational switch can show up differently in the CPS depending on the duration of nonemployment, both unemployment and not in labor force. The duration of nonemployment between two jobs could be anywhere from zero to permanence<sup>1</sup>. If the nonemployment lasts more than a month, then this person would report nonemployment, otherwise this person would report employment after occupation switch in the next monthly survey. The former case is hard to identify due to the limited longitudinal dimension of CPS, but the latter is straightforward to identify. I show both channels that a worker can switch occupation.

The occupation switch through first channel is approximated by the workforce turnover rate, which measures the rate at which workers leave the workforce. People leave the workforce because of laying off, retirement or other personal issues. If the likelihood of permanently separating from previous occupation after leaving the workforce remains relatively stable, then this rate can be used to construct the occupational mobility rate with nonemployment duration longer than one month. The occupation switch through second channel is measured by the employment-to-

---

<sup>1</sup>Some occupational changes are within the same firm, therefore the period of nonemployment is likely to be zero or very short. The duration is permanent if the worker never work again.

employment (EE) occupation mobility rate, which directly measure the rate at which workers change occupation with nonemployment duration less than one month.

The first four interviews from CPS are used to construct the occupational mobility rate. The fact that CPS is address based implies that workforce mobility is potentially correlated with replacement of households at the same address. Households who moved may be more likely change their labor force status or occupation than those who stayed at the same address. This issue, if it exists, can be mitigated by only using the interviews in their first 4 months of the survey. In theory about three eighths of each month's sample should be matched with the following month's, in reality 34% of the sample were retained.

In terms of data processing, I first compute the (seasonally unadjusted) monthly rate, and the (seasonally unadjusted) quarterly rate is simply the 3-month average of month rates. I then use linear regression with fixed quarterly effect to remove the seasonality, such that the quarterly rates can be compared to the other seasonally adjusted time series published by the Bureau of Labor Statistics (BLS). The labor productivity is defined as the real output per hour of all person in the non-farm business sector and available from BLS.

### **Workforce Turnover Rate**

Workforce turnover rate is defined as the fraction of workforce who become nonemployed in the following month. The monthly rate is estimated as:

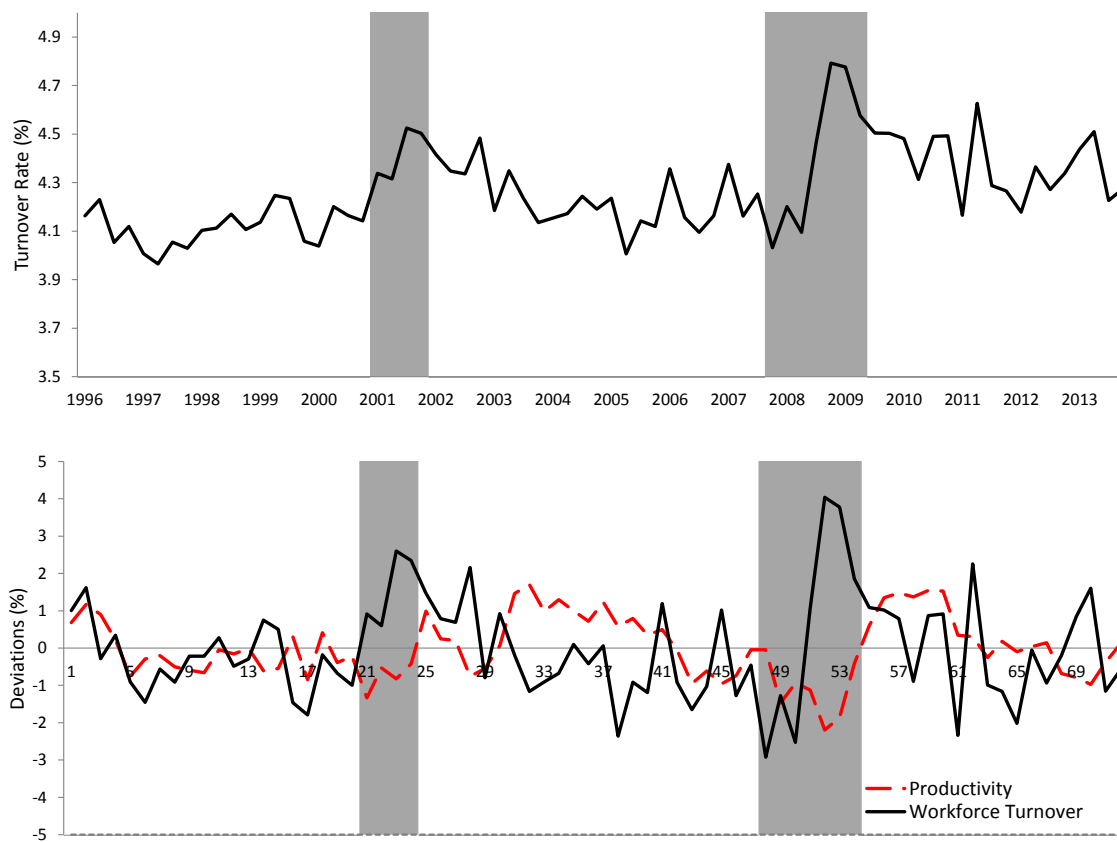
$$1 - \frac{\sum_{i \in E_t} EE_t \cdot w_i}{\sum_{i \in E_t} w_i} \quad (1.1)$$

where  $E_t$  is the subset of individuals who are employed during month  $t$ ,  $EE_t$  is an indicator function which equals one if individual  $i$  also reports employment in the following month. The  $w_i$  is the weight assigned by the CPS to reflect  $i$ 's representativeness in the population.

Figure 1.1 plots the quarterly workforce turnover rate and its comovement with labor productivity. The top panel plots the time series, where workforce turnover rate increased substantially during both recessions compared to the pre-recession levels, more than 0.3% during 2001 recession and more than 0.5% during the 2008-09 recession. The bottom panel plots the

percentage deviations of both workforce turnover rate and labor productivity from their HP trends. The workforce turnover rate shows mildly negative comovement with the productivity deviation with correlation of -15.5%.

**Figure 1.1 :Quarterly Workforce Turnover Rate**



a), The top panel plots the quarterly workforce turnover rate, which is the seasonally adjusted three-month average of monthly workforce turnover rate. The shaded areas are the NBER recession periods.

b), The bottom panel plots the percentage deviations of both workforce turnover rate and labor productivity from their HP trends. The black solid line represents the percentage deviation of workforce turnover rate and the red dash line represents the percentage deviation of productivity. A smoothing parameter of 1600 is used for HP filter. The productivity is the real output per hour of all person for non-farm business section published by BLS. The shaded areas are the NBER recession periods.

## EE Occupational Mobility Rates

Every month a small fraction of the workforce change occupation while reporting employment in consecutive months. High occupational mobility rate reduces the occupational tenure and the aggregate occupational specific human capital. I compute the monthly EE occupational mobility rate as the following:

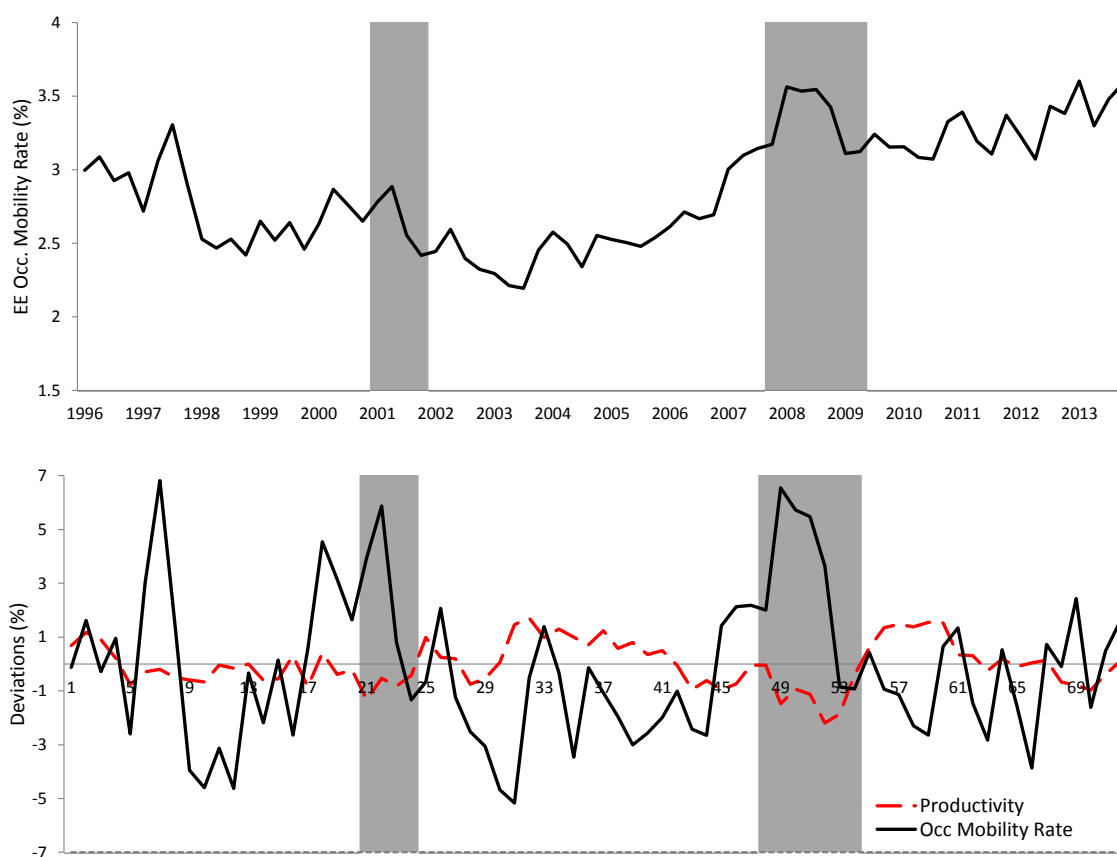
$$\frac{\sum_{i \in EE_t} OCC_i^d w_i}{\sum_{i \in EE_t} w_i} \quad (1.2)$$

where  $EE_t$  is the subsample who report employment in both month  $t$  and  $t + 1$ .  $OCC_i^d$  is an indicator function which equals one if respondent  $i$  changes occupation from month  $t$  to  $t + 1$ . I follow Moscarini and Thomsson (2007) to identify valid occupational changes. A valid occupational change requires change of the 3-digit occupation code for the primary work and also satisfying any of the following: a) class of workers (private firm, federal, government or self-employed) changed, b) the three-digit industry code changed, c) the respondent had looked for work in the past four weeks. The occupations are classified according to the standard occupational classification (SOC) published by BLS. According to 2010 SOC, the occupations are classified at four levels of aggregation: major group, minor group, broad occupation and detailed occupation. Four instance, football players share the same major group with fashion designers, although they require totally different skill sets. Football players even have the same minor group as actors, and only share the same broad occupation with coaches, although they have different detailed occupations. The broad occupation, or three-digit occupational code as in previous literature, is best suited for the definition of a career. In CPS, most respondents have their occupations coded at broad occupational, hence a change of their occupational code most likely corresponds to a career change. The occupation code is only available for respondents who are: a) employed, b) unemployed but have worked before, c) out of labor force but have worked in the last 12 months. The occupation code also changed twice in the sample period. In the appendix I show how to mitigate this issue.

Figure 1.2 plots the EE occupational mobility rate with labor productivity and NBER recession dates. The top panel plots the time series of the quarterly EE occupational mobility rate, the bottom panel plots its deviations from HP trend and compares it to the productivity deviation. The occupational mobility rate starts at around 3% and declines slightly to below 2.2% in 2003, and then rose afterwards. It hit 3.56% during the 2008-09 recession and remained at about

the same level for the rest of the sample. It seems to move in opposite direction with the labor productivity. Between two recessions, the labor productivity reaches its temporary peak right after EE occupational mobility rate reached all time lowest level. The EE occupational mobility reached its lowest point at 2.19% in 2003Q3, and the productivity had its largest positive deviation from trend at 1.71% in 2003Q4. During the 2008-09 recession, the EE occupational mobility rates had increased by about 0.5% from their pre-recession level, and the labor productivity reaches its largest negative deviation shortly afterwards.

**Figure 1.2 :EE Occupational Mobility Rate**



a), The top panel plots the quarterly EE occupational mobility rate. The shaded areas are the NBER recession periods.

b), The bottom panel plots the percentage deviations of both EE occupational mobility rate and labor productivity from their HP trends. The black solid line represents the percentage deviation of EE occupational mobility rate and the red dash line represents the percentage deviation of productivity. A smoothing parameter of 1600 is used for HP filter. The productivity is the real output per hour of all person for non-farm business section published by BLS. The shaded areas are the NBER recession periods.



## Workforce Occupational Mobility Rate

The workforce occupational mobility rate, which is defined as the fraction of workforce who leaves their current occupation regardless of duration of nonemployment in between, measures the rate at which workers are separated from their current occupation. It is the weighted average of occupational mobility rates of 50% among those who leave the workforce (workforce turnover) and the EE occupational mobility rate among those who stay in the workforce (one minus workforce turnover rate). The 50% is assumed since it cannot be estimated from CPS. 40% and 80% are also tested for robustness in the appendix, and the results are very similar other than the overall level of workforce occupational mobility rate. The mathematical expression is:

$$50\% \times \left(1 - \frac{\sum_{i \in E_t} EE_t \cdot w_i}{\sum_{i \in E_t} w_i}\right) + \frac{\sum_{i \in EE_t} OCC_i^d w_i}{\sum_{i \in EE_t} w_i} \times \left(\frac{\sum_{i \in E_t} EE_t \cdot w_i}{\sum_{i \in E_t} w_i}\right) \quad (1.3)$$

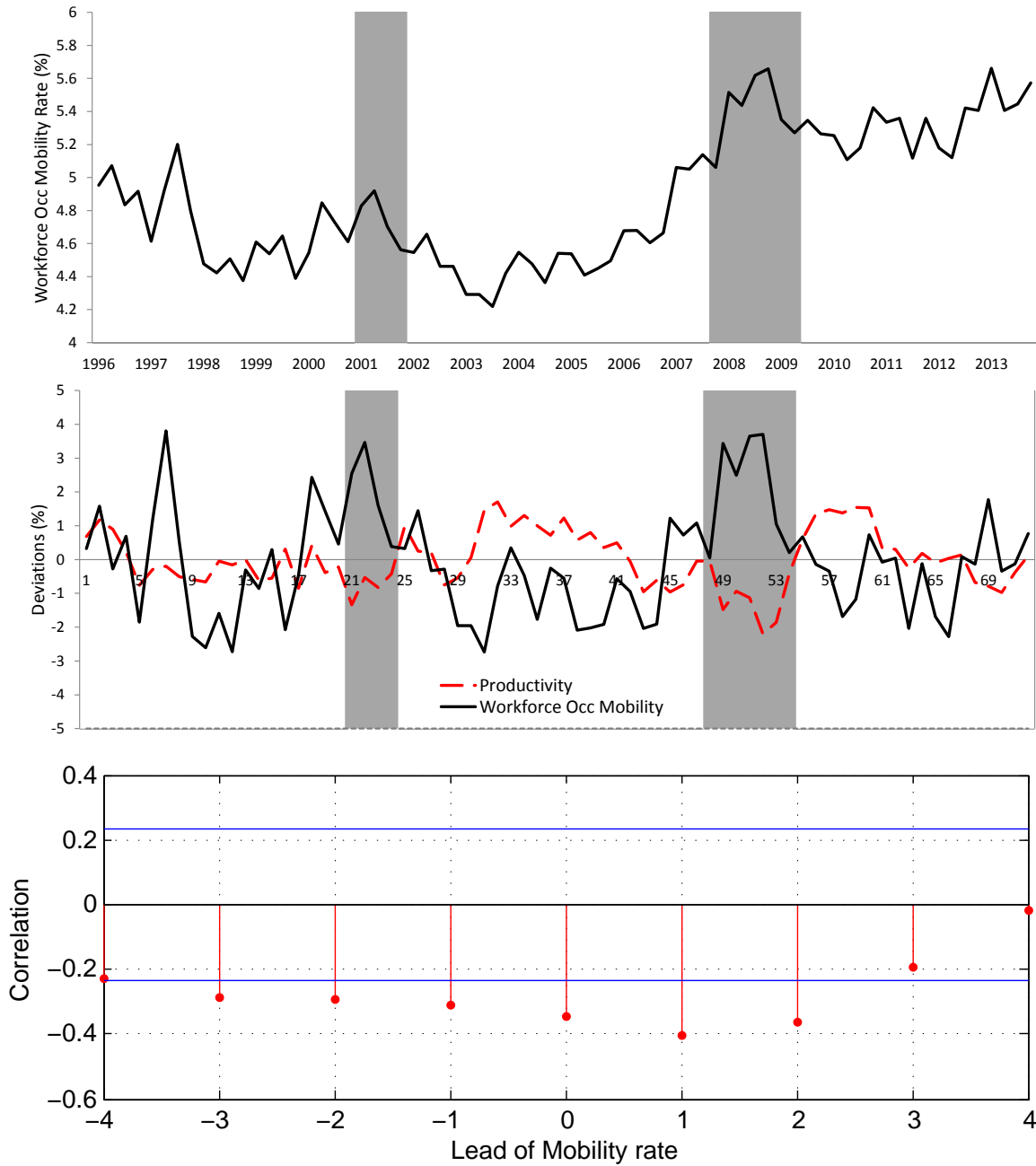
The workforce occupational mobility rate is plotted in figure 1.3. The top panel plots the time series of workforce occupational mobility rate. There was an upward trend after 2003, which may be due to overall demographic shifts like baby boomers retiring. The rate spiked briefly during both recessions. The middle panel plots the deviations of workforce occupational mobility rate and labor productivity from their HP trends. The mobility rate shows greater variations than the productivity, and two series are negatively correlated. The third panel plots the cross-correlation between two deviations. The contemporary correlation is -0.35 and the largest negative correlation is at lead one. This suggests the mobility rate may negatively lead the labor productivity.

Kambourov and Manovskii (2008) has several findings that are relevant to this study. They found that occupational switches are fairly permanent that only 20% of workers switching occupations return to their 3-digit occupation within 4 years. This means the mobility rates do not significantly overestimate the rate at which workers permanently switch occupations.

### 1.2.2 Additional Occupation Switches During Last Recession

In this part I investigate if deteriorated economic conditions adversely change the effect of an occupation switch on one's productivity, which is approximated by the hourly wage. The idea is that the fraction of various types of occupation switch should remain stable during normal

Figure 1.3 :Workforce Occupational Mobility Rate



a), The top panel plots the quarterly workforce occupational mobility rate. The shaded areas are the NBER recession periods.

b), The middle panel plots the percentage deviations of both workforce occupational mobility rate and labor productivity from their HP trend. The black solid line represents the workforce occupational mobility rate and the red dash line represents the labor productivity. A smoothing parameter of 1600 is used for HP filter. The productivity is the real output per hour of all person for non-farm business section published by BLS.

c), The bottom panel plot shows the cross-correlation of workforce occupational mobility rate with labor productivity in terms of their deviation from HP trends.

time, some people leave for a better paid occupation due to promotion, some leave for a lower paid occupation due to low productivity and some other people leave due to personal reasons like retirement. Therefore the average effect of an occupation switch on hourly wage should be stable too at aggregate level during normal time. Without being able to identify the cause of occupation separations at individual level, the drop of average effect of an occupation switch on hourly wage may indicate a greater fraction of the occupation switches are due to lower productivity and triggered by deteriorated economic conditions.

The CPS asks the outgoing respondents (month in survey of 4 and 8) about their weekly earnings as well as the number of working hours each week. The same person is asked twice with exactly one year apart. The subsample used are respondents who: a) age 25 and above at the time of 4th survey, b) take the 4th survey between January 2005 and June 2009, c) report positive weekly income and working hours in both 4th and 8th surveys.

I use two measures of economic conditions change faced by individual  $i$  between her 4th and 8th survey. The first one is a recession dummy based on NBER recession dates, the second one is the unemployment rate at state level in which individual  $i$  resided. The regressions are:

$$\Delta \log(W_i) = \alpha + \beta X_i + \sum_t \gamma_t I_t + \eta D_i^{Occ} + \theta D_i^{Occ} \cdot R_i + \phi D^{Ind} + \chi D^{Ind} \cdot R_i + \epsilon_i \quad (1.4)$$

$$\Delta \log(W_i) = \alpha + \beta X_i + \sum_t \gamma_t I_t + \eta D_i^{Occ} + \lambda D_i^{Occ} \cdot \Delta U_i^s + \phi D^{Ind} + \pi D^{Ind} \cdot \Delta U_i^s + \epsilon_i \quad (1.5)$$

where  $\Delta \log(W_i)$  is the percentage change of hourly wage for individual  $i$  between 4th and 8th survey, and hourly wage is defined as weekly earning divided by weekly working hours.  $\alpha$  is a constant,  $X_i$  is the set of demographic variables: age (a polynomial of 4th order), gender, race (white, black, Asian and others) and highest education attainment at both survey (no high school diploma, high school diploma, associated degree, bachelor's degree and above).  $\{I_t\}$  is the monthly dummy except for the first month, and its value equals one if individual  $i$  took the 4th survey in month  $t$ . The monthly dummy captures the change of hourly wage caused by aggregate economic conditions change like inflation, recession and seasonality. Once the monthly dummies are included, it is no longer necessary to add other monthly macroeconomic variables since the monthly dummies span all other aggregate monthly variables.  $D_i^{Occ}$  ( $D^{Ind}$ ) is a dummy variable that equals one if respondent  $i$  changed occupation (industry) between her 4th and 8th survey and zero otherwise.

The first regression (equation 1.4) uses the NBER recession as the measure of economic conditions, where  $R_i$  is a dummy variable which equals one if the 4th interview is taken between December 2007 and June 2009. The parameter of interest is  $\theta$ , which measures the average additional (or the change of) effect of an occupation switch during the last recession. The second regression (equation 1.5) uses the 12-month difference in state-level unemployment rate between 4th and 8th interviews to measure the change of economic conditions faced by individual  $i$  at state level. A positive change  $\Delta U_i^s$  indicates deteriorated economic condition. The parameter of interest is the  $\lambda$ , which measures the average additional effect of an occupation switch for each additional one percentage of unemployment rate. Both specifications include the industry switch and its additional effects during recession or higher unemployment as additional controls.

Table 1.1 summarizes the results. Column I and II correspond to equation 1.4 with and without controlling for industry switch, column III and IV correspond to equation 1.5 with and without controlling for industry switch. Coefficient estimates for age, education and monthly dummies are omitted from the table. In all four specifications, the coefficient estimates of monthly dummies become significantly negative since late 2007, which implies continuously reduction of hourly wage in the absence of occupation switch since the recession. Using the NBER recession as the economic condition indicator, column I and II show the effect of an occupation switch is associated with 0.18% or 0.3% increase in hourly wage on average without or with controlling for industry switch respectively, although the point estimates are not statistically significant in both case. More importantly, an occupation switch during last recession caused additional 1.61% or 1.43% loss of hourly wage on average respectively in each specification, and they are both statistically significant. Column III and IV use the change of state-level unemployment rate as the measure of economic condition change, and each additional 1% unemployment rate leads to 0.45% or 0.51% loss of hourly wage on average upon occupation switch, depending on whether controlling for industry switches or not. The average 12-month change of unemployment rate at state level was 2.23% during the recession, which means that 1% or 1.14% loss of hourly wage is associated with increased unemployment rate. All four regressions indicate that deteriorated economic conditions significantly altered the average effect of occupation switch, which suggests the additional occupation switches during recession may be triggered by worsened economic conditions.

Table 1.1 : Regression Results

Coefficients \ Specification	I	II	III	IV
$\eta \times 100 (D_i^{Occ})$	0.18	0.30	0.42	0.28
$\theta \times 100 (D_i^{Occ} \cdot R_i)$	<b>-1.61 **</b>	<b>-1.43 *</b>	.	.
$\lambda \times 100 (D_i^{Occ} \cdot \Delta U_i^s)$	.	.	<b>-0.45 **</b>	<b>-0.51 **</b>
$\phi \times 100 (D^{Ind})$	.	-0.48	.	-0.91 **
$\chi \times 100 (D^{Ind} \cdot R_i)$	.	-0.75	.	.
$\pi \times 100 (D_i^{Ind} \cdot \Delta_i^s)$	.	.	.	0.19

The table reports the estimate of  $\eta, \theta, \lambda, \phi, \chi, \pi$  and their statistical significance. Respondents are restricted to a) age 25 and above b) report positive weekly earnings and working hours in both 4th and 8th surveys c) the 4th survey is between 2005 January and 2009 June, d) have valid hourly wage for both dates. Observations are weighted by the final weight assigned by CPS and sum of weights from each month is the same. Estimated coefficients of control variables are omitted in this table. Specification I and II correspond to equation 1.4 without and with industry switch controls. Specification III and IV correspond to equation 1.5 without and with industry switch controls. The coefficient estimates for other controls (age, gender, education, time dummies) are omitted from this table. The statistical significance is denoted by asterisks: \* = 5%, \*\* = 1%. Coefficients of interested are emphasized in bold font.

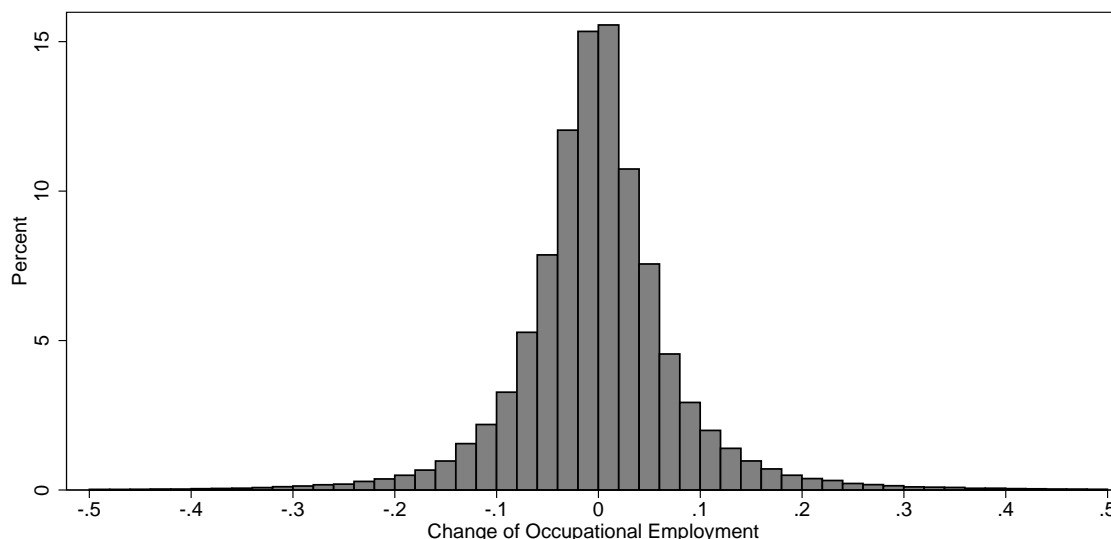
### 1.2.3 Employment at Occupation Level

In this section I explore the change of employment at occupation level. There were 500 occupational categories from 1996 to 1999 and over 1000 categories afterwards, and the employment varied greatly between different occupations. For instance, there were occupations that employed more than 2% of the workforce during most of the time in the sample, like sales supervisor & proprietors, sales cashiers, truck drivers and secretaries. On the other hand, a majority of these occupations hired only a tiny fraction of the workforce. For example, home economics teachers was associated with the lowest employment between 1996 and 2010, where less than 0.002% of the workforce reported working in this occupation.

The occupation employment also varied greatly across time. I plot the distribution of monthly change of occupational employment between 1996 and 2013 in figure 1.4. The monthly changes are adjusted within each month such that the weighted average is zero, and this eliminates the monthly seasonality and aggregate shifts. The range of monthly change is from -99%(almost diasperance) to 5312%, where the outliers are mostly due to the small employment at occupation

level. Changes of more than 50% are dropped, which only represents less than 0.3% of the sample. The standard deviation is 8.1%, and the shape of the distribution is quite symmetric, although the tails are much heavier than a normal distribution.

**Figure 1.4 : Occupational Employment Change**



The figure plots the distribution of monthly changes of employment at occupational level between 1996 and 2013. The occupational employment is defined as the fraction of workforce employed in a particular occupation. The change of occupational employment is adjusted within each month such that the weighted average is zero for each month, and this eliminates the monthly seasonality and aggregate change. Changes below -0.5 (less than 0.1% of the sample) and above 0.5 (0.2% of the sample) are not included in the plot. Data from 2002 December and 2010 December are dropped due to change of occupational codes. In total there are 214 months used to construct this distribution.

### 1.3 The Economy

In this part I calibrate a model with three types of productivity shocks and increasing productivity in occupational tenure. The purpose is to demonstrate that TFP shocks can cause time-varying endogenous occupational separation, which shifts the distribution of workforce occupational tenure over time. The workforce composition can also affect how the economy recovers from recession. The model is similar to the one in Nagypal (2007). The economy features a representative household who consumes and provides labor. On the production side, there is a continuum of occupations.

Production technology is partially segregated by occupation in the sense that occupation switch resets worker's occupational tenure to zero and productivity is increasing in tenure with other things being equal. Each occupation is modeled as a firm which produces with workers of different tenures, fires less productive workers and hire new workers. Time is discrete, each period corresponds to one months.

### 1.3.1 Household

The representative household lives infinitely, provides a continuum of workers and holds the share of the representative firm. The number of household members is normalized to 1. Momentary income includes labor income and dividend. Household derives utility from consumption only and solves the following optimization problem:

$$\max_{\{C_t, Z_{t+1}, B_{t+1}\}_t^\infty} J_t = E_t \sum_t \beta^t C_t \quad (1.6)$$

subject to:

$$C_t + Q_t Z_{t+1} = (Q_t + D_t) Z_t + W N_t + b U_t - T_t \quad (1.7)$$

where  $C_t$  is consumption,  $W$  is the fixed wage for all employed members,  $b$  is the unemployment benefit,  $Q_t$  is the ex-dividend share price of all firms,  $D_t$  is the dividend payment and  $Z_t$  is the equity position or period  $t$ . To make the model trackable  $W = b$  is assumed, members are indifferent between being employed or not. The government balances its budget such that  $T_t = U_t b$ .

The discount factor derived from this preference is simply  $\beta$ . As long as the firms uses the same discounting factor, the household is indifferent between different share positions. Of course in equilibrium the household chooses to hold all the shares to clear the financial market.

### 1.3.2 Representative firm in occupation $o$

There is a continuum of occupations indexed by  $o \in [0, 1]$ . Each occupation is represented by a firm of the same index. At the beginning of period  $t$ , firm  $o$  consists of  $N_{ot}$  workers who are indexed by

$i$ , each with  $\tau_i$  tenure. The output of worker  $i$  is:

$$y_i = A_t \cdot X_{ot} \cdot Z_{it} \cdot H(\tau_i) \quad (1.8)$$

where:

- $A_t$  is the aggregate TFP which follows a finite state first-order Markov process. The state space is  $A = \{A^1, A^2, \dots, A^M\}$  where  $A^1 < A^2 < \dots < A^M$ . The transition matrix is  $P^A$  with  $P_{ij}^A$  denoting the probability of transiting from  $A^i$  to  $A^j$ . Aggregate TFP affects the productivity of all workers in the economy, and is the only driving force for the long-term aggregate employment, output, productivity etc.
- $X_{ot}$  is the occupation-level productivity which follows a finite state first-order Markov process. The state space is  $X = \{X^1, X^2, \dots, X^K\}$  where  $X^1 < X^2 < \dots < X^K$ . The transition matrix is  $P_X$  with  $P_{ij}^X$  denoting the probability of transiting from  $X^i$  to  $X^j$ . With the assumption of continuum of occupations, the transition between different occupational productivity categories for all occupations does not create any aggregate uncertainty. This process does not capture the difference in productivities between occupations in the data, as some occupations are always more productive (or paid at a higher rate) than others. Instead, in the model all occupations are normalized to have the same long-run average productivity as well as same wage payment, and this variation in occupational productivity captures the time-varying demand for the same occupation, which drives the occupational employment and worker flows between occupations even in the absence of aggregate TFP shocks as in figure 1.4.
- $Z_{it}$  is the match quality for worker  $i$ . It follows a finite state first-order Markov process with state space  $Z = \{Z^1, Z^2, \dots, Z^S\}$  where  $Z^1 < Z^2 < \dots < Z^S$ . The transition matrix is  $P_Z$  with  $P_{ij}^Z$  denoting the probability of transiting from  $Z^i$  to  $Z^j$ . Different specifications can yield different hiring and separation dynamics. For instance, if  $Z$  is a singleton, then firms who fire incumbent workers do not post vacancies at all because all incumbent workers will always be more productive than a newly hired worker in the next period. With  $S > 1$  and appropriate parameterization, firms would hire new workers and fire incumbent workers at the same time because of low match quality for incumbent workers and relatively higher expected match quality for a new worker. Combining this with the occupation-level productivity fluctuations,



the model is able to replicate the distribution of occupational employment monthly change.

- $H(\tau_i)$  denotes the productivity gain because of occupational experience. Tenure is the number of months that worker  $i$  has worked in occupation  $o$ , including the current period.  $H(\cdot)$  is a strictly increasing, concave function starting at 1. A newly hired worker has a tenure of one. Tenure is capped at  $\bar{\tau}$  periods, therefore worker is separated from her current occupation with certainty after  $\bar{\tau}$  periods.

### 1.3.3 Search and Matching

Since the labor market is segregated by occupations and each is represented by a firm, I adapt the standard Diamond-Mortensen-Pissarides search friction mechanism to model the inflow of new workers to workforce. Firm  $o$  posts job vacancies,  $V_{ot}$ , to attract unemployed workers for a total cost of  $\kappa_0 V_{ot} + \kappa_1 V_{ot}^2$ . The convexity of cost ensures that not only the most productive occupations post vacancies.  $q_t$  is the probability for a vacancy to be filled at the end of period  $t$ . Unemployed members are ex-ante identical and cannot choose the occupation, therefore each firm faces the same probability  $q_t$ . Let  $V_t = \int_0^1 V_{ot} do$  denote the total number of vacancies posted. The total number of workers hired,  $G(U_t, V_t)$ , is specified as:

$$G(U_t, V_t) = \frac{U_t V_t}{(U_t^\iota + V_t^\iota)^{1/\iota}}, \quad (1.9)$$

in which  $\iota > 0$  is a constant parameter.

Define  $\theta_t \equiv V_t/U_t$  as the vacancy-unemployment ( $V/U$ ) ratio. The probability for an unemployed worker to get hired at the end of current period,  $f(\theta_t)$ , is:

$$f(\theta_t) \equiv \frac{G(U_t, V_t)}{U_t} = \frac{1}{(1 + \theta_t^{-\iota})^{1/\iota}}, \quad (1.10)$$

and the probability for a vacancy to be filled at the beginning of next period,  $q(\theta_t)$ , is:

$$q(\theta_t) \equiv \frac{G(U_t, V_t)}{V_t} = \frac{1}{(1 + \theta_t^\iota)^{1/\iota}}. \quad (1.11)$$

It follows that  $f(\theta_t) = \theta_t q(\theta_t)$  and  $\partial q(\theta_t)/\partial \theta_t < 0$ .  $\theta_t$  is a measure of labor market tightness from

the perspective of the firm. An increase in  $\theta_t$  leads to lower likelihood to fill a vacancy.

### 1.3.4 Separation and Hiring Decisions

After production, all matches face two forms of separation. Firstly, a fixed fraction  $s$  of workers leaves their occupation exogenously due to promotion, retirement or other reasons that are not related to poor performance and they will not return to the same occupation in the future. This also ensures that a decreasingly fraction of the workforce reaches next tenure level. After that, the firm chooses to fire some less productive workers. All workers are paid at the same wage rate of  $W$  and unemployed members are paid at the same rate  $b = W$ , therefore the surplus of turning an unemployed member into an employed member is zero for the household. The hiring and firing decisions depend solely on the surplus of the match for the firm. Similar to the specification in Nagypal (2007), this economy with constant wage is equivalent to an economy with a positive worker share of the surplus in the sense that both economies have the same optimal policies and distribution of workers across states. The proof is in Appendix section 1.B. Let  $S^\tau(A^m, X^k, Z^s)$  denote the continuation value for a match with tenure  $\tau$  after production and exogenous separation. It is defined recursively as (time script is omitted):

$$\begin{aligned}
S^{\bar{\tau}}(A^m, X^k, Z^s) &= 0 \\
S^\tau(A^m, X^k, Z^s) &= \beta \sum_{n=1}^M P_{m,n}^A \sum_{j=1}^K P_{k,j}^X \sum_{r=1}^S P_{s,r}^Z \{A^n X^j Z^r H(\tau+1) - W + (1-s) \max(0, S^{\tau+1}(A^n, X^j, Z^r))\} \\
S^0(A^m, X^k) &= \beta \sum_{n=1}^M P_{m,n}^A \sum_{j=1}^K P_{k,j}^X \sum_{r=1}^S \tilde{\pi}_r^Z \{A^n X^j Z^r H(1) - W + (1-s) \max(0, S^1(A^n, X^j, Z^r))\}
\end{aligned}$$

In the terminal period the match is separated with certainty, therefore the continuation value is zero. For all tenure  $\tau = 1, 2, \dots, \bar{\tau} - 1$ , the continuation value is the discounted expected value of the match going forwards. The value of the match in the next period is the sum of output net of wage payment and the higher value of separation or continuation taking into account the exogenous separation probability  $s$ . Separating in the next period yields a value of zero and continuation has a value of  $S^{\tau+1}(A^n, X^j, Z^r)$ . The expectation is taken over all possible realizations of  $A_{t+1}, X_{t+1}, Z_{t+1}$  weighted by their corresponding probabilities. New workers start with a random match quality according to the stationary distribution over  $Z$  implied by  $P^Z$  such that  $\tilde{\pi}^Z = \tilde{\pi}^Z P^Z$ .  $S(\cdot)$  can be

solved by backward iteration.

A firm keeps an existing match if and only if the continuation value is positive. In other words, existing match is separately endogenous if the present value is less than or equal to zero. Once a worker is separated her tenure is reset to zero, she becomes the same as any other unemployed member in the economy and available for hiring in the next period.

All new worker start with a random  $Z^r$  their first period,  $S^0(A^m, X^k)$  represents the present value of a new worker who starts to work in the next period given  $A_t = A^m$  and  $X_t = X^k$ . Firm  $o$  posts  $V_{ot}$  vacancies until  $\kappa_0 + 2\kappa_1 V_{ot} \geq q(\theta_t)S^0(A^m, X^k)$ , taking as given the hiring probability  $q(\theta_t)$ .  $\kappa_0 + 2\kappa_1 V_{ot}$  is the marginal cost of an additional post, and  $q(\theta_t)S^0(A^m, X^k)$  is the marginal benefit. If  $\kappa_0 \geq q(\theta_t)S^0(A^m, X^k)$ , i.e. the marginal cost of the first vacancy over-weighs the marginal benefit, the firm would optimally choose  $V_{ot} = 0$ .

### 1.3.5 The Evolution of Economy

Any two matches with the same  $\{X_t, Z_t, \tau\}$  have the same future probability space, regardless of the occupations they are associated with. This is apparent since the surplus function  $S(\cdot)$  does not depend on the occupation  $o$ . Let  $\Lambda_t$  be the distribution of all agents over  $\{X, Z, \tau\}$  at the beginning of period  $t$  with  $\lambda_t^{X, Z, \tau}$  denoting the mass of workers with the same  $\{X, Z, \tau\}$ .<sup>2</sup> Given that the number of exogenous states is finite and the number of tenure is bounded above,  $\Lambda_t$  can be described with a matrix of finite size. The economy is sufficiently summarized by  $\{A_t, \Lambda_t\}$ . After production, workers who reach  $\bar{\tau}$  are separated with certainty, then a fixed fraction  $s$  of all kinds of workers are separated exogenously, lastly workers with  $S^\tau(A_t, X_t, Z_t) \leq 0$  are fired endogenously.

The distribution  $\Lambda_t$  evolves as follows:

- For  $\tau = 2, 3, \dots, \bar{\tau}$ , the  $\lambda_{t+1}^{X^j, Z^r, \tau}$  is:

$$\lambda_{t+1}^{X^j, Z^r, \tau} = \sum_{k=1}^K P_{k,j}^X \sum_{s=1}^S P_{s,r}^Z (1-s) \lambda_t^{X^k, Z^s, \tau-1} \cdot I(S^\tau(A_t, X^k, Z^s) > 0) \quad (1.12)$$

The number of workers with  $X^j, Z^r, \tau$  at time  $t+1$  is the sum of all workers with  $X^k, Z^s, \tau-1$

---

<sup>2</sup>For convenience, unemployed members have  $\tau = 0$

at period  $t$  who are not separated either exogenously or endogenously, weighted by their probabilities of moving from  $\{X^k, Z^s\}$  to  $\{X^j, Z^r\}$  in one period.

- For  $\tau = 1$ , the number of new workers depends on the how many matches are created at the end of period  $t$ . Let  $\tilde{\pi}^X$  ( $O \times 1$  row vector) denote the stationary distribution of occupations over  $X$  such that  $\tilde{\pi}^X = \tilde{\pi}^X P^X$ , where  $\tilde{\pi}_k^X$  is the fraction of the occupations with  $X^k$ . Occupations with the same  $X^k$  choose to post the same number of vacancies. Let  $V_t^k$  denote the number of vacancies posted by a firm with  $X^k$ , then the total number of posting is  $V_t = \int_o^1 V_{ot} do = \sum_{k=1}^K \tilde{\pi}_k^X V_t^k$ . All new workers start with a random  $Z^r$  according to  $\tilde{\pi}_r^Z$ . The number of new workers is determined as:

$$\lambda_{t+1}^{X^j, Z^r, 1} = \tilde{\pi}_r^Z \sum_{k=1}^K P_{k,j}^X \tilde{\pi}_k^X V_t^k q(\theta_t) \quad (1.13)$$

where  $\tilde{\pi}_k^X V_t^k q(\theta_t)$  is the number of workers hired by all occupations with current  $X^k$ . The number of new workers with  $X^j, Z^r$  at period  $t+1$  is the sum of workers created at the end of last period  $\tilde{\pi}^X(k) V_t^k q(\theta_t)$  condition on  $X^k$ , weighted by the probability of moving from  $X^k$  to  $X^j$  and multiplied by the likelihood of the match quality  $\tilde{\pi}_r^Z$ .

### 1.3.6 Aggregation

The current employment is  $N_t = \sum_X \sum_Z \sum_{\tau \geq 1} \lambda_t^{X, Z, \tau}$ , i.e. the sum of masses of all workers.

The aggregate output  $Y_t = \sum_{k=1}^K \sum_s^S \sum_{\tau=1}^{\bar{\tau}} \lambda_t^{X^k, Z^s, \tau} A_t X^k Z^s H(\tau)$ . The average labor productivity is  $Y_t/N_t$ .

The total number of separation is:

$$s_t^{total} = \sum_{k=1}^K \sum_{s=1}^S \sum_{\tau=1}^{\bar{\tau}-1} \lambda_t^{X^k, Z^s, \tau} \cdot \left( s + (1-s) I(S^\tau(A_t, X^k, Z^s) \leq 0) \right) + \sum_{k=1}^K \sum_{s=1}^S \lambda_t^{X^k, Z^s, \bar{\tau}} \quad (1.14)$$

The total number of exogenous separation is simply  $sN_t$ . The number of endogenous separations is  $s_t^{total} - sN_t$ .

### 1.3.7 Competitive equilibrium

The economy is summarized by  $\{A_t, \Lambda_t\}$ .  $\Lambda_t$  evolves according to the aforementioned rules. Each occupations with  $X^k$  posts  $V_t^k$  vacancies taking as given  $\theta_t$ . At equilibrium  $\theta_t = \frac{\sum_{k=1}^K \tilde{\pi}^X(k) V_t^k}{1 - N_t}$ . The labor market is the only market that needs clearing each period at equilibrium. The financial market clears automatically since the firm use the same discounting factor  $\beta$ .

## 1.4 Model Calibration

The model is calibrated to mimic some key moments from the macro data as well as some findings in section 2. Each period corresponds to one month. The calibration is summarized in the following:

- The discount factor  $\beta = 0.96^{0.083}$  is chosen such that the annual risk-free interest rate is 4%.
- The aggregate TFP process  $A_t$  is modeled as the discrete-state version an AR(1) process with 101 states, persistence of 0.97 and innovation standard deviation of 0.004. The discretization method is adapted from Rouwenhorst (1995).
- The occupational level productivity is also modeled as the discrete-state version of an AR(1) process with 9 states, persistence of 0.97 and innovation standard deviation of 0.004. These are the same parameter values used for aggregate TFP.
- The match quality  $Z_t$  is assumed to take a two-state first-order Markov process with state space  $Z = [0.5, 1]$  and high persistence in each state. The transition matrix  $P^Z$  is:

$$\begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

- The productivity gain in occupational tenure  $H(\tau)$  is assumed to have a simple functional form  $H(\tau) = a(1 - \frac{b}{c+\tau})^d$ . Kambourov and Mankovskii (2009) finds that the wage increases by about 5.4% after 2 years, about 12% after 5 years and 17% after 8 years, and those increases represent 38% of the productivity gain. As the result  $H(\tau) = 2.2657 \cdot (1 - \frac{35.42}{\tau+136.01})^{2.7344}$ . This concave function starts at 1 and grows gradually, where  $H(1) = 1, H(25) = 1.142, H(61) = 1.316, H(97) = 1.447$ . The plot is in the appendix.

- The parameters are summarized in the table 1.2. The stochastic processes  $X$ ,  $Z$  and the matching parameters  $s, \kappa_0, \kappa_1, \iota$  are jointly chosen such that at steady state (no aggregate TFP shocks), the distribution of occupational employment change mimics that from CPS in figure 1.4.

**Table 1.2 : Model Parameters**

Parameter	Expression	Value
Discount factor	$\beta$	0.96 <sup>0.083</sup>
Wage, Unemployment Benefit	$W, b$	0.95
Matching parameter	$\iota$	1.25
Fixed Cost of Posting A Vacancy	$\kappa_0$	0.5
Quadratic Cost of Posting A Vacancy	$\kappa_1$	1

## 1.5 Model Results

The model moments are generated by simulation. Table 1.4 reports the main moments of the model. The data moments are based on published data from BLS and BEA except the workforce occupational mobility rate  $s$ , which is computed in section 2 based on the assumption of constant 50% occupational mobility rate for people who leave the workforce. In the appendix I show the data moments under different assumptions of occupational mobility rate for people who leave the workforce. Data moments are based on data from 1948Q1 to 2013Q4, except for those involving  $s$ , which are from 1996Q1 to 2013Q4. The relatively short time series of  $s$  renders  $cor(s, U)$  and  $cor(s, Y)$  not statistically significant from zero. The quarterly series is the three-month average of monthly series generated by the model. The key variables included are the unemployment rate  $U$ , output  $Y$ , labor productivity  $X$  and workforce occupational mobility rate  $s$ . Given the linear fashion of the model, certain model moments are more correlated than they appear in the data. For instance,  $U$  and  $s$  are almost perfectly correlated in the model, and it is because the unemployment is directly linked to the occupation separation.

A section of the simulation is plotted in Figure 1.5 for a total of 9 variables to demonstrate the comovements of some key moments with TFP. Subfigure (a) plots the time series of aggregate TFP, which drives all other aggregate variables and serves as the business cycle indicator. TFP

**Table 1.3 : Model Moments and Data Moments**

	Data					Model			
	$U$	$Y$	$X$	$s$		$U$	$Y$	$X$	$s$
Std.	0.14	0.02	0.01	0.04		0.30	0.03	0.01	0.13
Autocorr.	0.89	0.85	0.71	0.55		0.97	0.95	0.89	0.97
Corr.		-0.87	-0.01*	-0.20*	$U$		-0.97	-0.82	0.96
			0.40	0.06*	$Y$			0.92	-0.92
				-0.38	$X$				-0.83

The Data moments are based on data published by BLS and BEA. The seasonally adjusted quarterly unemployment ( $U$ ) is from 1948Q1 to 2013Q4 published by BLS. The per capita real output ( $Y$ ) is calculated as the real GDP divided by the population. The real GDP is from 1948Q1 to 2013Q4 and published by BEA and the population is the civilian non-institutional population of age 16 and older published by BLS. The productivity ( $X$ ) is the seasonally adjusted real output per hour of all person in the non-farm business sector from 1948Q1 to 2013Q4 published by BLS. The workforce occupational mobility rate ( $s$ ) is based on the workforce turnover rate and EE occupational mobility rate estimated in section 2 and available from 1996Q1 to 2013Q4. The standard deviation and first-order auto correlation is computed using the data from 1948Q1 to 2013Q4 except for  $s$ , which is from 1996Q1 to 2013Q4. The asterisk \* in correlation matrix denotes NOT statistically significant from zero at 5% level.

The model moments are generated by simulation. The model generates monthly time series of 9000 periods, and the quarterly series are computed as the three-month averages of the corresponding monthly series.

reverts to one in the long run. The theoretical unconditional standard deviation is 0.0165, hence TFP below 0.967 can be considered as a severe recession. The unemployment rate in (b) is clearly countercyclical and mimics the path of TFP in a reversed fashion. The total output in (c) is procyclical and also seems to mimic the path of TFP. The labor productivity in (d), which is the ratio of total output to employment, doesn't seem to follow TFP closely and is only weakly procyclical. The endogenous occupational separation rate in (e), which is the fraction of workforce that leave their current occupation because of non-positive continuation value, closely mimics the unemployment rate and is countercyclical. The workforce average tenure in (f) is primarily driven by the endogenous occupational separation rate, where higher separation rate drives down the average tenure eventually. The average tenure is also procyclical and slightly lags the TFP. The labor productivity excluding TFP, which is defined as the labor productivity in subfigure (d) divided by TFP. It measures the labor productivity contributed by occupational productivity  $x_i$ , match quality  $z_i$  and tenure  $\tau_i$ , and it is countercyclical. The employments by occupational productivity are plotted in (h). The  $k$ th line from the bottom represents the average number of workers employed

by an occupation with  $X^k$ . Note that  $X^1$  is the most contracting occupation and  $X^{15}$  is the most expanding occupation. Highly productive occupations hire more workers than less productive occupations. The employment by an occupation with low occupational productivity is procyclical, whereas the employment by an occupation with high occupational productivity is countercyclical. The employment by match quality are plotted in (i), the higher one corresponds to good quality and the lower one corresponds to bad quality. The number of good match workers are countercyclical and the number of bad-match workers are procyclical.

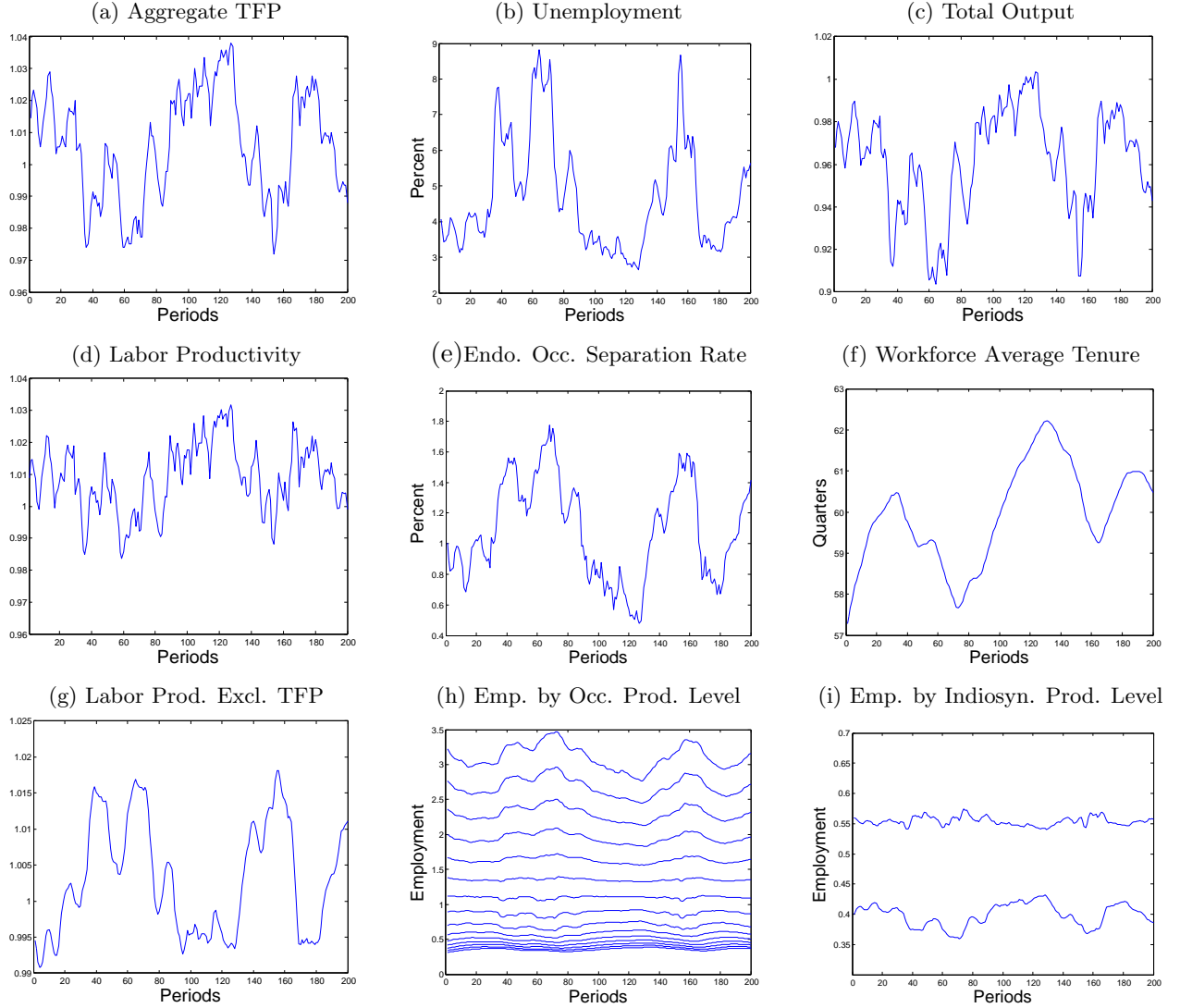
### 1.5.1 Impulse Response Analysis

In this part I show how the model economy response to a TFP shock by simulation. A negative shock hits the TFP once, and it drops immediately upon impact and then returns to its long-run median after a quarter. The size of the shock equals one conditional standard deviation (0.004). The economy responses to the negative TFP shocks of different magnitude the same way, although the magnitude of responses are different. The economy also responses to shocks symmetrically.

Figure 1.6 plots the impulse responses of some key variables of the economy. The same variables in Figure 1.5 are included. All figures are in terms of percentage deviations from their steady states. Subfigure (a) plots the time series of aggregate TFP which starts at steady state in the first quarter and is hit by a negative shock in the second quarter. The TFP reverts to steady states in the subsequent quarter. Subfigure (b) plots the unemployment rates. It spikes as soon as TFP dips, and recovers quickly once TFP returns to its steady state. However, the it does not return to its steady state until more than 10 quarters afterwards. This is due to a greater number of new workers who are more likely to leave their occupation quickly due to low tenure productivity. Subfigure (c) plots the total output. The initial impact is comparable to TFP, but it recoveries faster than the labor market. This is consistent with the data that the output recoveries faster than the labor market after a recession. Subfigure (d) plots the labor productivity defined as the ratio of output to employment. The labor productivity overshoot the steady state after TFP returns because of workforce composition change. More workers with bad match quality leave their occupation than workers with good match quality. When new workers are hired they have equal likelihood of good or bad match quality. Subfigure (e) plots the endogenous occupational separation



**Figure 1.5** A Simulated Sample



The model is simulated for 1600 months, and the first 1000 months are dropped to simulate a random starting point. The rest 600 monthly time series are converted to 200 quarterly time series.

Subfigure (a) plots the simulated quarterly aggregate TFP time series.

Subfigure (b) plots the quarterly unemployment rate in percentage.

Subfigure (c) plots the quarterly total output.

Subfigure (d) plots the quarterly labor productivity defined as the ratio of total output to employment.

Subfigure (e) plots the endogenous occupational separation rate.

Subfigure (f) plots the workforce average tenure in quarters.

Subfigure (g) plots the labor productivity excluding TFP, defined as the labor productivity divided by TFP.

Subfigure (h) plots the employment by occupational productivity level. The 15 lines, from top to bottom, correspond to occupational productivity from high to low. The  $k$ th line from the bottom represents the average number of workers employed by an occupation with current occupational productivity  $X^k$ . The aggregate employment would be the weighted average of these employments at different occupational productivity level, weighted by stationary fraction of occupations in each category.

Subfigure (i) plots the employment by idiosyncratic productivity level. The top line corresponds to the higher idiosyncratic productivity and the bottom line corresponds to lower idiosyncratic productivity. Those two employments sum up to the aggregate employment.

rate. The rate spikes immediately when TFP drops because the continuation value of all existing matches decreases at the same time, and eventually stabilises after TFP returns to its steady state. Subfigure (f) plots the workforce average tenure. Higher endogenous separation rate drives down the average tenure, which recovers very slowly after the negative TFP shock. Subfigure (g) plots the labor productivity excluding the TFP, which is defined as the labor productivity in subfigure (d) divided by TFP. It measures the labor productivity attributed to occupational productivity, match quality and tenure, and a higher value means a bigger fraction of the workforce is concentrated in expanding occupations and more workers have good match quality. Since the value of a match is monotonic in occupational productivity, match quality and tenure. Matches with contracting occupations, bad match quality and short tenures are more likely to be destroyed when TFP drops, and remaining matches are more likely to concentrate in expanding occupations, have good match quality or longer tenure, all of which lead to higher labor productivity excluding TFP. This shift in worker distribution over occupations, match qualities and tenure is the reason that output reverts back to steady state level much faster than the unemployment rate. Subfigure (h) plots the employment by occupational productivity. Expanding occupations increase their employment due to less tight labor market, whereas contracting occupations decrease their employment due to lower continuation values. TFP affects occupational employment through three channels. The first one is through endogenous separation which is negatively correlated with TFP, the second channel is through hiring where the value of a new match is increasing in TFP, the third channel is through the labor market where higher TFP leads to tighter labor market, which decreases the value of a vacancy. The first two channels reduce the employment and the third channel increases the employment after a negative shock to TFP. Employment associated with expanding occupations is countercyclical because the effect of less tight labor markets dominates the other two channels, whereas the employment associated with contracting occupations is procyclical because the first two channels dominate. Subfigure (i) plots the employment by match quality where the top one corresponds to good quality and the lower one corresponds to bad quality. Both drop when TFP is hit, but the number of workers with good match increases quickly afterwards, and the number of workers with bad match recovers slowly. The reason that the one corresponds to good quality increases is primarily because the labor market becomes less tight with lower TFP, and matches with good quality are more likely to survive once created.

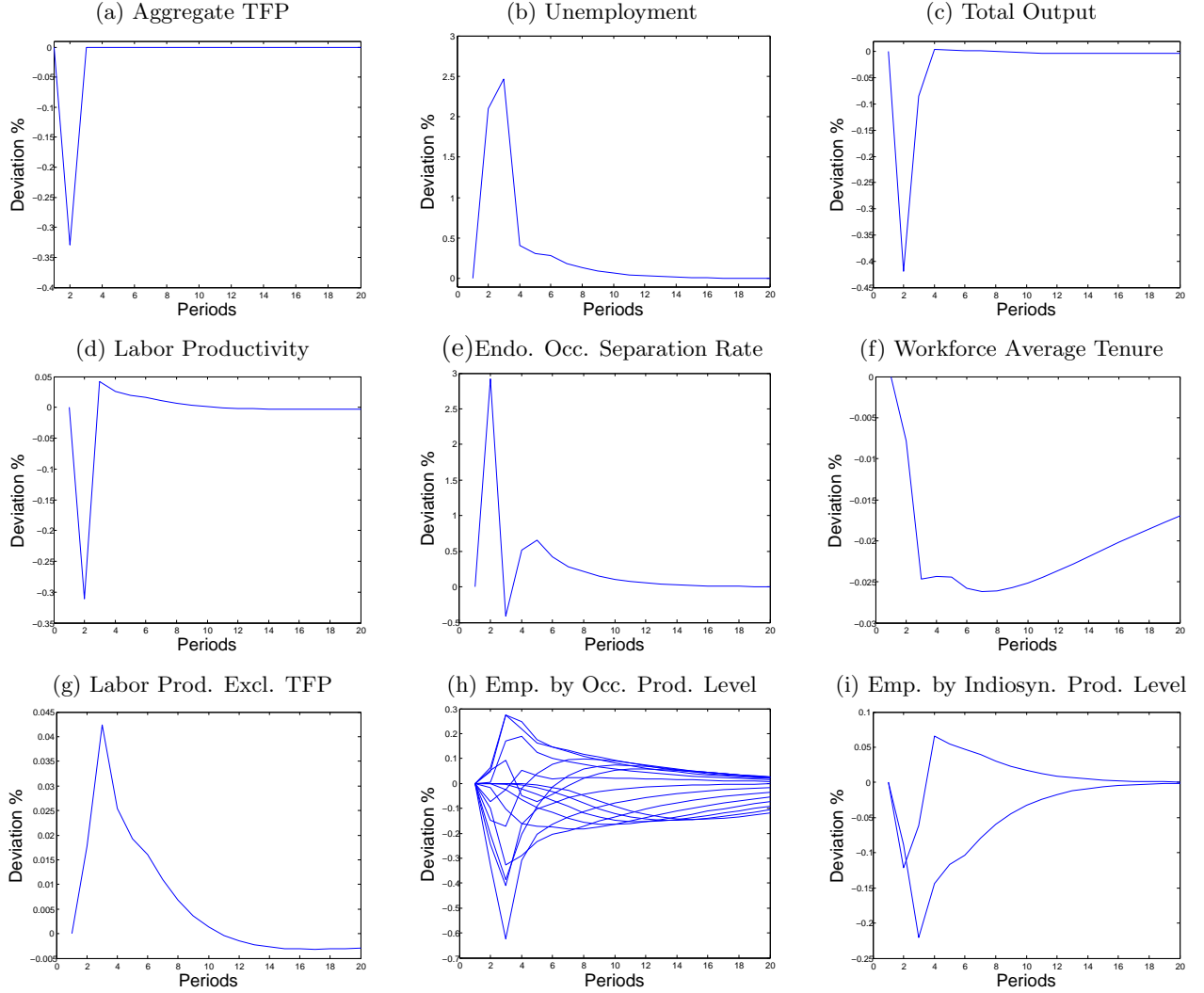
The impulse response shows by incorporating the occupational productivity, match quality

and occupational tenure, the model is able to generate the fast recovery of output and slow recovery of labor market after a recession. This is achieved by shifting the worker tenure distribution. Low TFP reduces the value of a match going forward, hence less productive workers, either because of occupational productivity, match quality or tenure, are more likely to be fired and replaced by new workers who are recruited in more productive occupation.

## 1.6 Conclusion

This paper argues that the countercyclical workforce occupational mobility could be caused by aggregate economic fluctuations, as a result, the workforce composition shifts over business cycle, which in turn can affect how the economy recovers from recession. Using the Current Population Survey, I find: a) workers were more likely to separate from their occupations during last two recessions, b) the deteriorated economic conditions during last recession significantly reduced the average productivity change upon occupation switch, which suggests the additional occupation switches were triggered by worsened economic conditions, c) the employment at occupation level fluctuates over time, and the distribution of monthly change looks like a bell shape with heavier tails than a normal distribution. To demonstrate how workforce composition could play a role in business cycle, I calibrate a model with aggregate TFP, occupational productivity, match quality and increasing productivity in occupational tenure. Workers can be separated endogenously from current occupation due to low continuation value for the firm. Aggregate TFP shifts the threshold of endogenous separation, and less productive workers are fired first in a recession, leaving workers who are in expanding occupations, have better match quality or longer occupational tenure in the workforce. The output recovers faster than the labor market because newly hired workers tend to concentrate in expanding occupations and have better match quality than before. The model is able to generate faster output recovery, slow labor market recovery after a recession and countercyclical endogenous occupation mobility rate. The result suggests that other than the total amount of labor input, the workforce composition change could also plays an important role in real business cycle and this channel shall be explored more in the future.

**Figure 1.6** Impulse Responses for a Negative TFP Shock



The model is simulated for 1060 periods. The first 1000 months are for the economy to reach its steady state and hence dropped. The rest 60 months are converted into 20 quarters. All graphs are in terms of percentage deviations from the steady state level.

Subfigure (a) plots the quarterly TFP. The negative TFP shock takes place at the fourth month, or the second quarter.

Subfigure (b) plots the response of quarterly unemployment rate.

Subfigure (c) plots the response of quarterly total output.

Subfigure (d) plots the response of quarterly labor productivity.

Subfigure (e) plots the response of endogenous occupational separation rate.

Subfigure (f) plots the response of workforce average tenure in quarters.

Subfigure (g) plots the response of labor productivity excluding TFP, defined as the labor productivity divided by TFP.

Subfigure (h) plots the response of employment by occupational productivity level. The 15 lines, from top to bottom, correspond to occupational productivity from high to low.

Subfigure (i) plots the response of employment by idiosyncratic productivity level. The top line corresponds to the higher idiosyncratic productivity and the bottom line corresponds to lower idiosyncratic productivity. Those two employments sum up to the aggregate employment.

## References

- Black, Lynch, 1996, Human-capital investments and productivity, *American Economic Review* 86, 263-267
- Burdett, Carrillo-Tudela, Coles, 2011, Human capital accumulation and labor market equilibrium, *International Economic Review* 52(3), 657-677
- Corrado, Hulten, 2010, Measuring intangible capital, how do yo measure a "technological revolution", *American Economic Review* 100, 99-104
- Corrado, Hulten, Sichel, 2005, Measuring capital and technology: an expanded framework, *NBER: Measuring capital in the new economy*, 11-46.
- Corrado, Hulten, Sichel, 2009, Intangible capital and US economic growth, *Review of Income and Wealth* 55(3), 661-685
- Danthine, Jin, 2007, Intangible capital, corporate valuation and asset pricing, *Economic Theory* 32, 157-177
- Fujita, Ramey, 2012, Exogenous versus endogenous separation, *American Economic Journal: Macroeconomics* 4(4), 68-93
- Kambourov, Manovskii, 2008, Rising occupational and industry mobility in the United States: 1968-97, *International Economic Review* 49(1), 41-79
- Kambourov, Manovskii, 2009, Occupational mobility and wage inequality, *Review of Economic Studies* 76, 731-759
- Kambourov, Manoovskii, 2009, Occupational specificity of human capital, *International Economic Review* 50(1), 63-115
- Kambourov, Manoovskii, 2013, A cautionary note on using march CPS and PSID data to study worker mobility, working paper.
- King, Rebelo, 1999, Resuscitation real business cycles, *Handbook of Macroeconomics* 1, 927 - 1007
- Madrian, Lefgren, 1999, A note on longitudinally matching current population survey (CPS) respondents, NBER working paper.

- Mankiw, Romer, Weil, 1992, A contribution to the empirics of economic growth, *The Quarterly Journal of Economics*, 407-437
- McCulla, Holdren, Smith, 2013, Improved estimates of the national income and product accounts - results of the 2013 comprehensive revision, *Survey of Current Business*, 14-45
- McGrattan, Prescott, 2010, Unmeasured investment and the puzzling U.S. boom in the 1990s, working paper.
- Mortensen, Pissarides, 1994, Job creation and job destruction in the theory of unemployment, *Review of Economic Studies* 61(3), 397-415
- Moscarini, Thomsson, 2007, Occupational and job mobility in the US, *Scandinavian Journal of Economics* 109(4), 807-836
- Moscarini, Vella, 2008, Occupational mobility and the business cycle, NBER working paper
- Nagypal, 2007, Learning by doing vs learning about match quality: can we tell them apart?, *The Review of Economic Studies* 74(2), 537-566
- Neal, 1995, Industry-specific human capital: evidence from displaced workers, *Journal of Labor Economics* 13, 653-677
- Parent, 2000, Industry-specific capital and wage profile: evidence from the National Longitudinal Survey of Youth and the Panel Study of Income Dynamics, *Journal of Labor Economics* 18(2), 306 -323
- Petrosky-Nadeau, Zhang, Kuehn, 2013, Endogenous disasters and asset prices, working paper.
- Shimer, 2005, The cyclical behavior of equilibrium unemployment and vacancies, *American Economic Review* 95(1), 25-49
- Sullivan, 2010, Empirical evidence on occupation and industry specific human capital, *Labour Economics* 17, 567-580
- Thompson, 2010, Learning by doing, *Handbooks in Economics* 1, 429 -476

# APPENDIX

## 1.A Occupation Mobility Estimation

### 1.A.1 Current Population Survey Brief Introduction

The Bureau of Labor Statistics conducts monthly survey of households called Current Population Survey (CPS), which is widely used for studies of the labor market. (more on the introduction of CPS..). About 50,000-60,000 households are selected for interview each month for a total of about 150,000 respondents, and household members are asked about their personal information, labor force status as well as details of their work. Each household stays in the survey for 4 consecutive months, and then leaves the survey for 8 months, and returns for another 4 months before permanently leaving the sample. This design ensures the continuity between consecutive months. Therefore in each monthly interview, about one eights of the households are in their first month in sample (MIS), one eights are in their second MIS, etc. When a new monthly survey is conducted, those households who had completed either their fourth or eighth survey in the previous month are replaced by: a) half of replacement are new to the survey and now in their first MIS, b) those households who finished their fourth survey 8 months ago are brought back. In theory, 75% of households are surveyed repeatedly between two consecutive months. This longitudinal aspect of CPS allows me to study the changes of occupation as well as labor force status by focusing on these repeatedly interviewed households.

### 1.A.2 Matching respondents from consecutive months

Since CPS was not designed to be a longitudinal dataset from the beginning, matching correspondents from consecutive months is not very straightforward. In any monthly survey, a correspondent is uniquely identified by the household identifier (HHID) and an individual line number (LINENO) within the household<sup>3</sup>. However, the same combination of HHID and LINENO

---

<sup>3</sup>The household identifier actually includes hrhhid (household id), hrsample (household sample id) and hrsersuf (household serial suffix). However, between 1994.4 and 1995.5, such combinations cannot uniquely identify a person, meaning that there are multiple observations with the same combination of aforementioned variables. Furthermore, 1995.5 cannot be matched with 1995.6. For now I only use data from 1996 to 2013 until I hear from Jean Roth, who

may not necessarily identify the same correspondent from one month to the next. This is because CPS samples are address based, meaning that if a household being interviewed moves to a new location, it will not be followed for interview at its new location, and the new household at the original location will be interviewed instead. The household number (HHNUM) variable is used to identify this situation: its value signifies the number of households being interviewed since the beginning at the same housing unit. A household has HHNUM of one when first entering the rotation, and HHNUM increases by one whenever a new household replaces the old household at the same housing units. In theory, the combination of HHID, HHNUM and LINENO should uniquely identify the same individual from one month to the next. (Add a table here to compare the key characteristics of total sample and matched sample, to show that the matched sample is still representative of the larger sample.) Madrian and Lefgren (1999) proposed a algorithm to match correspondents in the annual March survey. Since monthly CPS is used in this paper, a slightly revised version of their algorithm is used. Unfortunately, the BLS deliberately wanted to defeat linking between certain months in 1995 and such algorithm fails to identify any respondents.

The monthly data is available since 1976, but earlier surveys had a lot of noise in the occupation and industry classifications (more literature review here). Since 1994 the survey started to incorporate dependent interviewing in which labor force information reported in previous month is either confirmed or updated in the current month. Whereas before 1994, these questions were asked and coded independently in each month . The dependent interviewing procedure greatly reduces the coding errors in occupation. I only use data from 1994 for better estimates of separation rates.

Because the matching algorithm fails to identify respondents from certain consecutive months in 1995 because BLS deliberately broke the linking between certain months in 1995 to protect privacy, and the independent coding of occupation was too noisy before 1994, the sample period is between 1996 and 2013. The respondents are restricted to be 16 years old and above, and civilian household member.

The matching algorithm:

- Download original dat. and extraction do. files from NBER. The extraction do. files are
- 
- is responsible for maintaining CPS basic months.



modified to extract multiple months in a year using loop. The extracted files are saved in dta. format.

- The extracted dta files are further modified before matching. This procedure is to ensure that all monthly dta files have the same variables names and no duplicated entries. This step includes:
  - Matching id includes: *hh\_id*, *hh\_sample*, *hh\_serial*, *hh\_num*, *per\_line*.
  - Personal info includes: *mis*, *sex*, *age*, *race*, *per\_type*, *occ*, *worker\_class*, *ind*, *job\_search*, *lfs*, *emp*, *unemp*, *nlf*, *final\_weight*.
  - Rename and create above variables.
  - Only keep respondents who are at least 16 years old and being civilian population (not serving in the military). About 42,000-50,000 observations are dropped.
  - For completely duplicated entries, i.e. both matching id and personal info are the same, only keep one entry. No entry is dropped actually.
  - Drop all partially duplicated entry, i.e. same matching id but different personal info, since I don't know which one is true. No entry is dropped for the normal versions.
  - There is one issue with matching 200404 and 200405, part of the matching id, *hh\_serial*, changed its coding after 200404 and become incomparable between these two months. Therefore for these two months, I created additional versions for matching. These additional versions are matched using only *hh\_id*, *hh\_sample*, *hh\_num*, *per\_line*. As a result, the remaining matching identities may not uniquely identify a respondents, and partially duplicated entries do arise. Luckily, the numbers of duplication, 88 and 125, are tiny comparing to the size of the sample and hence should not have any effects on the following estimations.
- To ensure that the above cleaning process is correct, I compute the labor force status from the modified files using *final\_weight* as the weight, the implied participation and unemployment rates are almost the same as the ones published by BLS.
- To find out whether a worker works in the same occupation in the following month, I need to match CPS from month  $t$  with month  $t + 1$ . Furthermore, to avoid the possible correlations between occupation mobility and location change, I only look at mobility transitions in the

first four months of survey. The theoretical matching rate should be around three eighths and the actual average matching rate 34%. To ensure this subsample is still fairly representative of the population, I plot the implied participation and unemployment rates from the matched sample and compare them to the BLS published rates in figure 1.7. The implied participation rate matches well with BLS except for 2002 February where the implied participation is 4% higher than the BLS figure. The implied unemployment rate lines up very well with BLS figures. In general, the matching process does not introduce any significant selection bias and the matched sample still well represents the large sample as well as the population.

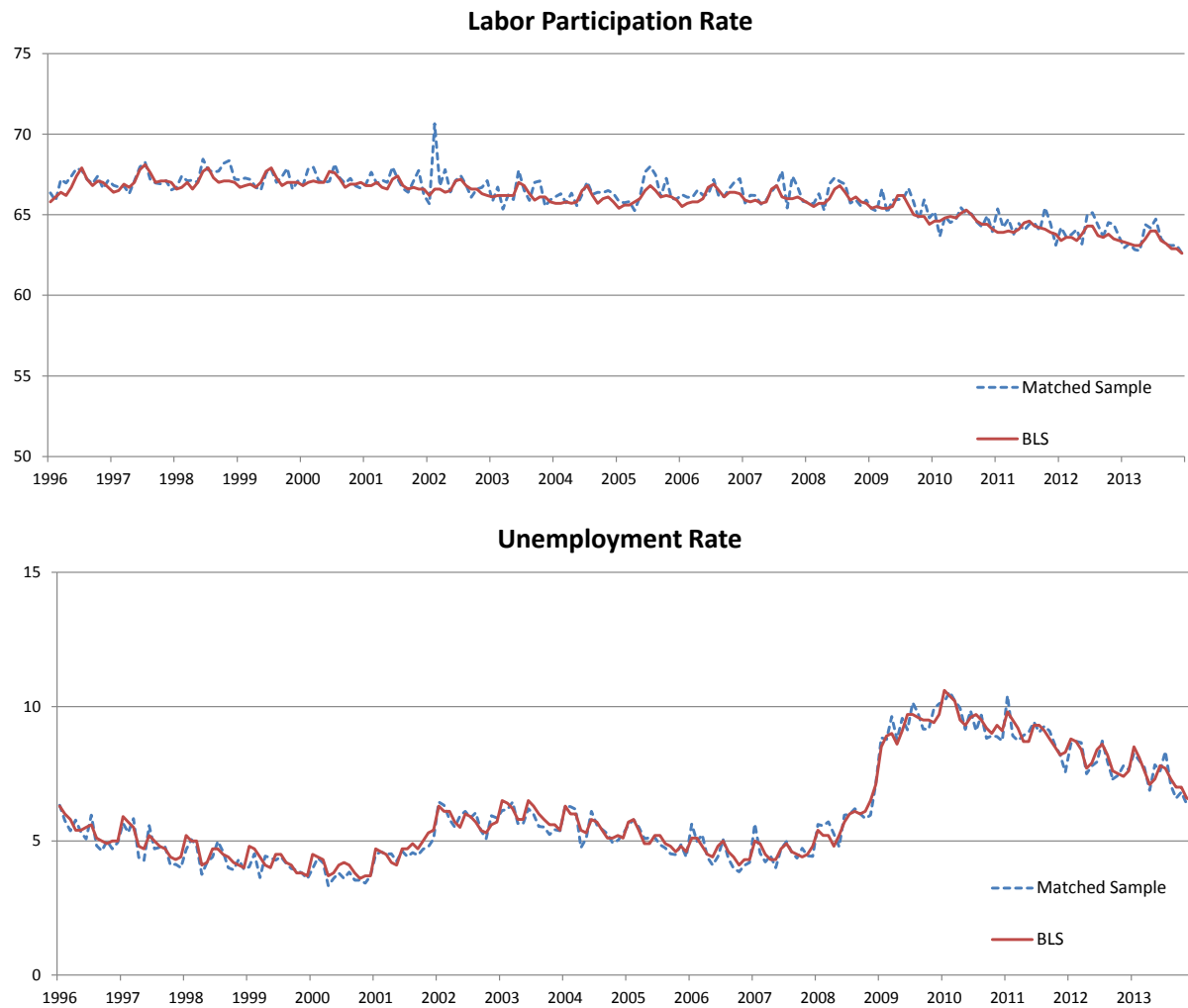
### **1.A.3 Estimate labor force and occupational transition rates**

To identify the current labor force status, I use the monthly labor force variable PEMLR. To identify the occupation, I use the occupation code (PEIO1OCD) for primary job. This occupation variable is available for respondents who are: a) employed, b) unemployed and have worked before, c) not in the labor force but have worked in the last 12 months. Composite final weight, PWCMPGTG, is used in all calculation. To identify a valid occupational change, I follow the algorithm proposed by Moscarini and Thomsson (2007). A valid occupational requires an occupational change from month  $t$  to  $t + 1$  and satisfies any of the following: a) class of workers (private firm, federal, government or self-employed) changed, b) the three-digit industry code has changed, c) the respondent had looked for work in the past four weeks.

The occupation code changed twice during the sample period, at January 2003 and January 2011. The occupation before January 2003 were coded into a three-digit number, but since January 2003 it became a four-digit code. At the result, almost all workers in January 2003 were identified as a new worker. The occupation code change in January 2011 was less comprehensive but still led to occupation change rate twice as higher as the average. Therefore I treat the new worker rate as missing values and imputed from other periods.

To remedy the issues caused by occupational code changes, I use the strong seasonality patterns of EE occupational mobility rate to impute the missing values in both December 2002 and December 2010. The EE occupational mobility rate in December is 0.22% higher than that in January of the following year on average except for year 2002 and 2010. The missing

**Figure 1.7** Comparing matched sample and BLS



The solid line is the BLS monthly seasonally unadjusted rate, and the dotted line is the implied rate from the matched sample.

value in December 2002 is imputed as the sum of 0.22% and the rate in January 2003, i.e.  $0.22\% + 2.04\% = 2.26\%$ . The missing value in December 2010 is imputed in the same way.

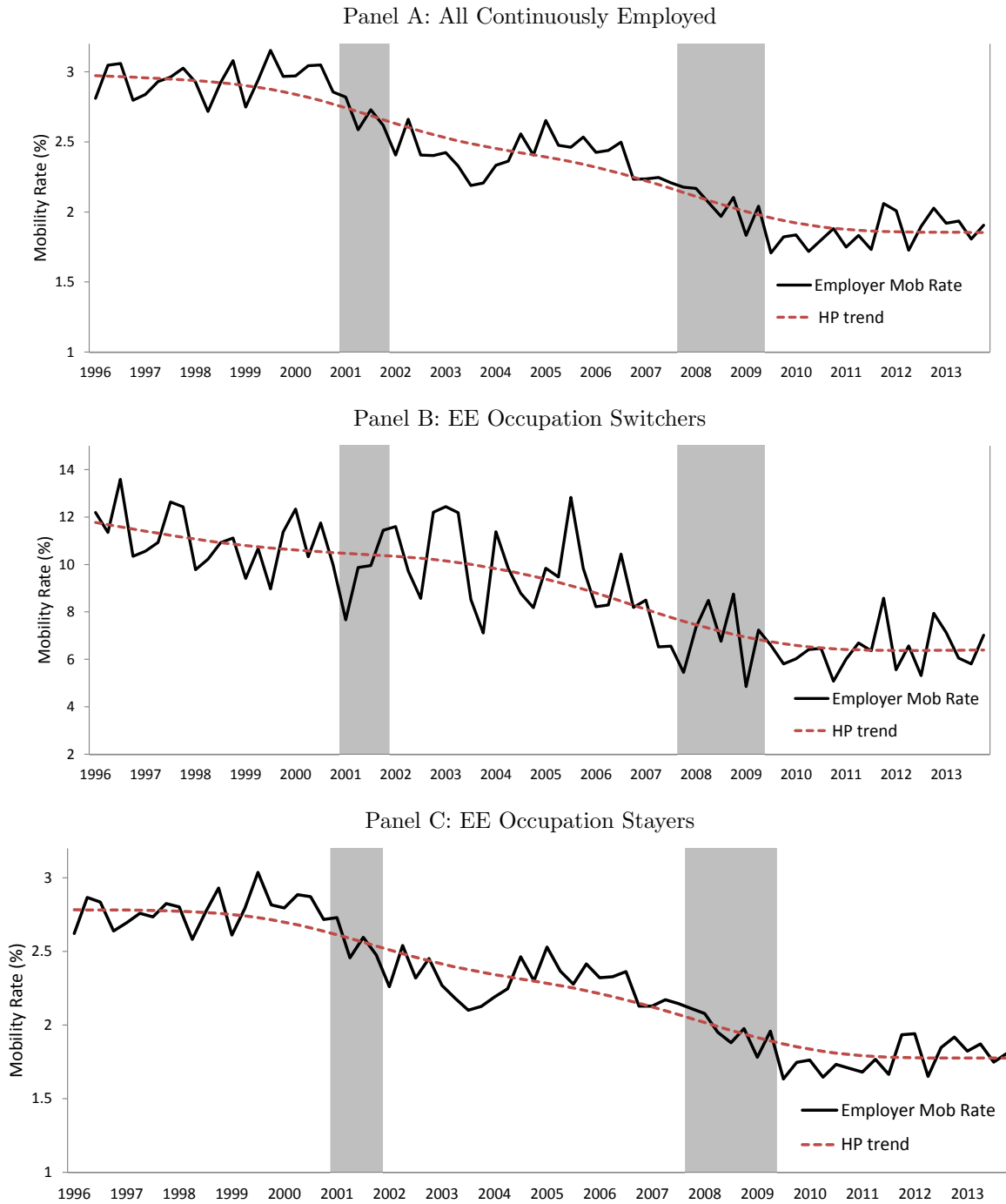
#### **1.A.4 EE employer mobility rate**

A large fraction of EE occupation switch take place within the same firm. Each month respondents are asked if they still work for the same employer as in last month, and the variable PUIODP1 records the answer to this question. However, a significant fraction of continuously employed respondents have blank entries for this variable. Therefore I report the EE employer mobility rate only based on the sample who had either yes or no response in Figure 1.8. Panel A is based on all continuously employed respondents. Panel B is based on the continuously employed respondents who switched occupation, and panel C is based on the continuously employed respondents who stayed in the same occupation. None of the graph exhibits any significant cyclicity at the business cycle. During both 2001-2002 and 2007-2009 recessions, the employer mobility rates did not deviate from the HP trend any further in both direction than during any other period. Of course, the low response rate for this question may undermine the reliability of estimate. The average response rate is 59.4% for all continuously employed respondents, 36.7% for those who have switched occupation and 60.1% for those who stayed in the same occupation.

#### **1.A.5 Data moments Robustness Check**

In the paper I assume a constant 50% occupational mobility rate for people who leave the workforce, and the workforce occupational mobility rate is computed as the sum of half of workforce turnover rate and EE occupational mobility rate multiplied by one minus workforce mobility rate. Here I use two different occupational mobility rate, 20% and 80%, for people who leave the workforce. The correlation between productivity and workforce occupational separation rate varies a little under different assumptions.

**Figure 1.8 :EE Employer Mobility Rate**



a), The top panel is based on all continuously employed respondents. The middle panel is based on the continuously employed respondents who switched occupation. The bottom panel is based on the continuously employed respondents who stayed in the same occupation.

b), The black solid line denotes the seasonally adjusted monthly employer mobility rate, averaged at quarterly frequency. The red dotted line denotes the HP trend.

**Table 1.4 : Workforce Occupational Separation Rate**

Occ. Mob. Rate. For Leavers	20%	50%	80%
Std	0.05	0.04	0.03
Autocorr.	0.57	0.55	0.52
$\text{corr}(s, U)$	-0.31	-0.20	-0.09
$\text{corr}(s, V)$	0.17	0.05	-0.05
$\text{corr}(s, V/U)$	0.24	0.13	0.02
$\text{corr}(s, Y)$	0.19	0.06	-0.05
$\text{corr}(s, X)$	-0.35	-0.38	-0.38

The table presents the key statistics for different occupational mobility rate for workers who leave the workforce. The sample is from 1996Q1 to 2013Q4.

## 1.B Equivalent Economy

The economy has zero worker share of the surplus and the firm retains all the surplus. In this section I will prove that it is equivalent to an economy with positive worker share of the surplus under a slightly different parameterization. Let  $\eta$  denote the worker share of the match surplus and  $\tilde{S}^\tau(A^m, X^k, Z^s)$  the continuation value for the firm after production and exogenous separation for the alternative economy. The total surplus from production in each period is  $y_i - b$ , and the payoff to the firm is  $(1 - \eta)(y_i - b)$ . The continuation value in the final period  $\bar{\tau}$  is zero by definition.

$$\tilde{S}^{\bar{\tau}}(A^m, X^k, Z^s) = 0 = (1 - \eta)S^{\bar{\tau}}(A^m, X^k, Z^s) \quad (1.15)$$

It can be proved by induction that for  $\tau = 1 \dots \bar{\tau}$ ,  $\tilde{S}^\tau(A^m, X^k, Z^s) = (1 - \eta)S^\tau(A^m, X^k, Z^s)$ . Assume this holds for  $\tau$ , then for  $\tau - 1$ :

$$\begin{aligned}
\tilde{S}^{\tau-1}(A^m, X^k, Z^s) &= \beta \sum_{n=1}^M P_{m,n}^A \sum_{j=1}^K P_{k,j}^X \sum_{r=1}^S P_{s,r}^Z \{ (1 - \eta)(A^n X^j Z^r H(\tau) - b) \\
&\quad + (1 - s) \max(0, \tilde{S}^\tau(A^n, X^j, Z^r)) \} \\
&= (1 - \eta) \beta \sum_{n=1}^M P_{m,n}^A \sum_{j=1}^K P_{k,j}^X \sum_{r=1}^S P_{s,r}^Z \{ (A^n X^j Z^r H(\tau) - b) \\
&\quad + (1 - s) \max(0, S^\tau(A^n, X^j, Z^r)) \} \\
&= (1 - \eta) S^{\tau-1}(A^m, X^k, Z^s)
\end{aligned}$$

The firing decision for a match with  $\{A^m, X^k, Z^s, \tau\}$  solely depends on the sign of  $\tilde{S}^\tau(A^m, X^k, Z^s)$ , which is non-positive if and only if  $S^\tau(A^m, X^k, Z^s)$  is non-positive. Hence the firing decisions for both economies are the same.

Similarly one can show that  $\tilde{S}^0(A^m, X^k) = (1 - \eta)S^0(A^m, X^k)$ . The hiring decision is a function of  $\kappa_0, \kappa_1$ . Given  $q(\theta_t)$ , the firm  $o$  posts vacancies until  $\tilde{\kappa}_0 + 2\tilde{\kappa}_1 V_{ot} \geq q(\theta_t)\tilde{S}^0(A^m, X^k)$ . Set  $\tilde{\kappa}_0 = (1 - \eta)\kappa_0$  and  $\tilde{\kappa}_1 = (1 - \eta)\kappa_1$ , the hiring condition in this alternative economy is the same as the hiring condition in the original economy.

To conclude, in order to generate time-varying procyclical wage rate and aggregate wage payment, one only needs to introduce a constant worker share of the surplus  $\eta$ , and re-scale  $\kappa_0$  and  $\kappa_1$ . The alternative economy has the same hiring and firing decisions as the original economy. In other words, the constant wage rate only serves to facilitate computations.

## 1.C Solve the Model Numerically

### 1.C.1 Calibrating $H(\tau)$

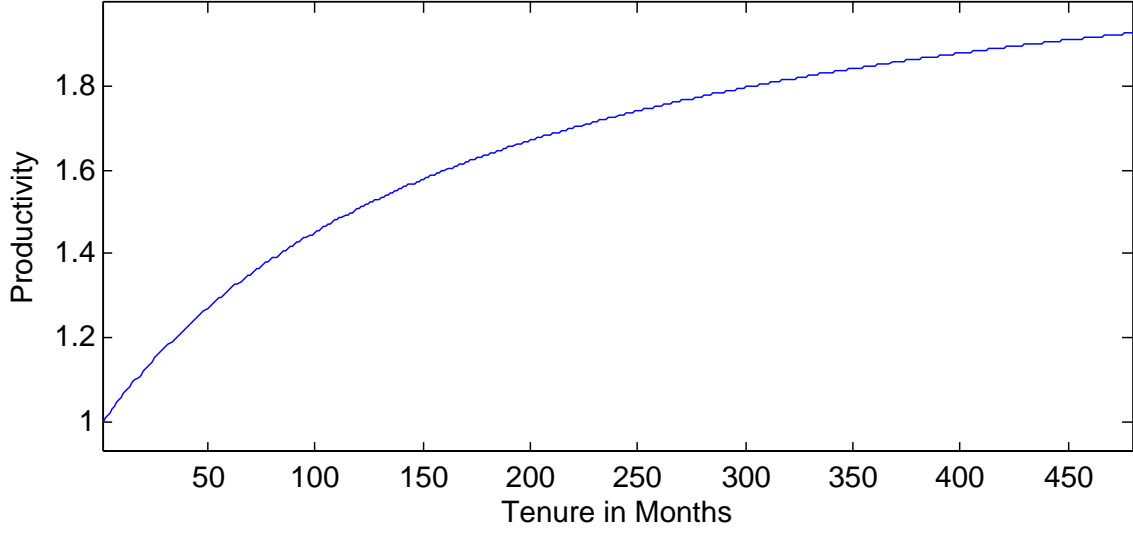
The productivity gain in occupational tenure  $H(\tau)$  is assumed to have a simple functional form  $H(\tau) = a(1 - \frac{b}{c+\tau})^d$ . Kambourov and Mankovskii (2009) finds that the wage increases by about 5.4% after 2 years, about 12% after 5 years and 17% after 8 years, and those increases represent 38% of the productivity gain. As the result  $H(\tau) = 2.2657 \cdot (1 - \frac{35.42}{\tau+136.01})^{2.2657}$ . This concave function starts at 1 and grows gradually, where  $H(1) = 1, H(25) = 1.142, H(61) = 1.316, H(97) = 1.447$ . It is plotted in figure 1.9.

### 1.C.2 Model Algorithm

In this part I show how I to solve for hiring decisions at equilibrium. The state variables are  $\{A_t, \Lambda_t\}$ . The unemployment is  $U_t = 1 - \sum_X \sum_Z \sum_{\tau \geq 1} \lambda_t^{X,Z,\tau}$ . The value of a new vacancy for a firm with occupational productivity  $X^k$  is  $q(\theta_t)S^0(A_t, X^k)$ . At equilibrium  $\theta_t = \frac{V_t}{U_t} = \frac{\sum_{k=1}^K \tilde{\pi}^X(k)V_t^k}{U_t}$ .

The hiring decision is solved numerically by bisection. For any given  $V_t$ , the implied total

**Figure 1.9** Increasing Productivity in Tenure



vacancy is:

$$V_t^{implied} = \sum_{k=1}^K \tilde{\pi}^X(k) V_t^k \quad (1.16)$$

$$V_t^k = \frac{1}{2\kappa_1} \max\left(0, q(V_t/U_t)S^0(A_t, X^k) - \kappa_0\right) \quad (1.17)$$

$V_t^{implied}$  is monotonically decreasing in  $V_t$ . The starting lower bound for  $V_t$  is 0, in this case the  $V_t^{implied} \geq 0$  where the equality holds if  $S^0(A_t, X^K) \leq \kappa_0$ . The starting upper bound for  $V_t$  is  $U_t((\frac{\kappa_0}{S^0(A_t, X^K)})^\iota - 1)^{\frac{1}{\iota}}$ , i.e. the highest  $V_t$  such that the firm with the highest occupational productivity  $X^K$  is just indifferent between posting or not.



## Chapter 2

# Shopping Time

### 2.1 Introduction

Frictions matter in markets. This viewpoint has proved successful in analysis of both the labor and credit markets. More recently, a body of research has modeled search frictional goods markets. This research has allowed for a better understanding of rationing in the goods markets and its role in the propagation of business cycles. However, there is no consensus on the cyclical importance of an important variable in this search frictional goods market, namely the cyclical importance of aggregate effort exerted by consumers.

In Bai et al. (2011), Gourio and Rudanko (2013) and Petrosky-Nadeau and Wasmer (2015), endogenous consumer shopping effort is procyclical. Notably, it increases with income. In Kaplan and Menzio (2013), consumer effort is exogenous and fixed over time but, by fixing the effort of the unemployed above that of the employed, as the former are assumed to search harder to find better prices, aggregate time shopping appears to be countercyclical. In this paper we first clarify the various interactions between income, shopping time, and working time, and review the theoretical ingredients needed to support the alternative patterns of covariations with the cycle. We then use the detailed daily time use diaries of the American Time Use Survey (ATUS) conducted by the Bureau of Labor Statistics (BLS) to measure the cyclical importance of consumer search in the goods market.

From a theoretical perspective, under standard assumptions on utility and cost functions of shopping effort, shopping time, thought of as a combination of effort before purchasing (e.g. comparing prices, hereafter pre-match search effort) and effort while purchasing (hereafter, purchasing effort), are procyclical. Purchasing effort increases with income because higher income reduces the opportunity cost of buying search goods at the cost of buying non-frictional goods. A less trivial results is that pre-match search effort depends positively on the surplus from consumption. The consumption surplus itself depends on income. Consumers therefore spend more time and effort to consume following a rise in income. Prices, when they are bargained, respond positively to income and attenuate the procyclicality result. When quantities can adjust, they respond positively to income and thus further raise the consumption surplus. This strengthens the procyclicality under price bargaining. However, the cyclicity of shopping time disappears under competitive pricing. There are also forces in the opposite direction. In the face of price dispersion and a reservation search strategy for consumers, an increase in income is associated with a higher reservation price less search effort for goods. Similarly, when working time can be chosen freely, shopping effort and working time covary negatively. Hence, a rise in the hourly wage - due for instance to a productivity innovation generating the business cycle - raises hours if the substitution effect dominates the income effect. This leads to less shopping time. Finally, pre-match search effort can be interpreted as also occur simultaneous with the effort undertaken while shopping for other goods. This would a case, say, when in a grocery store, an individual spends time searching for new yogurts after having filled the basket with salt, butter, and sour cream.

Whether procyclicality or countercyclicality dominates is, ultimately, an empirical question for which we use the American Time Use Survey from 2003 to 2013. The ATUS includes over 400 distinct time use categories. Our main task is to identify various components of shopping time. We settle on 25 time use categories that broadly encompass time spent shopping for consumer goods and services, and, separately, on groceries, gas and food (GFC).

We obtain three main results. First, we find that aggregate search by consumers in the goods market declined with the onset of the Great Recession. This is true for all labor market statuses, employed, unemployed, and nonparticipants. However, we find that the time allocation to finding and acquiring goods and services declined most for the unemployed. Prior to December 2007 the unemployed, and non-participants, spent more time searching in the goods market than

the employed. With the onset of the Great Recession the unemployed drastically reduced their time searching for goods and services, spending less time on this activity than the employed by 2012.

Second, there is a positive relation between cross-state variations in GDP per capita and our different measures of search effort in the goods markets. States with the largest declines in GDP per capita tended to have the largest declines in time spent shopping for goods and services. In Michigan, for instance, there was 21% decline in time spent in this shopping category and a 10% decline in GDP per capita. Oklahoma, with a very different experience over the period in question, experienced a 2% increase in GDP per capita and a 20% increase in shopping time.

Third, we find that search effort in the goods market is increasing in individual income and household income. This result is robust to controlling for state of residence and various demographic characteristics such as age, gender, education, and marital status. The one exception is time spent shopping for groceries, gas, and food, which is unrelated to either income variable. Overall, we don't find much evidence in favor of a negative correlation between income and shopping time.

We also find that nonparticipants and the employed spent more time than the unemployed in this activity. This may indicate that most of the effort made by consumers in the goods market is unrelated to uncovering better prices, although it also suggests that it is difficult to properly measure this activity using the ATUS.

Overall, this body of evidence supports a conclusion that price comparisons cannot be a driver of business cycles, which contradicts recently published work by Kaplan and Menzio (2015) where fluctuations and multiple equilibria arise precisely from the fact that in recessions the unemployed search more for lower price goods, depressing the economy further. Nothing in our investigation of the ATUS data supports this mechanism.

In contrast, our result confirm a negative correlation between working hours and shopping time found in Aguiar et al. (2013). The opposite finding would have been surprising. The time budget constraint is less tight in a recession. Households have more time to allocate to various non-work activities. However, this does not imply that forces pushing towards a countercyclicality of shopping time dominate over the business cycles, for the reasons indicated above. Our conclusion is that models where the consumption surplus and search effort in the goods market are procyclical

are more relevant for discussing business cycles.

Section 2.2 reviews the various mechanisms at play between income, shopping time, prices, and working time and classifies them into procyclical forces and countercyclical forces. Section 2.3 describes the ATUS and the time use categories we employ in this study. Section 2.4 describes aggregate trends in shopping time over the sample period 2003-2012 and according to labor force status. Section 2.5 then measures the business cycle and income elasticity of time spent searching for goods and services, and discusses some robustness issues. Section 6 discusses the individual regression of shopping time and income. Section 2.7 concludes.

## 2.2 The cyclical pattern of shopping time

The early search and matching literature considered the market for goods and services as the prototypical setting in which buyers and sellers are engaged in a costly and time consuming process to find and establish trading relationships (Diamond, 1982). Firms exert effort and resources advertising their products and maintaining their customer relationships. Consumers spend time selecting and waiting to obtain goods and services, adding and removing items from their consumption basket.<sup>1</sup> Departures from market clearing assumptions introduce a range of possible price determination mechanisms that have important implications for equilibrium allocations in the long run and over the business cycle.

### 2.2.1 A setup for thinking about shopping time

Consumers may spend more or less time shopping. A good framework for thinking about shopping time is to disentangle this time between pre-match effort denoted by  $e_0$  (which may be interpreted as the match to a marginal good while purchasing other goods) and purchasing time  $e_1$ , which transforms time into actual purchasing. For the sake of simplicity, we assume that the relation between the quantity of search good purchased denoted by  $x$  is linear in time spent  $e_1$ .

---

<sup>1</sup>On the procyclicality of advertising and customer relationship building expenditures, along with their implications for product market frictions, see Hall (2012) and Gourio and Rudanko (2013). For a flow approach to a typical household's consumption basket, see Broda and Weinstein (2010) for empirical evidence and Petrosky-Nadeau and Wasmer (2011), Bai et al. (2011), and Michaillat and Saez (2014) for a modeling approach. den Haan (2013) consider the role of product market imperfections for the business cycle of inventories.

The purpose of this section is to identify the mechanisms that lead to either procyclical or countercyclical shopping time over the business cycle. Although the relevant search margin is  $e_0$ , we start by showing that both  $e_0$  and  $e_1$  contain strong procyclical forces. Then we focus more on  $e_0$ . The latter search effort can be thought of as time spent locating products in the goods market. This effort is driven mostly by two different incentives. First, it may be used to match faster with a good and thereby enjoy the associated consumption surplus more quickly. Alternatively, it may be used to locate the best price in the presence of price dispersion across sellers. The two motives may be at play simultaneously. Further, an individual consumer's probability of finding a product in the goods market will depend, in general, on individual effort relative to the effort by other consumers, as well as the amount of consumers and sellers out there in the market.

At the individual level, the probability of finding a particular good is given by  $(e_0/\bar{e}_0)\lambda$ , where  $\lambda$  is an aggregate contact rate with a product in the goods market. This rate depends on the aggregate amount of consumers and sellers in the market. If the aggregate amounts of consumers and sellers in the goods market are denoted by, respectively,  $C$  and  $S$ , the matching literature typically postulates an increasing aggregate matching function  $M(\bar{e}_0 C, S)$  that governs the number of matches per unit of time. As such, the aggregate rate of matching for consumers is given by  $\lambda = M(\bar{e}_0 C, S)/C$ .<sup>2</sup>

Assume that the consumer's preferences are summarized by an increasing and concave utility function  $v(x, y, l)$  associating two goods: the search good denoted by  $x$ , all other goods acquired on frictionless markets,  $y$ , and finally leisure, which is the total time budget less hours worked less shopping time  $e_0$  and  $e_1$ . Let  $h$  be hours worked, and the time budget is normalized to 1.

Once in a match with a seller, a consumer spends  $px$  out of disposable income  $\omega$  to consume the search good. The remaining income is spent on the consumption of other goods. We use the following compact notation:  $v(\mathbf{1}) = v(x, \omega - px, 1 - h - e_1)$  denotes the utility from consuming the search good, and  $v(\mathbf{0}) = v(0, \omega, 1 - h - e_0)$  denotes the utility from consuming the only the other goods.  $\Delta v \equiv v(\mathbf{1}) - v(\mathbf{0})$  denotes their difference.<sup>3</sup> Hence, the value of the consumer's surplus is

<sup>2</sup>Appendix 2.A.1 explores a more general specification of individual consumption match probability where individual and aggregate efforts are complementary, allowing for different consumption effort externalities.

<sup>3</sup>We also denote by a subscript  $x$  or  $y$  the first-order derivative of any function of  $x$  and  $y$ . For instance, the full differentiation of the consumption surplus with respect to a change in income  $\omega$  and price  $p$  is given by:

given as:

$$\Omega = v(\mathbf{1}) - v(\mathbf{0}) = \Delta v \quad (2.2)$$

The properties of  $\Delta v$  and its derivatives with respect to  $y$  and  $x$  will be very important for what follows. In particular, we will assume marginal decreasing utility in both goods ( $v_{yy} < 0$  and  $v_{xx} < 0$ ). Under separability between the two types of goods, this leads in particular to

$$\Delta v_y \equiv (v_y(\mathbf{1}) - v_y(\mathbf{0})) > 0 \quad (2.3)$$

A marginal unit of consumption of good 0 has a greater value when the agent diverts part of its income to also consume the search good. The same is true under nonseparability when the degree of substitutability between the two goods is not strong enough. Appendix 2.A.2 derives this more general case.

Consumers incur a cost to search effort in the goods market described by the returns to leisure  $v_l$ . Given this environment, the generic optimality condition for consumer search effort states that effort is increasing in the expected surplus from finding a good in the product market:

$$v_l(\mathbf{0}) = \frac{\lambda}{\bar{e}_0} \Omega_c \quad (2.4)$$

while the other first order condition for purchasing time would be:

$$v_l(\mathbf{1}) = \frac{\partial x}{\partial e_1} [v_x(\mathbf{1}) - p v_y(\mathbf{1})]$$

This condition states that increasing purchasing time of the search good raises the opportunity cost of time (less leisure), reduces the consumption of the non-search good, and equals the marginal utility from consumption of the search good.

The following can already be learned about forces leaning towards procyclicality of shopping time. First, Pre-match effort  $e_0$  has the the same cyclicity as the surplus  $\Omega_c$ . This

---


$$\begin{aligned} d\Delta v &= (v_y(x, \omega - px) - v_y(0, \omega)) d\omega - x v_y(x, \omega - px) dp = (v_y(\mathbf{1}) - v_y(\mathbf{0})) d\omega - x v_y(\mathbf{1}) dp \\ &= \Delta v_y d\omega - x v_y(\mathbf{1}) dp \end{aligned} \quad (2.1)$$

will be shown to be procyclical. Second, consumption effort  $e_1$  is procyclical as well in the simple case where prices, hours and  $x$  are a-cyclical. This arises from the fact that higher income  $\omega$  leads to a lower marginal utility of consumption of the numeraire, such that the return on consumption effort  $e_1$  is to reduce the consumption of  $x$  which has lower value.

Conversely, if hours are procyclical, then the marginal value of leisure is higher, and hence both  $e_0$  and  $e_1$  may be countercyclical. Moreover, if prices are procyclical, then return on consuming goods  $x$  is lower (it has a greater opportunity cost in terms of numeraire  $y$ ) and thus there is an additional force towards countercyclicity. We show this formally in the Appendix.

Whether procyclicality or countercyclicity dominates depends on the utility function and price determination mechanism, but it ends up ultimately in being an empirical question, explored in next Section.

### 2.2.2 Surplus sharing with a fixed supply of the search good $x$

Assume for the moment that the supply of the search good is fixed at some level. We explore two alternative pricing rules.

#### Price setting - simple sharing rule:

The simplest price setting mechanism is a price that equalizes the surplus to the consumer and the firm, i.e., such that  $\Omega_c = \Omega_f$ . This results in a simple price rule:

$$xp = \Delta v + \kappa(x) \tag{2.5}$$

A positive, procyclical, shock to income  $d\omega$ , keeping the cost  $\kappa$  constant, leads to:

$$x \frac{dp}{d\omega} = \frac{\Delta v_y}{1 + v_y(\mathbf{1})} > 0 \tag{2.6}$$

The assumption that utility is increasing, concave and separable in  $x$  and  $y$ , implies that the price varies positively with income.<sup>4</sup>

The cyclicalty of the consumer and firm match surpluses follows easily from their definitions in equations (2.2) and (??). We have under the inequality (2.3):

$$\begin{aligned}\frac{d\Omega_f}{d\omega} &= x \frac{dp}{d\omega} > 0 \\ \frac{d\Omega_c}{d\omega} &= \Delta v_y - x v_y(\mathbf{1}) \frac{dp}{d\omega} = \frac{\Delta v_y}{1 + v_y(\mathbf{1})} > 0\end{aligned}$$

Hence, a positive income shock raises the consumer's surplus and, from the optimal search effort condition (2.4), individual search effort in the goods market. Moreover, the same shock will raise the number of sellers, thus raising  $\lambda$ . Therefore, consumer search effort is procyclical.

In Appendix, we explore other dimensions for completeness: price bargaining ; the endogeneity of the supply of the search good, the existence of significant price dispersion and the endogeneity of the choice of hours worked.

## 2.3 Searching for goods and services in the ATUS

We use data from the 2003-2012 waves of the ATUS, conducted by the BLS drawing on individuals from the exiting sample of the Current Population Survey (CPS). The types of activities recorded in the ATUS are described in detail in Hammermesh et al. (2005), and have been used to document changes in overall time use during the Great Recession (Aguiar et al. 2013), with a particular emphasis on how individuals reallocate decreased hours of market work to other activities.

We focus on time spent in the process of selecting and acquiring goods and services. Overall we select 25 categories out of more than 400, which include time spent traveling associated with purchasing *marketized* goods and services. A potential 26th category, travel time related to relaxing and leisure, was excluded even if it may include some market activities. Each category is mutually exclusive and sum to total time spent shopping:<sup>5</sup>

<sup>4</sup>See Appendix 2.A.2 for a more general case where the inequality holds under imperfect substitutability between  $y$  and  $x$ . See also Petrosky-Nadeau and Wasmer (2014) for an application of this property to fiscal shocks in a frictional goods market.

<sup>5</sup>Appendix 2.B.1 provides the ATUS time use codes that compose each category.



- 1 - **Consumer goods and services** is divided into three subcategories. The first is time spent shopping other than for groceries, gas, and food. The second is time spent researching purchases. The third corresponds to waiting associated with purchasing services.
- 2 - **Groceries, Gas, and Food (GGF)** includes all time spent shopping for groceries, gas, and food. We present results for all three subcategories.
- 3 - **Travel time** includes all travel associated with the purchasing of goods and services.

We focus on the population ages of 24 and 55. We exclude the population aged 16 to 24 because labor force status, such as unemployment and participation, line up very closely with the CPS based rates published by the BLS for all age groups except this younger group.<sup>6</sup> We remove those aged 55 and over because labor force participation rates for these individuals decline dramatically, whereas our main results emphasize differences across labor force status. We also exclude respondents with a positive amount of unclassified time. In total we have 66,958 individuals in our sample for the baseline results.<sup>7</sup> We use the sample weights provided by the ATUS to aggregate responses to either year or state-year averages.

## 2.4 Aggregate trends in shopping time

Figure 2.1 plots the annual data for our main components of shopping time, in the aggregate – averaging across all individuals in the sample (solid black line) – and by labor force status. Time spent by the employed is plotted by the red dashed line, time spent by the unemployed by the triangle green line, and time spent by nonparticipants by the blue circled line. Table 2.1 provides the average values for the sample period, 2003 to 2012. The table also reports averages by gender and marital status, as well as for the population over age 55.

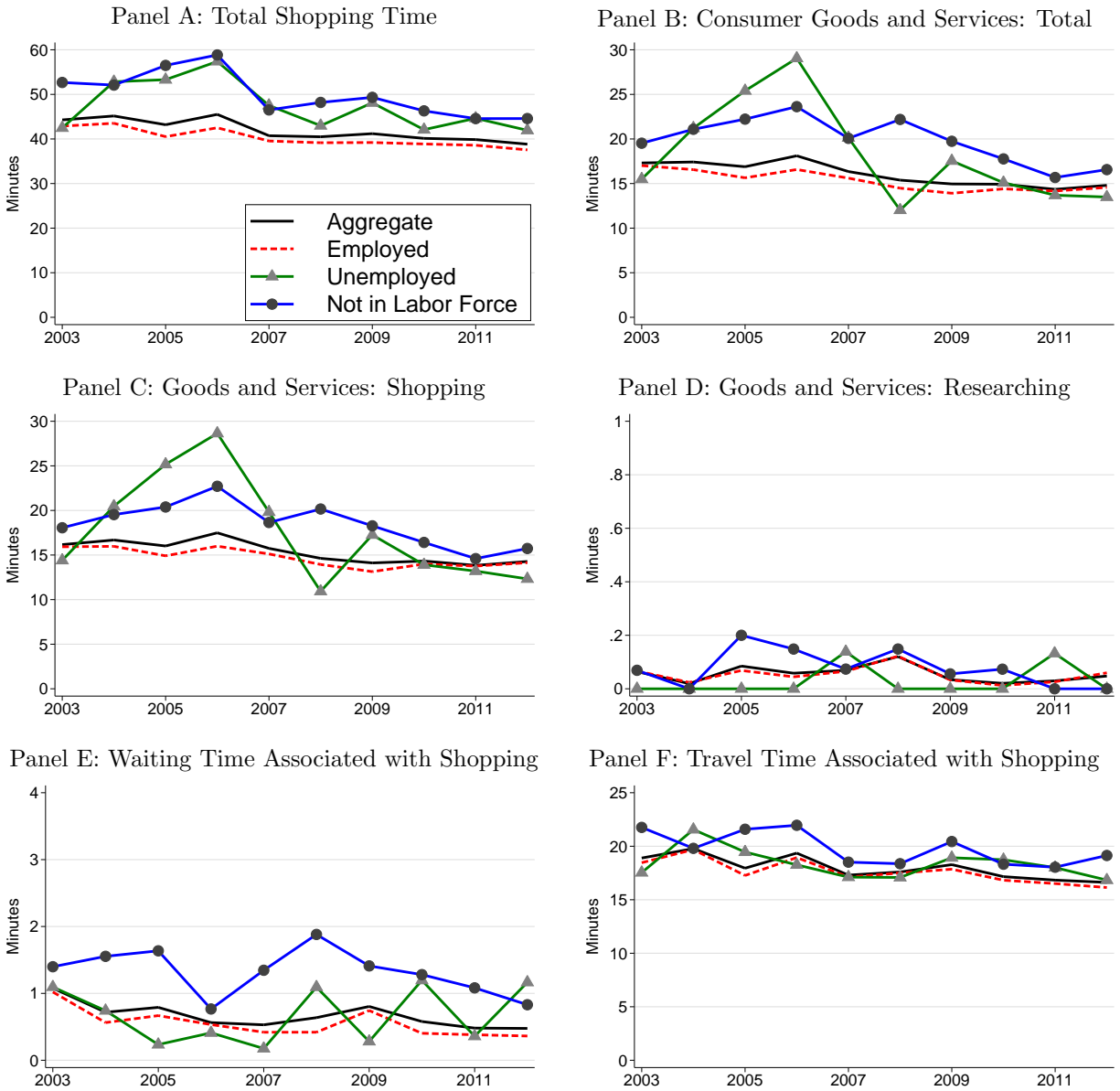
Aggregate time spent shopping for goods and services, plotted in Panel A of Figure 2.1, averaged 42 minutes per day over the sample period. Two main observations arise from this first look

---

<sup>6</sup>We have attempted to identify the reasons for this discrepancy with the BLS in private communications but have so far been unsuccessful.

<sup>7</sup>Appendix B provides more details for these deletions. The results for the whole age sample change very little. The results are available upon request.

**Figure 2.1** Shopping Time: Age 25 to 54 by Labor Force Status



The solid black line represents the aggregate for the age category 25 to 54 years. Time spent by the employed is plotted by the red, dashed line, time spent by the unemployed by the triangle green line and time spent by non-participants by the blue, circled line. The ATUS definition for time spent in Panels A through F are detail in the appendix.

at the data. First, individuals not in the labor force spend the most time shopping for goods and services, an average of 50 minutes a day. Second, individuals not in the labor force and unemployed display a similar pattern of total time spent shopping for goods and services, with a pronounced decline in time spent starting in 2007.

**Table 2.1** Shopping Time by Labor Force Status and by Age - Minutes per Day

Categories	Agg.	Ages: 25 to 54						Age:			
		Labor force status			Men		Women				
		Emp.	Unem.	NLF	All	Married Single	All		Married Single		
Total Shopping Time	42.0	40.2	47.3	50.1	34.2	34.6	33.6	49.8	52.5	45.2	44.9
Consumer Goods and Services	16.1	15.3	18.3	20.0	12.3	12.6	11.9	19.8	20.8	18.1	17.4
-Consumer Goods	15.3	14.7	17.6	18.4	11.8	12.1	11.4	18.9	20.0	17.0	15.7
-Researching G&S	0.07	0.05	0.03	0.20	0.06	0.07	0.04	0.09	0.07	0.13	0.07
-Waiting for services	0.67	0.55	0.68	1.32	0.51	0.51	0.51	0.82	0.72	1.00	1.55
Groceries, Gas and Food	7.91	7.31	10.7	10.3	5.82	5.82	5.82	10.0	10.8	8.67	8.48
-Groceries	6.25	5.56	9.36	9.06	4.22	4.24	4.18	8.31	9.10	6.89	7.32
-Gas	0.41	0.43	0.35	0.30	0.42	0.44	0.39	0.39	0.37	0.42	0.38
-Food	1.25	1.32	0.94	0.95	1.18	1.13	1.24	1.33	1.31	1.36	0.78
Travel time	18.0	17.6	18.4	19.8	16.0	16.1	15.9	20.0	20.9	18.4	19.0

Notes: ATUS, sample restricted to respondents with no unclassified time. "Agg.", "Emp.", "Unem." and "NLF" correspond to "Aggregate", "Employed", "Unemployed" and "Not in Labor Force". For the baseline sample of individuals aged 25 to 54, there were 10,434 respondents per year, on average. For both gender, married status refers to spouse present only and does not include unmarried partner present.

The second panel of Figure 2.1 plots time spent shopping for goods and services other than groceries, gas, and food, and excluding the time spent traveling associated with shopping activities. This is the core measure of search effort by consumers in the goods market. The average time spent in a day over the period 2003-2012 is 16 minutes, with nonparticipants spending the most time, 20 minutes a day. Interestingly, the unemployed start out resembling nonparticipants early in the sample time period, even spending more time shopping for goods and services by a significant margin in 2006. In the second half of the sample, after the onset of the Great Recession, the unemployed are very similar to the employed in time spent shopping for goods and services. By 2012, the unemployed spend less time shopping for goods and services than the employed.

Average time spent in researching goods and services is small, averaging 0.07 minute a day. However, we note that nonparticipants spent the most time, 0.20 minute, and that the employed spent on average more time than the unemployed researching goods and services (see Table 2.1). Finally, time spent waiting associated with the acquiring of goods and services averages 0.67 minute a day. The employed and unemployed spent about the same amount of time waiting, while individuals not in the labor force wait twice as long, or 1.3 minutes.

Individuals spend on average 8 minutes a day purchasing groceries, gas, and food, the bulk of which is spent purchasing groceries (6 minutes), and an average of 18 minutes a day in travel associated with purchasing goods and services.

The last columns of Table 2.1 report the time spent shopping for goods and services for men and women, each according to marital status, as well as for individuals age 55 and older. Women spend 15 more minutes a day shopping for goods and services relative to men. The largest difference is in the shopping for goods and services category. Conditioning on marital status reveals little difference in the pattern of time spent searching for and acquiring goods and services across married and single men. Married women spend 7 more minutes a day in total shopping time than single women. Time spent shopping for groceries show the most important differences across married and single women, along with travel time associated with shopping for goods and services. The last column reports that individuals over age 55 spend, on average, 3 more minutes a days shopping than individuals aged 25 to 54.

## 2.5 Shopping time and the business cycle

The ATUS does not cover an entire business cycle at the moment. This renders the discussion of trends and cycle in time use data delicate. We address this question in the first subsection by comparing average time spent shopping for goods and services by individuals in the three years leading up to and three years following the start of the Great Recession. Next, we estimate the elasticity of time spent on different goods market search categories over the business cycle in the manner of Aguiar et al (2013). That is, we exploit cross-state variations in time spent and a measure of state business cycles. We then look at the relationship between individual and household income and our measure of shopping time, as well as changes in hours of market work. In each section we emphasize differences across labor market statuses.

### 2.5.1 The cycle and goods market search

A first step in examining the cyclicity of search time in the goods market, presented in Table 2.2, is to compare the average time spent on our different subcategories over two time periods. This also has the advantage, as argued by Aguiar et al. (2013), of smoothing some year-to-year noise in the ATUS.<sup>8</sup> The periods we compared are the expansion years of 2005-2007 prior to the onset of the Great Recession in December 2007, to the following three years, 2008-2010. The sample in Table 2.2 Panel A are all individuals in the age category 25 to 54, our baseline. Panel B reports the results for employed individuals, while Panels A and B of Table 2.3, respectively, report the results for unemployed individuals and persons not in the labor force.

The first two columns of Table 2.2 present the average time spent in each category for the two periods. The third column presents the unconditional difference in time spent over the two periods. The last column presents the difference in time spent conditioning on age, education, race, gender, marital status, and the presence of children.

Overall time spent shopping saw a statistically significant decline of 2.5 minutes per day in

---

<sup>8</sup>It does not appear appropriate to remove a time trend from the data when the sample does not cover an entire business cycle. Such a transformation will introduce a bias in the direction of the differences that reflects the fraction of the cycle actually covered by the said sample, and on the state of the cycle at the start of the sample. The next section will use cross-state changes in GDP and time use in order to further investigate the cyclicity of shopping time.

**Table 2.2** Shopping Time by Period - Minutes per Day

Panel A: Full sample				
Categories			Difference	
	Average	Average	2008-2010 vs 2005-2007	
	2005-2007	2008-2010	Unconditional	Conditional
	(1)	(2)	(3)	(4)
Total Shopping Time	43.1	40.7	-2.49 **	-2.44 **
Consumer Goods and Services	17.1	15.1	-1.97 ***	-1.92 ***
-Consumer goods	16.4	14.4	-2.06 ***	-2.01 ***
-Researching G&S	0.07	0.11	0.04	0.04
-Waiting for services	0.63	0.67	0.05	0.04
Groceries, Gas and Food	7.83	7.84	0.01	0.03
- Groceries	6.17	6.13	-0.04	-0.01
- Gas	0.38	0.46	0.08 ***	0.08 ***
- Food	1.28	1.25	-0.03	-0.04
Travel time	18.2	17.7	-0.53	-0.55

Panel B: Employed Individuals				
Categories			Difference	
	Average	Average	2008-2010 vs 2005-2007	
	2005-2007	2008-2010	Unconditional	Conditional
	(1)	(2)	(3)	(4)
Total Shopping Time	40.9	39.1	-1.78 *	-1.87 *
Consumer Goods and Services	15.9	14.3	-1.68 ***	-1.68 ***
-Consumer goods	15.3	13.7	-1.66 ***	-1.65 ***
-Researching G&S	0.06	0.06	-0.004	-0.002
-Waiting for services	0.56	0.52	-0.02	-0.02
Groceries, Gas and Food	7.14	7.42	0.28	0.27
- Groceries	5.42	5.6	0.18	0.18
- Gas	0.39	0.48	0.09 **	0.09 ***
- Food	1.32	1.34	0.02	0.005
Travel time	17.8	17.4	-0.38	-0.46

Notes: ATUS using individual demographic and time use information. Sample restricted to respondents of age 25 to 54 (Panel A), and employed (Panel B), and have zero amount of unclassified time. Column 1 and 2 report average minutes per day spent on various shopping activities during 2005-2007 and 2008-2010. Column 3 reports the unconditional difference and Column 4 reports the conditional difference in a regression controlling for age, education, race, gender, marriage status, and number of children. The asterisks \*, \*\* and \*\*\* denote statistical significance at 10%, 5% and 1% levels.

**Table 2.3** Shopping Time by Period - Minutes per Day

## Panel A: Unemployed Individuals

Categories	Difference			
	Average	Average	2008-2010 vs 2005-2007	
	2005-2007	2008-2010	Unconditional	Conditional
	(1)	(2)	(3)	(4)
Total Shopping Time	52.7	44.4	-8.34	-9.01 *
Consumer Goods and Services	24.9	14.9	-9.98 ***	-10.3 ***
-Consumer goods	24.5	14	-10.5 ***	-10.9 ***
-Researching G&S	0.05	0	-0.05	-0.05
-Waiting for services	0.27	0.86	0.58	0.64
Groceries, Gas and Food	9.59	11.3	1.67	1.6
- Groceries	8.1	10	1.94	1.88
- Gas	0.27	0.49	0.21	0.22
- Food	1.22	0.73	-0.49 *	-0.5 *
Travel time	18.3	18.2	-0.03	-0.28

## Panel B: Individuals Not in the Labor Force

Categories	Difference			
	Average	Average	2008-2010 vs 2005-2007	
	2005-2007	2008-2010	Unconditional	Conditional
	(1)	(2)	(3)	(4)
Total Shopping Time	54	48.3	-5.67 *	-5.09 *
Consumer Goods and Services	22	20.3	-1.7	-1.45
-Consumer goods	20.6	18.3	-2.32	-1.89
-Researching G&S	0.14	0.5	0.36	0.28
-Waiting for services	1.25	1.52	0.27	0.17
Groceries, Gas and Food	11.3	8.99	-2.3 ***	-2.23 **
- Groceries	9.96	7.75	-2.21 **	-2.15 **
- Gas	0.33	0.35	0.03	0.03
- Food	1.01	0.89	-0.11	-0.11
Travel time	20.7	19	-1.67	-1.41

Notes: A TUS using individual demographic and time use information. Sample restricted to respondents of age 25 to 54 who are unemployed (Panel A) or not in the labor force (Panel B), and have zero amount of unclassified time. Column 1 and 2 report average minutes per day spent on various shopping activities during 2005-2007 and 2008-2010. Column 3 reports the unconditional difference and Column 4 reports the conditional difference in a regression controlling for age, education, race, gender, marriage status, and number of children. The asterisks \*, \*\* and \*\*\* denote statistical significance at 10%, 5% and 1% levels.



2008-2010 compared with 2005-2007. A closer examination reveals that the decline is concentrated in the consumer goods and services category. None of the increases in time spent shopping is statistically significant, with the exception of time spent purchasing gas. Presumably individuals are willing to commute greater distances to work at scarce jobs. Moreover, the conditional difference in time spent is almost identical to the unconditional difference.

We find that the largest decline in time spent shopping for goods and services is by the unemployed. Time spent shopping for consumer goods and services declines by 10 minutes for the unemployed, compared with 1.7 minutes for the employed. Both declines are highly statistically significant. This decline is slightly larger for the unemployed after controlling for other individual characteristics. The corresponding time use categories for nonparticipants also declines, but the difference is not statistically significant.

This finding relates to the composition of individuals across labor market statuses and aggregate search effort in the goods market. Prior to the onset of the Great Recession, the unemployed spent more time searching for goods and services than the employed: 24.5 minutes against 15.3 by the employed. As such, the rapid increase in unemployment in 2008 and 2009 could have led to an increase in aggregate search effort in the goods market. The ATUS data show that this was not the case. Shopping time of the unemployed declined to an average of 14 minutes in the 2008-2010 period, below the average of 13.7 for the employed. Since shopping time declined for every labor force status, aggregate search time declined (Panel A of Table 2.2). Again, the search effort of the unemployed declined most during the recession. By the end of the sample, the unemployed spent less time than the employed searching in the goods market.

Average time spent shopping for groceries, gas, and food, just as travel time associated with shopping, saw essentially no change across time periods. This is true across labor force statuses, with one exception. There is a 2.3 minute decline in time spent shopping for groceries by individuals not in the labor force.

### 2.5.2 Cross-state variations in shopping time

We define the state-level aggregates of time use of category  $j$  as follows:

$$\tau_{st}^j = \sum_{i=1}^{N_{st}} \left( \frac{\omega_{ist}}{\sum_{i=1}^{N_{st}} \omega_{ist}} \right) \tau_{ist}^j, \quad (2.7)$$

where  $\tau_{ist}^j$  represents minutes per day by individual  $i$  from state  $s$  during period  $t$  spent on time use category  $j$ .  $N_{st}$  is the total number of individuals in our sample from state  $s$  during period  $t$  and  $\omega_{ist}$  is the ATUS sampling weight.<sup>9</sup> We then construct state-level differences in shopping between the 2005-2007 period and 2008-2010 period, along with corresponding changes in state real GDP per capita. Figure 2.2 plots the log change in state GDP per capita against the log change in different categories of shopping time in the corresponding state. Panel A corresponds to total shopping time, panel B corresponds to time spent shopping for goods and services, panel C for groceries, gas and food, and panel D corresponds to the travel time associated with shopping. Each panel also plots a cross-state regression line using state population as regression weights.

The pattern in panel A of Figure 2.2 is clear. States with the strongest decline in GDP per capita experienced the most significant decline in total shopping time. Likewise, states that saw an increase in GDP per capita also saw an increase in total shopping. In Michigan for instance, there was a 21% decline in time spent and a 10% decline in GDP per capita. Oklahoma, with a very different experience over the period in question, experienced a 2% increase in GDP per capita and an 20% increase in shopping time. The regression line has a positive slope with coefficient 1.09, implying that, on average, a 1% decline in state GDP per capita coincided with a 1.1% decline in total time spent shopping.

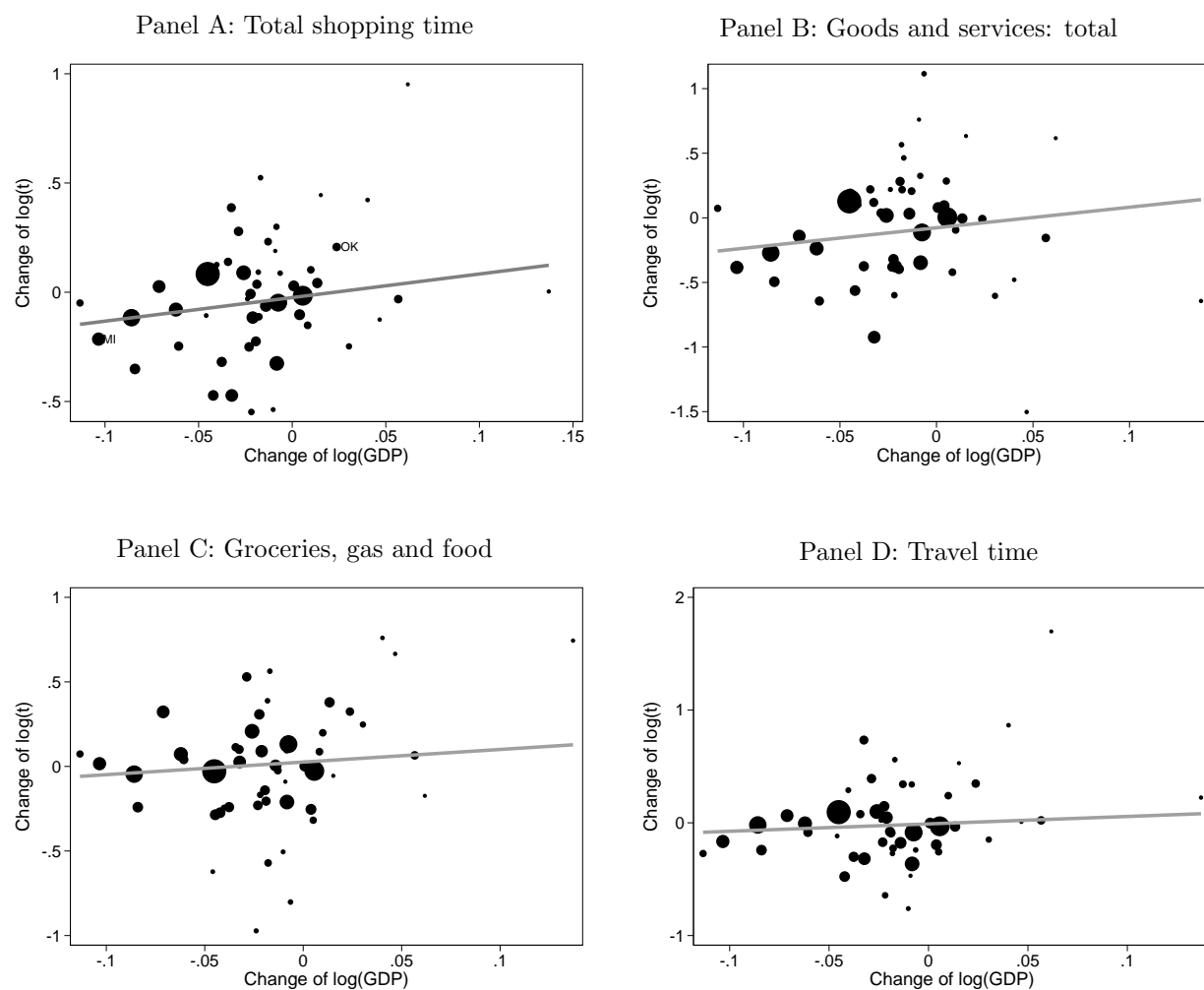
Time spent shopping for goods and services, in Panel B, shows a similar pattern of decline with the contraction of state GDP. The regression line has a slope of 1.60, implying a stronger positive relation than in the case of total shopping time.

The state-level changes in time spent shopping for groceries, gas, and food or travel time

---

<sup>9</sup>Unlike the CPS, which is designed to produce reliable estimates at both the state and national level, the ATUS only has a national reliability requirement. Less populous states constitute a smaller proportion of the ATUS sample and will not produce estimates as reliable as for the more populous states. We use the average state population between 2003 and 2012 as weights in the regressions to account for this concern regarding the state-level estimates.

**Figure 2.2** Cross-State variations in shopping time: 2008-2010 vs. 2005-2007



associated with shopping have a weaker positive relation to changes in state GDP per capita, but neither is significant. Most states, especially the most populous, saw virtually no change in the time spent shopping for groceries, gas, and food (see Panel C).

## 2.6 Individual regressions

This section examines the relationships between two measures of income, household and individual, and search activity in the goods market in the ATUS. In the first case we use reported household income brackets. In the second case we use reported weekly earnings in the CPS files. Both approaches uncover a positive relation between income and search effort in the goods market. The results are strongest for shopping for goods and services and time spent traveling associated with shopping. We find no evidence in the time use data that individuals with lower incomes search far and wide.

### 2.6.1 Household income

We consider the relation between household income and consumer search in the goods market using six income categories: (1) \$0 to \$24,999; (2) \$25,000 to \$49,999; (3) \$50,000 to \$74,999; (4) \$75,000 to \$99,999; (5) \$100,000 to \$149,999, and; (6) \$150,000 and over, and running the following regression:<sup>10</sup>

$$\tau_{ist}^j = \alpha^j + \beta^j F_{ist} + D_t + S_i + \delta X_{ist} + \epsilon_{ist}, \quad (2.8)$$

where  $F_{ist}$  is a vector of family income categorical variables,  $D_t$  is a time dummy,  $S_i$  is a state dummy, and  $X_{ist}$  is a vector of demographic and educational variables.

Table 2.4 reports the estimated coefficients on income categories 2 to 6. Each coefficient represents the increment in shopping time for an increment in the household income categories relative to the first household income category \$0-\$24,999. For instance, total shopping time for households with an income between \$100,000 and \$150,000 (category 5) is 5 to 7 minutes greater than average shopping time by an individual in a household with income in the \$0 to \$24,999

---

<sup>10</sup>The ATUS provides two family income variables – hufaminc for 2002-2009 and hefaminc for 2010-2012 – with the distinction that the second variable has been edited to leave no empty responses.

**Table 2.4** Individual level regression of time use on household income

Category	HH. Income	I	II	III	
Total Shopping Time	2	1.53	1.52	1.41	3.00 ***
	3	1.34	1.42	1.37	3.30 ***
	4	4.01 ***	3.97 ***	3.62 ***	5.73 ***
	5	4.36 ***	4.99 ***	4.45 ***	6.59 ***
	6	4.86 ***	5.69 ***	4.90 ***	6.90 ***
Consumer Goods and Servies	2	0.86	0.84	0.82	1.52 ***
	3	0.41	0.43	0.47	1.32 **
	4	1.85 ***	1.79 ***	1.76 **	2.69 ***
	5	2.75 ***	3.06 ***	2.93 ***	3.86 ***
	6	1.44	1.86 **	1.65 *	2.52 ***
Groceries, Gas and Food	2	-0.02	0.00	-0.04	0.45 *
	3	-0.43	-0.41	-0.45	0.14
	4	-0.34	-0.31	-0.45	0.20
	5	-0.35	-0.30	-0.51	0.15
	6	-0.32	-0.26	-0.54	0.09
Travel time	2	0.69	0.68	0.63	1.03 **
	3	1.36 ***	1.40 ***	1.35 ***	1.84 ***
	4	2.51 ***	2.49 ***	2.31 ***	2.84 ***
	5	1.96 ***	2.23 ***	2.03 ***	2.58 ***
	6	3.73 ***	4.09 ***	3.79 ***	4.29 ***
Demo. controls		Yes	Yes	Yes	Yes
Time dummy		No	Yes	Yes	Yes
State dummy		No	No	Yes	Yes
LFS dummy		No	No	No	Yes

There are six income categories: i) \$0 to \$24,999; ii) \$25,000 to \$49,999; iii) "\$50,000 to \$74,999; iv) \$75,000 to \$99,999; v) \$100,000 to \$149,999, and; vi) \$150,000 and over. The table reports the estimated coefficients on income categories 2 - 6. The asterisks denote the statistical significance: \* = 10%, \*\* = 5% and \*\*\* = 1%. The demographic controls are for age, education, gender, marital status, race and number of children.

bracket (category 1). After controlling for various individual characteristics (column I), this is robust to including both time and state (columns II and III), as well as labor force status dummy variables (column IV).

The three broad subcategories show that both time spent shopping for consumer goods and services, as well as the associated travel time, are increasing in household income, and the differences are highly statistically significant. For instance, households in the \$50,000 to \$75,000 income bracket (category 3) spend an extra 1.3 minutes shopping for goods and services and an extra 1.8 minutes waiting. Households in the \$100,000 to \$150,000 income bracket (category 5) spend an extra 3.9 minutes shopping for goods and services and an extra 2.6 minutes waiting. There appears not to be any statistically significant relationship between time spent shopping for groceries, gas, and food and household income.

### 2.6.2 Individual income

Reported weekly income for employed respondents averages \$37,856 per year. We use this information to estimate the following relation between income and search effort in the data:

$$\tau_{ist}^j = \alpha^j + \beta^j I_{ist} + D_t + S_i + \delta X_{ist} + \epsilon_{ist}, \quad (2.9)$$

where  $j$  is the time use category,  $I_{ist}$  is annual personal income (in thousands of dollars) for individual  $i$  in state  $s$ ,  $D_t$  is a time dummy,  $S_i$  a dummy for the individual's state, and  $X_{ist}$  is a vector of demographic and educational variables (age, gender, race, education, marital status, and presence of children). The coefficient of interest,  $\beta^j$  is reported in Table 2.5, with standard errors in parentheses, for total shopping time and the three broad subcategories: consumer goods and services; groceries, gas, and food; and travel time.

The first row of Table 2.5 reveals that total shopping time is increasing in individual income but that none of the coefficients are statistically significant. The second row indicates that time spent shopping for consumer goods and services is increasing in individual income, yet none of the specifications yield a statistically significant coefficient. Time shopping for groceries, gas, and food, however, declines with income, and the coefficient is highly significant in each of the

**Table 2.5** Individual level regression of time use on personal income

Category	I	II	III
Total Shopping Time	0.75 ( 0.55 )	1.34 ( 0.28 )	0.64 ( 0.61 )
Consumer Goods and Services	0.40 ( 0.60 )	0.69 ( 0.37 )	0.41 ( 0.60 )
Groceries, Gas and Food	-1.00 ( 0.00 )	-0.99 ( 0.00 )	-1.23 ( 0.00 )
Travel time	1.35 ( 0.03 )	1.65 ( 0.01 )	1.46 ( 0.02 )
Demo. controls	Yes	Yes	Yes
Time dummy	No	Yes	Yes
State dummy	No	No	Yes

The regression sample is restricted to respondents who are between age 25 and 54, have zero unclassified time and positive personal income (employed). The table reports the estimate of  $\beta$  and its p-value in parentheses. Personal incomes are in thousands of dollars. Estimates of  $\beta$  are multiplied by 100.

specifications. Last, we find that travel time associated with shopping is increasing with individual income, and the coefficient is highly statistically significant in each specification. This suggests that individuals with lower income do not travel further and search wider for goods and services.

### 2.6.3 Market work and shopping time

In this section we examine the relation of shopping time to time in market work. We apply the same definition of market work in the ATUS as Aguiar et al. (2013) and run the following regression:<sup>11</sup>

$$\tau_{ist}^j = \alpha^j + \beta^j \tau_{ist}^M + D_t + S_i + \delta X_{ist} + \epsilon_{ist}, \quad (2.10)$$

where  $\tau_{ist}^M$  is the time spent on market work,  $D_t$  and  $S_i$  are the time and state dummies, respectively, and  $X_{ist}$  is a vector of demographic and educational variables (age, gender, race, education, marital status, and the presence of children). Table 2.6 reports the resulting estimates  $\beta^j$ .

The results indicate that a 1 minute decline in market hours is associated with a 0.04

<sup>11</sup>The ATUS time use categories used to measure market hours of work are: 05-01, 05-02, 05-99, 18-05.

**Table 2.6** Individual level regression of shopping time on market work time

Category	I	II	III
Total Shopping Time	-8.17 ( 0.00 )	-8.17 ( 0.00 )	-8.18 ( 0.00 )
Consumer Goods and Services	-3.85 ( 0.00 )	-3.85 ( 0.00 )	-3.86 ( 0.00 )
Groceries, Gas and Food	-1.24 ( 0.00 )	-1.24 ( 0.00 )	-1.24 ( 0.00 )
Travel time	-3.08 ( 0.00 )	-3.08 ( 0.00 )	-3.08 ( 0.00 )
Demo. controls	Yes	Yes	Yes
Time dummy	No	Yes	Yes
State dummy	No	No	Yes

The table reports the estimate of and its p-value in parentheses. Market work is in terms of minutes per day. Estimates of are multiplied by 100 for readability. An estimate of 1 means that the time use is expected to be 0.01 minutes per day if market work is 1 minute higher.

minute decline in shopping time. The impact of diminished hours of the employed on shopping time is very small.

## 2.7 Conclusion

The recent availability of new data sets measuring the inflows and outflows of goods and services in a household's consumption basket (Broda and Weinstein, 2010) and the presence and investment in customer relationships (Gourio and Rudanko, 2013) has lead to a renewal of theories in which search frictions in the goods market play an important role for macroeconomic outcomes. The cyclical properties of the intensive margins in this market, by consumers and firms, have quickly shown themselves to be important to our deeper understanding of dynamics of consumption, employment, and business cycles in general.

A precursor paper by Hall (2012) had shown that advertisement by firms is very procyclical. In this paper we find, on the other side of the market, that consumers spend a varying amount of time for purchasing goods and products. This shopping time declined significantly with the onset of the Great Recession across all types of individuals, and it is positively correlated with



individual and household income. In addition, consumption effort dropped more in states where economic activity decreased relatively more following the financial crisis. We also find that a decline in working time raises shopping time. Nonetheless, employed individuals spent less time shopping during the recession.

Overall, we don't find much evidence in favor of a negative correlation between income and shopping time, particularly comparison shopping possibly motivated by locating better prices. As a matter of fact, whereas total time spent purchasing goods and services is about 20 minutes per day, the component of shopping time devoted to comparing prices and products seems to be extremely low in the data. It is about 4 seconds a day on average, given the large number of respondents declaring zero. Moreover, nonparticipants and the employed spent more time than the unemployed in this activity. This may indicate that most of the effort made by consumers in the goods market is unrelated to uncovering better prices.

The data used in this paper are the best available from a macro perspective. More precise information about consumer shopping efforts should be obtained from microeconomic data such as consumer surveys. This is left for future works.

## References

- Aguiar**, M, Hurst, E. and Karabarbounis, L. (2013). Time Use During the Great Recession. *American Economic Review* 103(5), 1664–1696.
- Bai**, Y., Rios-Rull, V., and Storesletten, K. (2011). Demand shocks that look like productivity shocks. Mimeo University of Minnesota.
- Broda**, C. and Weinstein, D. E. (2010). Product creation and destruction: Evidence and price implications. *American Economic Review* 100(3), 691–723.
- den Haan**, W. (2013). Inventories and the role of goods-market frictions for business cycles. mimeo London School of Economics.
- Diamond**, P. A. (1982). Aggregate demand management in search equilibrium. *Journal of Political Economy*, 90(5):pp. 881–894.
- Gourio**, F. and Rudanko, L. (2013). Customer capital. *Review of Economic Studies*, forthcoming.
- Hall**, R. E. (2012). The cyclical response of advertising refutes counter-cyclical profit margins in favor of product-market frictions. NBER Working Papers 18370, National Bureau of Economic Research, Inc.
- Hamermesh**, D., Harley, F. and Stewart, J. 2005. Data Watch: The American Time Use Survey. *Journal of Economic Perspectives* 19 (1), 221–32
- Kaplan**, G. and Menzio, G. (2013). Shopping Externalities and Self-Fulfilling Unemployment Fluctuations. NBER Working Papers 18777, National Bureau of Economic Research, Inc.
- Michaillat**, P. and Saez, E. (2014). An Economical Business-Cycle Model. NBER Working Papers 19777, National Bureau of Economic Research, Inc.
- Petrosky-Nadeau**, N. and Wasmer, E. (2011). Macroeconomic Dynamics in a Model of Goods, Labor and Credit Market Frictions. IZA Discussion Papers 5763, Institute for the Study of Labor (IZA).
- Petrosky-Nadeau**, N. and Wasmer, E. (2014). Market Frictions and Fiscal Multipliers. Mimeo, Sciences Po, Paris.

# APPENDIX

## 2.A Theory

### 2.A.1 Individual consumption probability: a more general specification

At the individual level, one could assume that the probability of finding a particular good is given by  $\lambda e_c / (\bar{e}_c^{1-a})$ , in which  $a \geq 0$ . When  $a = 0$ , the most frequent assumption explored in the text, only the ratio of individual to aggregate effort matters for the individual finding rate. In contrast, assuming a positive value for  $a$  allows for a potential positive externality from consumer search effort. In essence, if my neighbors search for a good, they may transmit some information to me thereby increasing the probability of locating a particular good.

A key insight here is that allowing for a positive  $a$  does not change the results on cyclicity. Indeed, the aggregate level of effort does not change the cyclicity of individual search effort. In a symmetric equilibrium, we have  $d(\ln \bar{e}_c)/d\ln \Omega_c = \eta_c - a$ , where  $\eta_c$  is the elasticity of the search cost function with respect to effort, assumed larger than 1. Thus,  $d(\ln \bar{e}_c)/d\ln \Omega_c > 0$  under the sufficient condition that  $a < 1$  even for positive consumption externality  $a > 0$ .

### 2.A.2 Non-separable utility

One could write the utility function as:

$$v(x, y) = X(x) + Y(y) + bZ(x, y) \tag{2.11}$$

where  $X$  and  $Y$  are increasing concave functions of their input, and  $Z$  is a complementary function with  $\partial^2 Z / (\partial y \partial x) > 0$  and  $Z(0, y) = Z(x, 0) = 0$ . This reflects the interaction between  $x$  and  $y$ .  $b$  is a scalar that may be zero in the case of perfect separability, and positive (negative) depending on the degree of complementarity (substitutability) between the two consumption goods. The

important property for the main discussion on consumer search effort is the sign of

$$\Delta v_y = [Y_{yy}(\omega - px) - Y_{yy}(\omega)] + b \frac{\partial^2 Z}{\partial y \partial x}(x, \omega - px)$$

The first term in brackets is positive due to the concavity of utility in goods  $y$ . If the second term is assumed to be positive, e.g., if  $Z(x, y) = xy$ , it is sufficient that the value of  $b$  is not too negative, e.g.  $Z(x, y) = xy$ , it suffices that  $b$  is larger than the negative of  $Y_{yy}(\omega - px) - Y_{yy}(\omega)$  to preserve a positive sign of  $\Delta v_y$ .

### 2.A.3 Price setting - Nash bargaining:

The results are similar under a formal Nash-sharing rule with the exception that the marginal utility of the composite good  $y$  now affects the bargained price. Assuming equal bargaining weights, the price is set to maximize the Nash-product  $[\Delta v]^{1/2} [px - \kappa(x)]^{1/2}$ , and must satisfy a sharing rule similar to equation (2.5) but with the marginal utility of goods  $y$  appearing in the denominator of the consumption surplus  $\Delta v$ :

$$xp = \Delta v / v_y(\mathbf{1}) + \kappa(x) \quad (2.12)$$

A positive, procyclical, shock to income  $\omega$ ,  $d\omega$ , keeping the marginal cost  $\kappa$  constant, leads to (see Appendix for the derivation):

$$x \frac{dp}{d\omega} = \frac{\Delta v_y v_y(\mathbf{1}) - v_{yy}(\mathbf{1}) \Delta v}{2v_y^2(\mathbf{1}) - v_{yy}(\mathbf{1}) \Delta v}$$

with the same conclusion as we found previously by the fact that  $v_{yy} < 0$ . The effect of an income shock on the price and surplus remains positive. It is still the case that consumer search effort is procyclical.

#### 2.A.4 Endogenous supply of the search good $x$

On the other side of the market, a seller has a cost of providing the search good  $\kappa(x)$ , with  $\kappa_x, \kappa_{xx} > 0$ . The seller's surplus to the match is

$$\Omega_f = px - \kappa(x) \quad (2.13)$$

Fixing  $e_1$ , we now assume that, at the time of the match, sellers can react to the demand of matched consumers by varying instantaneously their production of  $x$  at cost  $\kappa(x)$ . Different pricing and bargaining assumptions may be introduced here.

#### Nash-bargaining on both price $p$ and quantity $x$

Assume that the seller and the consumer can bargain over two dimensions:  $p$  and  $x$ . The Nash problem is:

$$p, x = \operatorname{argmax} [v(x, \omega - px) - v(0, \omega)]^{1/2} [px - \kappa(x)]^{1/2}$$

The maximization of the Nash-product with respect to the price leads to a pricing rule identical to equation (2.12). The maximization of the Nash-product with respect to the quantity produced leads to

$$\frac{\Delta v}{v_x(\mathbf{1}) - pv_y(\mathbf{1})} + \frac{xp - \kappa(x)}{p - \kappa_x} = 0 \quad (2.14)$$

Combining (2.12) and (2.14) leads to the equality between the marginal production cost and the marginal utility of the good divided by its marginal utility cost:

$$\kappa_x(x) = \frac{v_x(\mathbf{1})}{v_y(\mathbf{1})} \quad (2.15)$$

Again, the key for the elasticity of consumer search effort to an income shock is the response of the surplus  $\Omega_c = \Delta v$ . This can be decomposed by differentiating equation (2.2) into

income, quantity, and price effects:

$$\frac{d\Omega_c}{d\omega} = \frac{d\Delta v}{d\omega} = \underbrace{\Delta v_y}_{\text{Change in income: } > 0} - \underbrace{xv_y(\mathbf{1}) \frac{dp}{d\omega}}_{\text{Change in price}} + \underbrace{[v_x(\mathbf{1}) - pv_y(\mathbf{1})] \frac{dx}{d\omega}}_{\text{Change in quantity}} \quad (2.16)$$

The direction of search effort with endogenous price and quantity is ambiguous. The income effect raises the surplus but, if the price increases with income, this will reduce the surplus. Moreover, the quantity effect may increase or decrease the surplus depending on the sign of the bracketed term, which is positive if the marginal utility from consumption 1 (being able to consume the search good) exceeds its opportunity cost  $pv_y$ . One may presume that the production of the search good in a match is increasing in income  $\omega$  and decreasing in the price  $p$ , while the price increases with income. In that case, under small marginal utility of the composite good  $y$  (small  $v_y$ ) and large marginal utility from the search good (large  $v_x$ ), the positive effect would dominate.

### More specific cases

Some additional results can be obtained in specific cases. First, if utility is linear in  $y$ , then both  $x$  and  $p$  become independent of income  $\omega$ . Therefore the consumer's surplus and search effort in the goods market become independent of the business cycle.<sup>12</sup>

Second, a similar result holds under the alternative assumption of an exhaustion of profit margins. This would be the case if the market price is such that  $p = \kappa(x)$ . The firm's surplus vanishes to zero. If in addition workers and firms bargain on the quantity, the consumption surplus will also vanish. Hence, search effort will not react to the cycle due to the surplus. Consumers may however respond to further entry of firms. An increase in  $S$  raises the contact rate  $\lambda$ , implying once again that search effort will be procyclical.

Third, if one assumes that the quantity  $x$  is chosen so as to share the surplus equally, while the price is equal to the marginal cost of production with a markup to account for ex post monopoly power, again a large utility from the search good and small utility from the composite leads to a positive effect of income on the surplus and search effort.<sup>13</sup>

---

<sup>12</sup>This is shown in Appendix 2.A.7. However, this result does not hold in the general case derived in Appendix 2.A.7.

<sup>13</sup>See Appendix 2.A.7 for details.

### 2.A.5 Price posting with homogeneous consumers and a given distribution of prices

Assume that the search good has different prices in different locations. Consumer search effort is now also motivated by finding better prices. Let  $G(p)$  be the c.d.f. of prices. The program of a searching, unmatched consumer with income  $\omega$  and search effort cost  $\sigma(e_c)$  can be represented by the Bellman equation  $D_U(e_c)$ , while  $D_M(p)$  describes the asset value of being matched at price  $p$ :

$$rD_U(e_c) = v(0, \omega) - \sigma(e_c) + \frac{e_c}{\bar{e}_c} \lambda \int^{p_R} (D_M(p) - D_U(e_c)) dG(p) \quad (2.17)$$

$$rD_M(p) = v(x, \omega - px) + \tau [D_U(e_c) - D_M(p)] \quad (2.18)$$

where  $p_R$  is the reservation price above which the consumer keeps searching,  $\tau$  is an exogenous destruction of the consumption match and  $r$  is a constant discount rate.

The value of searching in the goods market,  $D_U$ , is maximized by varying the level of effort. The reservation price  $p_R$  satisfies an indifference condition  $rD_M(p_R) = rD_U$ . In a symmetric equilibrium where  $e_c = \bar{e}_c$ , optimal search effort and reservation price are given by the conditions:

$$\eta_c \sigma_c(e_c^*) = \lambda \int^{p_R} (D_M(p) - D_U(e_c^*)) dG(p) \quad (2.19)$$

$$\sigma(e_c^*) + v(x, \omega - p_R x) - v(0, \omega) = \lambda \int^{p_R} (D_M(p) - D_U(e_c^*)) dG(p) \quad (2.20)$$

which combined yield a relation between search effort, income, and the reservation price:

$$(\eta_c - 1) \sigma_c(e_c^*) = v(x_1, \omega - p_R x) - v(0, \omega) \quad (2.21)$$

For a given level of the reservation price, the right-hand side is procyclical under concavity of the utility function with respect to income. This is the procyclical channel of income on search effort outlined in earlier sections. The key countervailing force is the movement in the reservation price. If  $p_R$  varies procyclically, then the second term has a countercyclical component. That is, an increase in the reservation price leads to a decline in the expected payoff to search in the goods market, and reduces the incentive to exert effort in searching. Variations in the reservation price must therefore

be studied jointly with variations in effort. Differentiating equation (2.21), we have:

$$(\eta_c - 1)\sigma'_c(e_c^*)de_c^* = [v_y(x, \omega - p_R x) - v_y(0, \omega)]d\omega - x v_{yy}(x, \omega - p_R x)dp_R \quad (2.22)$$

It can be shown that  $dp_R/d\omega > 0$ . A reservation price increasing in income limits the positive response of consumer search effort to an income shock.

The link between the distribution of prices and search effort in the goods market is ambiguous. The mechanisms are similar to the relation between wage dispersion and job search. In the labor market, more dispersed wages lead workers, *ceteris paribus*, to be more picky about wage offers. This raises their reservation wage and reduces the so-called hazard rate (the exit rate from unemployment to employment). To compensate, workers may reduce their search effort if the lower hazard rate discourages them, or raise it if the surplus value of getting better jobs dominates. Transposed to search for products, price dispersion may reduce the reservation price  $p_R$  with an ambiguous effect on shopping effort.

### 2.A.6 Correlation between income, shopping time, and working time

So far, we have assumed that working time was indivisible and the supply of shopping time was elastic, only limited by the cost of supplying effort. We now relax this assumption and generate a setup that allows us to discuss the findings of Aguiar et al. (2013) and how they differ from ours.

Let  $h$  and  $e_c$  be the two relevant time inputs, working time and shopping effort, respectively. We remain close to the previous analysis, ignoring other nonworking time, and consider an environment similar to Section 2.2.2 to determine a consumer's optimal choices. We assume that  $e_c$  and  $h$  enter additively in the cost of effort function, and that the choice of hours is made before the consumer is actually matched with a good. The consumer knows the matching probability in the goods market  $m(e_c) = e_c\lambda/\bar{e}_c$  and the separation probability in the labor market (denoted by  $s$ ) so that (s)he expects to be matched  $\alpha(e_c) = m/(m + s)$  of the time.<sup>14</sup> The first-order conditions

---

<sup>14</sup>The special case where agents choose hours worked while they search for goods corresponds to the case  $\alpha(e_c) = 0$ . The special case where agents choose hours worked while they consume corresponds instead to  $\alpha(e_c) = 1$ .



for  $e_c$  and  $h$ , respectively, are given by:

$$\sigma'(e_c + h) = \frac{\lambda}{\bar{e}_c} \Omega_c(h) \quad (2.23)$$

$$\sigma'(e_c + h) = \alpha(e_c) \frac{dv}{dh}(x, h\omega - p(h)x) + (1 - \alpha(e_c)) \frac{dv}{dh}(0, h\omega) \quad (2.24)$$

where  $\Omega_c(h) = v(x, h\omega - px) - v(0, h\omega)$ . Inspection of equation (2.23) shows that an increase in  $h$  increases the cost of an extra unit of shopping time. However, it also affects the right-hand side. In particular, it raises  $\Omega_c$  by  $\omega [v_y(x, h\omega - px) - v_y(0, h\omega)]$  which, under the same separability and concavity assumptions as in the benchmark case of Section 2.2.2, is positive. Hence, we expect an increase in hours worked to decrease shopping time and, reciprocally, a drop in hours worked to raise shopping effort. The response of search effort in the goods market to an income shock, however, is ambiguous and explored empirically in the rest of the paper.

## 2.A.7 Supplemental Appendix to Section 2.A.4: Optimal consumption choice when $x$ is endogenous

### Simpler case of $y$ as a numeraire

Start from equations (2.12) and (2.14). Under separability and constant marginal utility for the composite (the case of a numeraire) these equations simplify to:

$$xp = \kappa(x) + \Delta v \quad (2.25)$$

$$\frac{v_x(\mathbf{1}) - p}{\Delta v} = \frac{p - \kappa_x}{xp - \kappa(x)} \quad (2.26)$$

where  $v(\mathbf{1}) = X(x) + (\omega - px)$  and  $\Delta v = X(x) - px$ . This implies that  $\Delta v$  is independent of income, and therefore replacing  $xp - \kappa(x)$  in the equation (2.25):

$$xp = \frac{xX_x(x) + x\kappa_x}{2} = \frac{X(x) + \kappa(x)}{2}$$

This last equality implies that  $x$  is such that the elasticity of the function  $g(x) \equiv X(x) + \kappa(x)$  is 1. It follows that both the price and quantity of  $x$  are independent of income, and so is the surplus.

### General case under separability of $x$ and $y$

The full system represented by equations (2.12) and (2.14) must be differentiated to obtain the variations of  $p$  and  $x$  in response to a change in  $\omega$ . Before doing so, one can rewrite the utility function  $v(x, y) = X(x) + Y(y)$  and manipulate the two equations to obtain:

$$xp = \kappa(x) + \frac{\Delta v}{Y_y(\omega - px)} \quad (2.27)$$

$$X_x(x) = \kappa(x)Y_y(\omega - px) \quad (2.28)$$

$$\Delta v = X(x) - Y(\omega - px) + Y(\omega) \quad (2.29)$$

The second equation can be differentiated to obtain:

$$Y_y\kappa_{xx}dx + \kappa_x Y_{yy}(d\omega - d(px)) = X_{xx}dx$$

The functions are evaluated at consumption point **1**. That is,  $Y$  and its derivatives are evaluated at  $\omega - px$ , and  $X$  and  $\kappa$  and their derivatives are evaluated at  $x$ . This equation defines a linear relation between  $dp$ ,  $dx$ , and  $d\omega$ . The first equation can also be differentiated to provide another linear relation between these variations:

$$[d(xp) - \kappa_x dx] Y_y + (xp - \kappa(x)) Y_{yy}(d\omega - d(px)) = d\Delta v$$

where  $d\Delta v$  can be obtained as  $X_x(x)dx + Y_x(\omega - px)d(\omega - px) - Y_y(\omega)d\omega$ . In this general case, both  $x$  and  $p$  vary with income  $\omega$ .

### Markup pricing and bargaining over quantity

If one assumes that the quantity  $x$  is chosen so as to divide the surplus into equal shares, while the price is equal to the marginal cost of production with a markup to account for ex post monopoly power, a large utility from the search good and small utility from the composite lead to a positive effect of income on the surplus and search effort. Set the price to  $p = \kappa_x(1 + \epsilon)$ , such that the consumer's surplus is  $\Delta v = xp - \kappa(x) = x\kappa_x(1 + \epsilon) - \kappa(x)$ . Then, differentiation of this last equation

leads to:

$$\frac{dp}{dx} = \kappa_{xx}(1 + \epsilon) = (p/x)(\eta_c - 1) > 0$$

where  $\eta_c$  is the elasticity of  $\kappa$  with respect to  $x$ . Therefore, using equation (2.16), and replacing  $dp/dw$  with  $(dp/dx) \times (dx/dw)$ , and using the second part of the equality above, one has:

$$\begin{aligned} \frac{d\Omega_c}{d\omega} = \frac{d\Delta v}{d\omega} &= \underbrace{\Delta v_y}_{\text{Change in income: } > 0} - \underbrace{v_y(\mathbf{1})p(\eta_c - 1)}_{\text{Change in price}} \frac{dx}{d\omega} + \underbrace{[v_x(\mathbf{1}) - pv_y(\mathbf{1})]}_{\text{Change in quantity}} \frac{dx}{d\omega} \\ &= \Delta v_y + [v_x(\mathbf{1}) - v_y(\mathbf{1})p\eta_c] \frac{dx}{d\omega} \end{aligned}$$

Again, large utility from the search good and small utility from the composite lead to a positive effect of income on the surplus.

## 2.B Data

### 2.B.1 Shopping Time in the American Time Use Survey

The ATUS classifies diary activities into 18 major categories, and each category has additional second- and third-tier categories. According to the 2003-2012 multiyear lexicon, each activity has a code in the form of “x-y-z,” where “x” denotes the first-tier, “y” , the second tier, and “z,” the third tier classification. For example, major categories include personal care activities 01, household activities 02, work and work-related activities 05, education 06, etc. The second- and third-tier categories further break down the major categories. For example, under the major category of work and work- related activities, second-tier categories include working 05-01, working-related activities 05-02, other income-generating activities 05-03, job searching and interviewing 05-04, etc. And under the second-tier category of working, the third-tier categories are main job work 05-01-01, other job(s) work 05-01-02, security procedures related to work 05-01-03, etc.

We group the time spent shopping for market goods and services as follows:

Total time shopping for market goods and services:

1. Consumer Goods and Services shopping other than Groceries Gas and Food:

(a) Shopping for consumer goods:

- 07-01-04, shopping except groceries, food and gas.
- 07-01-05, waiting associated with shopping.
- 07-01-99, shopping, n.e.c.
- 07-99-99, consumer purchase, n.e.c.

(a) Researching goods and services:

- 07-02-01, comparison shopping.
- 07-02-99, researching purchases, n.e.c.

(a) Waiting associated with shopping for goods and services:

- 08-01-02, waiting associated with purchasing childcare services.
- 08-02-03, waiting associated with banking/financial services.
- 08-03-02, waiting associated with legal services.
- 08-04-03, waiting associated with medical services.
- 08-05-02, waiting associated with personal care services.
- 08-06-02, waiting associated with purchasing/selling real estate.
- 08-07-02, waiting associated with veterinary services.
- 09-01-04, waiting associated with using household services.
- 09-02-02, waiting associated with home maintenance/repair/decoration/construction.
- 09-03-02, waiting associated with pet services.
- 09-04-02, waiting associated with using lawn & garden services.
- 09-05-02, waiting associated with vehicle maintenance/repair services.
- 12-05-04: waiting associated with arts & entertainment.

2. Purchasing Groceries, Gas, and Food (GG&F):

- 07-01-01, grocery shopping.

- 07-01-02, purchasing gas.
- 07-01-03, purchasing food (not groceries).

### 3. Travel Time associated with shopping for goods and services:

- 18-07-01, travel related to grocery shopping.
- 18-07-82, travel related to shopping (except grocery shopping).
- 18-08-01, travel related to using childcare services.
- 18-08-02, travel related to using financial services and banking.
- 18-08-03, travel related to using legal services.
- 18-08-04, travel related to using medical services.
- 18-08-05, travel related to using personal care services.
- 18-08-06, travel related to using real estate services.
- 18-08-07, travel related to using veterinary services.
- 18-08-99, travel related to using professional & personal care services, n.e.c.
- 18-09-01, travel related to using household services.
- 18-09-02, travel related to using home maintenance/repair/decoration/construction services.
- 18-09-03, travel related to using pet services (not vet).
- 18-09-04, travel related to using lawn and garden services.
- 18-09-05, travel related to using vehicle maintenance & repair services.
- 18-09-99, travel related to using household services, n.e.c.
- 18-12-04, travel related to arts and entertainment.

## 2.B.2 Labor force status in the ATUS

The ATUS sample is a subset of the CPS sample. We verify how the estimates of labor force status – employed, unemployed, and nonparticipant – line up with the population estimates provided by the BLS based on the CPS. The five possible labor force statuses in the ATUS are: (1) employed

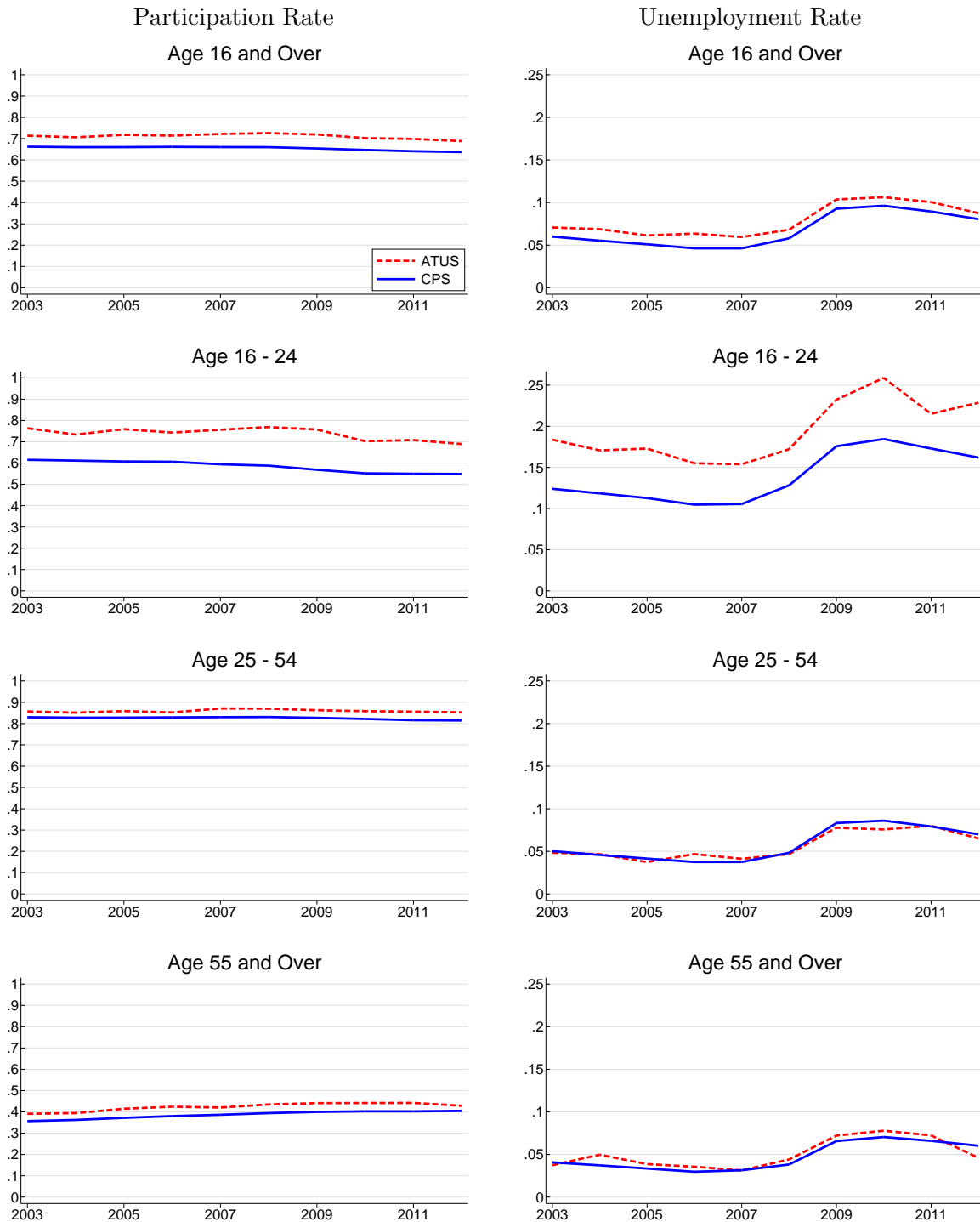
– at work; (2) employed – absent; (3) unemployed – on layoff; (4) unemployed – looking, and; (5) not in the labor force. Let  $D_i$  be the dummy variable for labor status, which equals one when the correspondent is in a given labor status and zero otherwise. The formula for calculating the percentage of the population being in one labor status is given by:

$$P = \frac{\sum_i I_i wt_i}{\sum_i wt_i} \quad (2.30)$$

where  $wt_i$  is the final weight assigned by ATUS. The formula can also be applied to a subsample of the same interview year and age group.

We compare the implied labor participation rate and unemployment rate from ATUS to those provided by the BLS. The ATUS labor participation rate is defined as the sum of employed and unemployed as percentages of population, and the unemployment rate is defined as the ratio of unemployed to the sum of employed and unemployed. The CPS participation rate is the ratio of the civilian labor force to the civilian non-institutional population, and the unemployment rate is the ratio of the unemployment level to the civilian labor force. Annual participation and unemployment rates from both ATUS and BLS are plotted in Figure 2.B.2. The ATUS participation rate for the age group 16 and over is 5.6% higher than the CPS rate, on average. A closer look and different age groups reveals that the difference is 15.2% for the 16 to 24 age category, 3.3% for ages 25 to 54, and 3.7% for ages 55 and over. The ATUS unemployment rate for ages 16 and over is 1.1% higher than the CPS rate, on average. Again, the difference is largest for the 16 to 24 group, 5.5%, while it negligible for those between 25 and 54, -0.1%, and for those over 55 is 0.4%.

**Figure 2.3** Participation and Unemployment Rates from ATUS and CPS



The solid line and dash line represent the CPS and ATUS respectively. The left column of figures compare the participation rates from ATUS and CPS, and the right column of figures compares the unemployment rates from ATUS and CPS. For ATUS, the participation rate is calculated as the sum of employed and unemployed as percentage of the population, the unemployment rate is the ratio of unemployed in percentage to the participation rate in percentage. For CPS, we calculate the participation rate as the ratio of civilian labor force to the civilian noninstitutional population, and unemployment rate as the unemployment level to the civilian labor force.

## Chapter 3

# Individual Occupation Switch: the Determinants and the Impact on Income

### 3.1 Introduction

The personal income process has long been a key factor for economic decisions. Individual life-cycle consumption, saving and asset portfolio choice rely heavily on the current and expected future incomes (Browning and Lusardi, 1996; Gourinchas and Parker, 2002; Storeletten, Telmer, Yaron, 2004). Recent discussions on the appropriateness of various reduced-form modeling methods have been centered on heterogeneity across agents (Browning, Ejrnaes, Alvares, 2010; Meghir, Pistaferri, 2004), especially in terms of personal income profiles (Guvenen, 2007; Guvenen, 2009). They have found evidence that at the beginning of one's career, not everyone has the same expected life-time income profile in terms of both starting income and income growth rate. This could potentially help us differentiate heterogeneity from uncertainty. Econometricians do not have all the information that individuals have, therefore what may be perceived as uncertainty by econometricians could actually be heterogeneity among individuals with the same observables included in the model. To understand the income risk faced by a individual, it is very important to isolate the uncertainty



from heterogeneity, this paper incorporates information in occupation choice and switches to shed further light on this question.

This paper shows that the occupation and personal demographics can be an important sources for heterogeneous income profiles as in Guvenue (2007, 2009). First of all, occupation is a key determinant for one's income. In a cross-sectional regression where log income is regressed on personal demographics and occupation dummies, the range of occupational effect is 1.16 while the standard deviation of error term is 0.51, and this is when only a handful of personal variables are included in the regression. The result means the wage dispersion between different occupations can be as great as or even greater than that within the same occupation. Secondly, occupations differs not only in terms of current income, but also in terms of the likelihood of transiting to other occupations and labor force statuses. Each occupation has a very distinct career path due to heterogeneous transition probabilities between occupations and labor force statuses, which determines the future incomes. Occupations are associated with different probabilities of transiting to other occupations or even nonemployment, and the future income depends on the occupations or labor force statuses that one will be in. Furthermore, demographics also play an important role in determining the income growth rate through the overall likelihood of different career movements. For instance, individuals with college degree are more likely to be promoted than individuals with high school degree with other things held constant. The income growth rate is jointly determined by one's current occupation and personal demographics. This approach to modeling the income process is very distinct from conventional approaches which mostly focus on the time series of income itself (Chamberlain and Hirano, 1999; Gustavsson and Osterholm, 2014). Prior literature (Moscarini and Vella, 2008; Kambourov and Manovskii, 2009) on occupational mobility focuses on the estimation of the overall mobility rate at aggregate level without classifying different types of occupation switch.

To construct the monthly transition probabilities at occupational level, a large amount of transition data is required. Monthly CPS is used for its large sample size. The transition probability depends on both demographics and current occupation if employed, so ideally the transition probability between any two states is estimated independently for all combination of demographics, labor force status and current occupation. However, the CPS is far from large enough to allow such estimations given that there are about 500 occupational classifications which

leads to about 250,000 combinations of employment-to-employment (EE) transitions, thousands of demographics combinations and the fact that transitions are rather rare in the data. Therefore I have to impose certain structure on the transition. Based on the results from cross-sectional regression, all occupations are ranked according to their effect on one's income and the ranking is used to determine if an EE transition is a promotion or demotion. For an employed person, the monthly transitions are classified into three categories: promotion, demotion and nonemployed. Such categorical probabilities are studied and evaluated using a series of logistic regressions. The regressions are able to capture the general relation between the probabilities and one's personal demographics as well as the heterogeneity at occupation level. Once a promotion or demotion arises, the new occupation is determined by the empirical distribution constructed using all EE transitions from the same occupation. For a nonemployed person, all monthly transitions are classified into one category: re-employed in any occupation, and its probability is evaluated using a logistic regression. A new occupation is chosen according to the empirical reemployment occupation distribution constructed using all nonemployment-to-employment transitions in the CPS. By combining all above results, the transition probability can be evaluated for all demographics and current occupations if employed.

Given the transition probabilities, I am able to compute the distribution of income at any time for any given initial occupation and personal demographics. The results show that both factors play significant roles in determining income growth and uncertainty. The first example uses the same personal profile with different starting occupations, and the results show the starting occupation has a long lasting effect on one's income. The second example uses four personal profiles which differ in education level. People with higher degree have greater dispersion in income profiles and higher wage growth rate which is consistent with findings in Guvenen (2009). The results highlight that heterogeneous people with the same initial income can have very different income growth rate due to different starting occupation and personal demographics. The initial occupation becomes less and less important as time goes by due to the Markov property of the stochastic income process, but the effect of personal demographics lasts for the whole career. Lastly, by decomposing the variance of income growth rate into expected individual uncertainty and variance of heterogeneous income profile, the model predicts that 55-57% of the income growth variance is due to heterogeneous income profile before age 55.

The paper is organized in the following order: section 2 describes the statistical personal income process model, section 3 introduces the data and occupation ranking, section 4 studies and estimates the transition probabilities, section 5 illustrates the heterogeneous income profiles and section 6 concludes the paper.

## 3.2 Personal Income Process

A person stays in the economy for  $T$  months. The monthly income  $y_{it}$  for individual  $i$  at month  $t$  is determined by the following Mincer earning equation:

$$\log y_{it} = \begin{cases} \phi X_{it} + \sum_{n=1}^N \lambda_n O_{nit} & \text{if employed} \\ b & \text{if nonemployed} \end{cases} \quad (3.1)$$

where  $X_{it}$  denotes the personal demographics which include: a) gender, b) marital status, c) a polynomial of order 4 in age to capture the non-linearity of age effect, d) 3 additional educational dummies for high school degree, some college and college degree respectively, and e) 3 additional racial dummies for black, Asian and other races respectively.  $\{O_n\}_{n=1}^N$  is a set of occupation dummies which captures the overall level of income associated with each occupation. Without loss of generality, occupations are ordered and indexed by the values of  $\lambda_n$  from the smallest to the largest, i.e.  $\lambda_1 < \lambda_2 < \dots < \lambda_N$ .  $O_{nit}$  equals one if individual  $i$  works in occupation  $n$  at time  $t$  or zero otherwise. For nonemployed people, I assume the income is a fixed payment  $b$  and use an additional code  $O_0$  for nonemployment. Hence  $\{O_n\}_{n=0}^N$  can also be used as to describe one's state in terms of labor force status and occupation if employed.

The life-cycle income stream,  $\{y_{it}\}_{t=1}^T$ , depends on one's future labor force status and occupation. The log income while employed in Equation (3.1) is decomposed into two components, and I assume the only source of income uncertainty comes from the stochastic transitions between different states.  $X_{it}$  evolves from time  $t = 1$  to  $T$  deterministically, meaning that the age grows and all other aspects remain the same. Let  $P(O_m|O_n, X_{it})$  denote the one-period transition probability of being in state  $O_m$  at time  $t + 1$  given  $X_{it}$  and current state  $O_n$  at time  $t$ , where  $n, m \in \{0, 1, \dots, N\}$ . The multi-period transition probability  $P^{(\tau)}(O_m|O_n, X_{it})$ , which denotes

the probability of being in state  $O_m$  at time  $t + \tau$  given  $X_{it}$  and current state  $O_n$  at time  $t$ , can be computed by induction as  $P^{(\tau)}(O_m|O_n, X_{it}) = \sum_{l=0}^N P(O_m|O_l, X_{it+\tau-1})P^{(\tau-1)}(O_l|O_n, X_{it})$ . Therefore it suffices to construct the one-period transition probabilities, and multi-period transition probabilities can be inferred from one-period transition probabilities. The stochastic income process  $\{y_{it}|O_{ni1}, X_{i1}\}_{t=1}^T$  depends on the initial occupation and personal demographics. The income profile is defined as  $\{E(y_{it}|O_{ni1}, X_{i1})\}_{t=1}^T$ , and the uncertainty of income at month  $t$  is defined as the s.d. of  $(y_{it}|O_{ni1}, X_{i1})$ .

The income shocks are in the form of random transitions between states  $\{O_0, O_1, \dots, O_N\}$ , which is governed by the one-period transition probabilities. By incorporating heterogeneous transition probabilities at occupational level, the income process is able to generate very different income dynamics for similarly paid starting occupation. Section 3 estimates  $\{\lambda_n\}_{n=1}^N$ , and section 4 describes the estimation procedure for the one-period transition probabilities. Section 5 studies the heterogeneous income profiles.

### 3.3 The CPS Data

In order to have a large enough sample for the estimation of occupation transitions, the monthly Current Population Survey (CPS) is used in this paper. The monthly CPS is the most detailed survey data with occupation and income information at monthly frequency. For a brief introduction of CPS, check Madrian and Lefgren (1999), Kambourov and Manovskii (2013), Moscarini and Thomsson (2007) and my other paper. In order to monitor the change of states for the same person at monthly frequency, consecutive monthly CPSs are merged. I follow Madrian and Lefgren (1999) to match respondents based on household ID, household number and individual line number. Anyone with different sex, different race or excessive change of age between consecutive months is dropped. The monthly CPS from 2003 to 2010 used the same 2000 Standard Occupational Code. A consistent coding system is necessary for the analysis in this paper. The other dataset that can be used is the Survey of Income and Program Participation (SIPP) conducted by the Census Bureau. The advantage of SIPP is that respondents always report income and occupation in each core survey, whereas in CPS only the outgoing respondent report their income. The disadvantage of SIPP is its much smaller sample size and non-rotational respondents, where a panel of respondents

are surveyed for about 4 years until replaced by a new panel. For example the 2008 panel starts with about 105,000 respondents in the first wave and the the sample decreased gradually to less than 72,000 respondents at the end of 15th wave. Given there are about 500 occupations and the occupation switch rarely happens, it is difficult to reliably estimate the transition probabilities at occupational level. The other potential data source is the Panel Study of Income Dynamics (PSID), but it is an annual survey which is not suitable for estimation highly frequent mobilities.

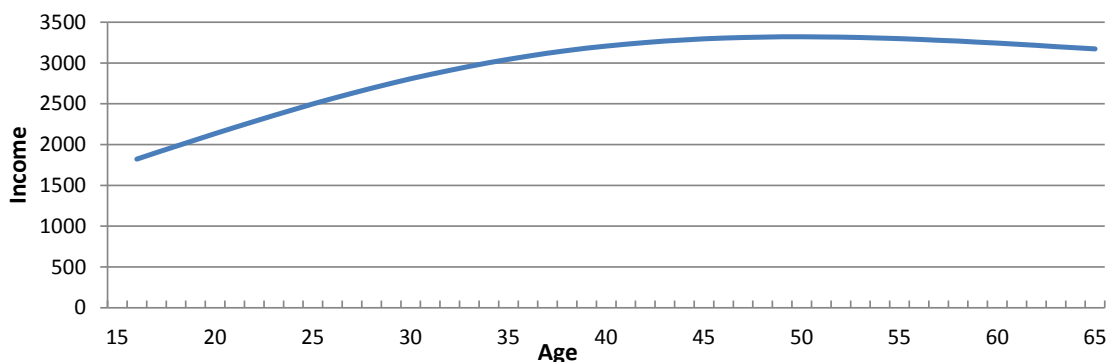
### 3.3.1 Occupation Rank

To understand the effect of an occupation switch on one's income, one needs to first understand the relation between one's income and occupation. Using the subsample of CPS in which outgoing employed respondents report both occupation and weekly income, I regress the log monthly income on personal demographics and occupation dummies as in Equation (3.1). The occupational effects  $\{\lambda_n\}_{n=1}^N$  are sorted from the smallest to the largest, and the ranking is used as the benchmark for either promotion or demotion. Since the survey spans 8 years, I use the growth rate of nominal GDP per capita from 2003 to 2010 to adjust incomes from different times. The monthly income is computed as weekly earning multiplied by 4.34 and is in 2009 dollar. Based on the subsample in which employed outgoing respondents reported no less than 30 weekly working hours and positive income, a total of 502 occupations have been reported. Not all occupations have the same number of observations in the sample, occupations like wood model maker, hunter, railroad police appeared less than 50 times in the sample. On the other hand, occupations like manager, sales supervisor, administrative secretary have more than 30,000 observations. In order to have a reliable estimate of the occupational effect, this paper focuses on 482 occupations ( $N = 482$ ) which have at least 50 observations. In total there are 1,230,230 observations, and the number of observations ranges from 50 to 32,565 with an average of 2,552.

The estimation results are reported in Table (3.1). Three specifications differ in whether controlling for monthly and state fixed effects or not, and the coefficient estimations are very close among all three specifications. Based on specification III, male workers are associated with 21% higher wage, and being married is associated with 4.4% higher wage. Each additional education level increases one's wage, where college degree increases wage by 64% compared to people who have less than high school degree. As for race, white people is associated with highest wage and

black people are associated with lowest wage. To illustrate the effect of age, I plot the income for a white single male with high school degree and works in the average paid occupation in Figure (3.1). The polynomial in age is able to capture the hump shape of income where the income peaks at age 50. The fixed effects of both monthly dummies and state dummies are omitted from the table. As for state fixed effects, Montana is associated with the lowest income whereas the D.C. is associated with the highest, the difference is as large as 37.8%.

**Figure 3.1 :Regression Implied Monthly Income by Age**



This is for a single white male with high school degree and work in the average paid occupation.

The estimates of occupational effects are too many to report in the table, and the range is 1.16, meaning that the highest ranked occupation pays 3.19 times as much as the lowest ranked occupation. This means the wage dispersion between different occupations can be greater than that within the same occupation. All 482 occupations are sorted and ranked according to  $\{\lambda_n\}_{n=1}^{482}$ . Based on a white single male with a high school degree, the income is plotted against the occupation rank in Figure (3.2). Food serving related occupations like dishwashers, cafeteria attendants and fast food workers are among the lowest paid occupations. On the other hand, petroleum engineers, chief executives and lawyers are among the most paid occupations in the sample.

### 3.4 Occupation and Labor Force Status Transition

Because individuals stay in the survey for consecutive months, I am able to track and study the same individuals in two consecutive months. In total there are over 5.7 millions individual-entries which appeared again in the following month for the whole sample. To estimate  $P(O_m|O_n, X_{it})$  and

Table 3.1 : Individual Annual Income Regression Results

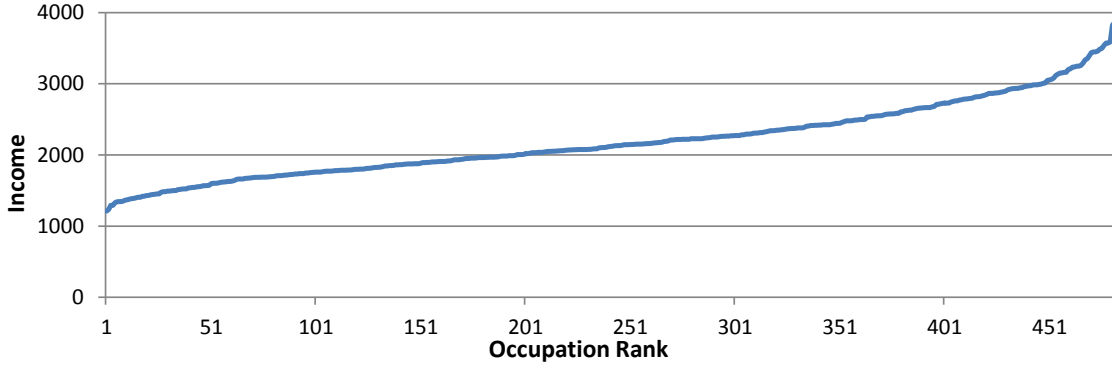
Coefficients \ Specification	I	II	III
<i>Male</i>	0.1945 ***	0.1943 ***	0.1911 ***
<i>Married</i>	0.0376 ***	0.0371 ***	0.0427 ***
<i>Age</i>	0.0808 ***	0.0816 ***	0.0818 ***
<i>Age</i> <sup>2</sup> /100	-0.1303 ***	-0.1344 ***	-0.1384 ***
<i>Age</i> <sup>3</sup> /1000	0.0063 *	0.0070 *	0.0081 **
<i>Age</i> <sup>4</sup> /100000	0.0004	0.0000	-0.0008
<i>Education Category</i>			
<i>high school</i>	0.1905 ***	0.1909 ***	0.1968 ***
<i>some college</i>	0.2466 ***	0.2476 ***	0.2517 ***
<i>college degree</i>	0.4942 ***	0.4957 ***	0.4941 ***
<i>Racial Category</i>			
<i>black</i>	-0.0764 ***	-0.0759 ***	-0.0735 ***
<i>asian</i>	-0.0108 ***	-0.0100 ***	-0.0468 ***
<i>other races</i>	-0.0210 ***	-0.0203 ***	-0.0240 ***
Monthly dummy	No	Yes	Yes
State dummy	No	No	Yes

a) The baseline for regression is a white single female with less than high school education. The regression is based on the subsample of monthly CPS in which outgoing respondents report income and occupation at the same time. Only respondents who worked at 30 hours and earned positive income are included. Only occupations with at least 50 observations are included.

b) The sample consists of all respondents who were 16-65 years old, civilian population and not disabled at the time of survey. CPS final weights are used to adjust for the representativeness of each observations. \* denotes statistically significant at 10% confidence level, \*\* at 5% and \* at 1%.

c) The monthly dummy consists of 95 monthly indicator variables, the state dummy consists of 50 state indicator variables. The coefficient estimations for monthly dummies and state dummies are omitted from the table. In specification II, the constant term incorporates the average monthly effects. In specification III, the constant term incorporates both average monthly effects and average state effects.

**Figure 3.2 :Regression Implied Monthly Income by Occupation Rank**



This is for a 30-years old single white male with a high school degree.

especially its dependence on  $X_{it}$ , one can use separate subsample with the same  $X_{it}$  to estimate the transition probabilities. However, there are 233,289 transition probabilities to estimate for each of 1,600 combinations of personal demographics, so estimating the transition probabilities separately for each demographic subsample would introduce too many zero probabilities. For instance, in the full sample there are four transitions from economist to higher paid occupations. Three of them became lawyers and one of them became a chief executive, and in all four cases they respondents were male. Therefore the sample probability for a female economist to become a lawyer or chief executive is zero because there is no such transitions in the sample. In order to interpolate such probabilities, I need to impose a structure on the transition probabilities.

All transition probabilities are classified and aggregated into four categories based on their implications for one's career development. For employed people with  $O_n$ ,  $n > 0$ , transitions can be: a) promotion,  $\Pi^P(O_n, X_{it}) = \sum_{l=n+1}^N P(O_l|O_n, X_{it})$ , b) demotion,  $\Pi^D(O_n, X_{it}) = \sum_{l=1}^{n-1} P(O_l|O_n, X_{it})$ , or c) nonemployed:  $\Pi^{EN}(O_n, X_{it}) = P(O_0|O_n, X_{it})$ . For nonemployed people with  $O_0$ , they can either remain nonemployed or work in any occupation in the following month, and the transitions are aggregated to  $\Pi^{NE}(O_0, X_{it}) = \sum_{l=1}^N P(O_l|O_0, X_{it})$ . A transition probability can be rewritten as the product of a categorical transition probability and a conditional transition



probability which can be estimated separately:

$$P(O_m|O_n, X_{it}) = \Pi^P(O_n, X_{it})Q^P(O_m|O_n, X_{it}) \quad \text{if } m > n > 0 \quad (3.2)$$

$$P(O_m|O_n, X_{it}) = \Pi^D(O_n, X_{it})Q^D(O_m|O_n, X_{it}) \quad \text{if } n > m > 0 \quad (3.3)$$

$$P(O_0|O_n, X_{it}) = \Pi^{EN}(O_n, X_{it}) \quad \text{if } n > 0 \quad (3.4)$$

$$P(O_m|O_0, X_{it}) = \Pi^{NE}(O_0, X_{it})Q^{NE}(O_m|O_0, X_{it}) \quad \text{if } m > 0 \quad (3.5)$$

where  $Q^P(O_m|O_n, X_{it}) = \frac{P(O_m|O_n, X_{it})}{\sum_{l=n+1}^N P(O_l|O_n, X_{it})}$  is the conditional probability from  $O_n$  to  $O_m$  conditioned on a promotion,  $Q^D(O_m|O_n, X_{it}) = \frac{P(O_m|O_n, X_{it})}{\sum_{l=1}^{n-1} P(O_l|O_n, X_{it})}$  is the conditional probability from  $O_n$  to  $O_m$  conditioned on a demotion, and similarly  $Q^{NE}(O_m|O_0, X_{it}) = \frac{P(O_m|O_0, X_{it})}{\sum_{l=1}^N P(O_l|O_0, X_{it})}$  is the conditional probability of arriving at occupation  $O_n$  conditioned on re-employment.  $\Pi^P(O_n, X_{it})$ ,  $\Pi^D(O_n, X_{it})$ ,  $\Pi^{EN}(O_n, X_{it})$ ,  $\Pi^{NE}(O_0, X_{it})$  are referred to as categorical probabilities.

In the following I will fit the categorical probabilities with a series of logistics regressions, and use the fitted categorical probabilities to replace the true values in Equations (3.2) - (3.5). For EE conditional probabilities, I assume that they do not depend on  $X_{it}$  such that I can replace them by the sample conditional probabilities constructed from the whole EE transition sample. For NE conditional probabilities, based on the fact that different demographic groups have different distribution of the occupation upon re-employment, I assume that individuals return to their previous occupation with a fixed probability before sampling a new occupation randomly.

### 3.4.1 Employed Transitions

Firstly, The sample categorical probabilities are analyzed to establish the fact that they exhibit strong heterogeneity in both demographics and occupations. The relations between the categorical probabilities and demographics as well as occupations are analyzed using logistic regressions. Secondly, the sample conditional probabilities are studied, and by using Chi-square goodness of fit test, it can be shown that they do not differ significantly in demographics on average. The results imply that the transition probability depends on demographics mostly through its relation with the categorical probabilities but not with conditional probabilities. Therefore I assume that conditional probabilities only depend on initial occupation but not on demographics. This assumption allows

me to study the categorical probabilities and conditional probabilities separately, where the later can be constructed using the whole EE transition sample regardless of demographics. By combining these two results, the employed transition probabilities can be constructed for any demographics and occupation.

## Employed Categorical Probabilities

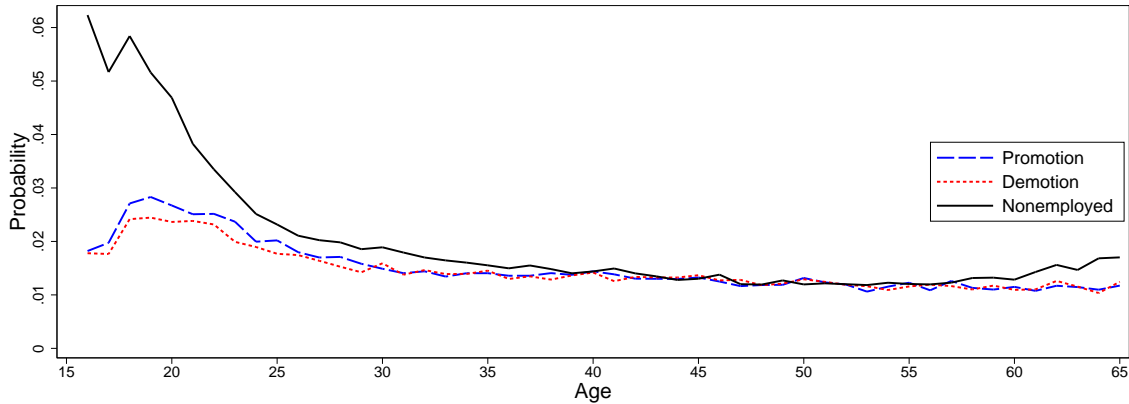
I first describe the sample categorical probabilities for different demographics. Table (3.2) summarizes the sample averages constructed from the whole sample. Male workers have higher mobilities than female workers for all three types of transitions. Single workers are also associated with high mobilities than married workers for all three types of transitions, partially due to younger age. As for different education categories, employed people with higher education are associated with lower mobility for promotion, demotion and nonemployment compared to lower educated people. For different racial groups, black employed people are associated with the highest promotion and demotion probabilities, and other racial group is associated with highest nonemployment probability. White employed people are associated with lowest promotion and demotion probabilities while Asian people are associated with lowest nonemployment probability. Figure (3.3) plots the sample average monthly transitions by age based on employed subsample. The blue dash line is the promotional probability, and older age is associated with lower likelihood for promotion. The red dotted line is the demotional probability, where older workers are less likely to be demoted. The demotional probability lines up with the promotional probability closely except for age below 30, where promotional probability is slightly greater. The black solid line is the nonemployed probability, and it starts at very high level for young workers and eventually decreases to about 1.2% since late 40s and picks up a bit after age 55 due to retirement. In general, young workers below age 30 face greater mobility, workers between age 30 and age 55 face relatively stable mobility and workers above age 55 are slightly more likely to become nonemployed. Overall, the employed transition probabilities exhibit strong patterns in personal characteristics  $X_{it}$ .

Employed transitions also exhibit strong heterogeneity in occupation. Figure (3.4) plots the sample promotional, demotional, nonemployment and employed in same occupation probabilities by current occupation rank for employed subsample. Panel A plots the fraction of the sample who will be promoted to a better paid occupation in the next month. In general,

**Table 3.2 : Monthly Categorical Probability Summary, Employed Subsample**

Characteristics \ Transitions	$\Pi^P$	$\Pi^D$	$\Pi^{EN}$
Full Sample	1.48 % ( 0.01 % )	1.44 % ( 0.01 % )	1.81 % ( 0.01 % )
<i>Gender</i>			
<i>Male</i>	1.53 % ( 0.01 % )	1.48 % ( 0.01 % )	1.95 % ( 0.01 % )
<i>Female</i>	1.42 % ( 0.01 % )	1.38 % ( 0.01 % )	1.65 % ( 0.01 % )
<i>Marital Status</i>			
<i>married</i>	1.26 % ( 0.01 % )	1.24 % ( 0.01 % )	1.26 % ( 0.01 % )
<i>single</i>	1.80 % ( 0.01 % )	1.73 % ( 0.01 % )	2.61 % ( 0.01 % )
<i>Education Category</i>			
<i>no high school</i>	1.88 % ( 0.02 % )	1.77 % ( 0.02 % )	3.77 % ( 0.03 % )
<i>high school</i>	1.52 % ( 0.01 % )	1.46 % ( 0.01 % )	2.14 % ( 0.01 % )
<i>some college</i>	1.46 % ( 0.01 % )	1.43 % ( 0.01 % )	1.71 % ( 0.01 % )
<i>college degree</i>	1.33 % ( 0.01 % )	1.31 % ( 0.01 % )	0.95 % ( 0.01 % )
<i>Racial Category</i>			
<i>white</i>	1.40 % ( 0.01 % )	1.35 % ( 0.01 % )	1.72 % ( 0.01 % )
<i>black</i>	1.93 % ( 0.02 % )	1.99 % ( 0.02 % )	2.57 % ( 0.03 % )
<i>asian</i>	1.77 % ( 0.03 % )	1.61 % ( 0.03 % )	1.24 % ( 0.03 % )
<i>other races</i>	1.66 % ( 0.04 % )	1.66 % ( 0.04 % )	2.70 % ( 0.05 % )

This table summarizes the three classifications of monthly mobility for different demographic groups based on employed subsample. CPS final weights are used to adjust for representativeness of each observation. Standard deviations are reported in parentheses.

**Figure 3.3 :Monthly Categorical Probability by Age, Employed Subsample**

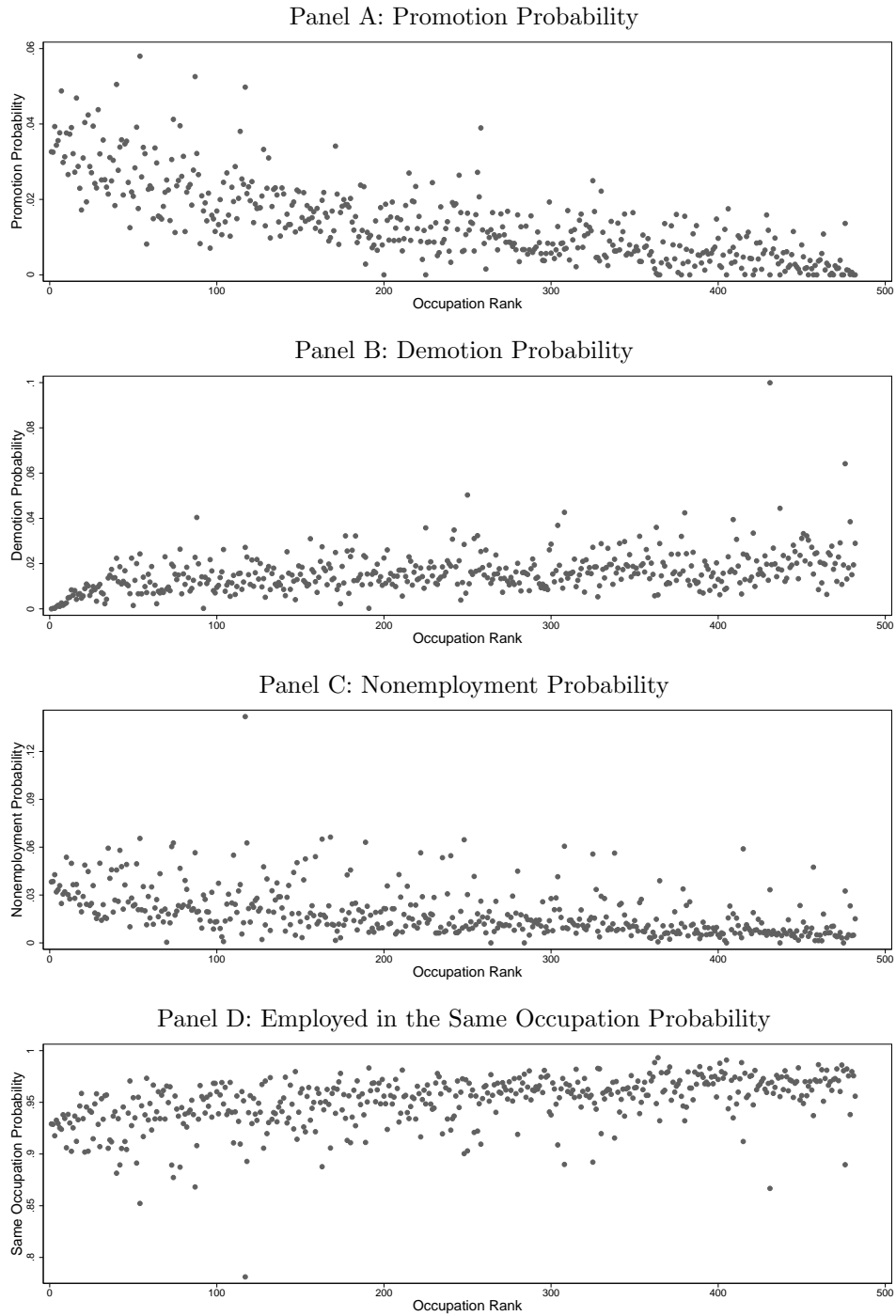
The figure plots the three categories of transition for employed subsample: employed with promotion, employed with demotion and nonemployed in the next month. CPS final weights are used to adjust for representativeness of each observation.

lower paid occupations have greater potentials and are more likely to be promoted. There are a few occupations with promotional probabilities significantly higher than the trend. For instance, telemarketers with rank of 54 have the highest promotion probability of 5.8%. Panel B plots the fraction of the sample who will be demoted to a lower paid occupation in the next month. People working in higher paid occupations were generally more likely to be demoted because of greater downward potential, however, the likelihood seems to be rather constant once occupation is ranked above 100. There are few exceptions that deviations from the trend. for instance, roustabouts with an occupation rank of 431 have a demotion probability of 10%, and oil extraction workers and entertainers have the second and third highest demotion probabilities of 5.4% and 5.0% respectively. Panel C plots the fraction of the sample who will become nonemployed in the following month. Lower paid occupations face greater risk of nonemployment in general, and they frequently deviate from the trend and even reach above 6%. Actors have the highest probability of 14.2%, and while the majority of occupations have nonemployment probability below 3%, there are still some above 3% regardless of the rank. Panel D plots the fraction of the sample who will be employed with the same occupation in the next month which is computed as one minus all three probabilities above. Lower paid occupations are generally associated with lower occupational stability. This could because younger workers are more likely to be in one of the lower paid occupations and they frequently sample different occupations during their early stage of career. A few occupations have significantly lower probabilities than the rest. For example actor has the lowest probability of 78.1% while the average among all occupations is 95.1%. To conclude, there are two main findings from these figures, the first is that people working in lower paid occupations generally have greater mobility mostly due to higher promotion and nonemployment probabilities, the second is that the transition probabilities vary greatly even for occupations with similar incomes. The latter could introduce a large amount of heterogeneity for future income uncertainty which may be well perceived by economic agents but not by econometricians.

### **Employed Categorical Probability Regressions**

It has been shown that the categorical probabilities differ greatly with demographics  $X_{it}$  as well as current states  $O_n$ , therefore it is important to estimate the transition probabilities while taking into account both the heterogeneous demographics and occupation. A series of logistic regressions

Figure 3.4 :Sample Categorical Probabilities by Occupation Rank for Employed



- a) A total of 482 occupations are included, each has at least 50 observations for estimating the occupational income.  
b) 4,172,597 observations are included, The CPS final weight is used to adjust for the representativeness of each observation.

are used to fit the data and study the effect of demographics and heterogeneous probabilities at occupations. Theoretically a multinomial logistic regression with four outcomes can also be used to achieve the same goal, but it is computationally impossible with over 4.1 million observations and more than 500 variables. Instead a series of binary logistic regressions are used. The regression equation is as followed:

$$\log\left(\frac{P_i}{1-P_i}\right) = \theta X_{it} + \sum_{n=1}^N \gamma_n O_{nit} \quad (3.6)$$

where the log of odd is a linear function in personal demographics and occupational effect.  $X_{it}$  includes the same variables as in Equation (3.1) and  $\gamma_n$  captures the heterogeneous effects at occupation level. The same regression equation is used three times with different subsample and interpretations of the coefficients. The first logistic regression is to determine whether employed individual  $i$  will change states or not due to promotion, demotion or nonemployment and the whole employed subsample is included. The  $P_i$  is the probability of changing states to  $m \neq n$ . This binary logistic regression is still computationally impossible, therefore I use a two-step procedure. The first step is to estimate  $\theta$  with a polynomial of order 3 in occupation rank, where the later is used to approximate the general trend in occupation effect. The second step is to estimate  $\{\gamma_n\}_{n=1}^N$  while holding  $\theta$  constant from the first step. The second logistic regression is to determine whether individual  $i$  will remain employed (i.e. either promotion or demotion) conditioned on change of states, and only the employed subsample with promotion, demotion and nonemployment in the second months is included.  $\theta, \{\gamma_n\}_{n=1}^N$  can be estimated jointly due to much smaller sample size. The third logistic regression is to determine whether individual  $i$  will be promoted conditioned on EE transitions, and the employed subsample with either promotion or demotion in the second month is included. All coefficients can be estimated jointly due to even smaller sample size. By combining the results from all three logistic regressions, I am able to construct  $\Pi^P(O_n, X_{it})$ ,  $\Pi^D(O_n, X_{it})$ ,  $\Pi^{EN}(O_n, X_{it})$  for any  $O_n$  and  $X_{it}$ , which will be used later in constructing probability space of life-cycle income stream.

Table (3.3) reports the results for all three logistic regressions. Log ratios are reported in the table and the heterogeneous occupational effect is omitted due to large number of coefficients. Regression I estimates the probability of changing states due to either promotion, demotion and nonemployment in the second month. Male is more likely to change state and married people are less likely to, also higher educated people are less likely to change too. Black people are associated with

highest probability of changing states and white people are with the lowest probability. Regression II estimates the conditional probability of remaining employed conditioned on change to states. Male has slightly higher conditional probability, so do married people. People with higher education have higher conditional probability of remaining in the workforce. Lastly, Asian people have the highest conditional probability while other races have the lowest. Regression III estimates the conditional probability of promotion conditioned on EE transitions. Male and married are associated with higher conditional probability. Higher education significantly increases the conditional probability. White people are associated with the highest conditional probability whereas black people are associated with the lowest. Combining the results from all three logistic regressions, I can study the age effect on three types of transition. Figure (3.5) plots the probabilities by age for two personal profiles. Panel A is based on a single white male with a high school degree working in the average occupation. The blue long dashed line represents the probability of promotion, which increases from age 16 until late 20s, and remains relatively constant before falling slightly after age 60. The red dotted line represents the probability of demotion, and it starts at a high level at early age and decreases eventually throughout the whole range. The black solid line is the probability of nonemployment, which starts high at early age and then decreases until age 55 before picking up slightly again. The overall level and relative position of these three lines depend on the personal characteristics as well as current occupation. Panel B is based on a single white male with a college degree. The shape of each line is very similar to that in Panel A, but the relative positions are very different, and the promotional probability is greater than both demotional probability and nonemployment probability for age 21 and above.

Figure (3.6) plots the regression implied probabilities by occupation rank for a 30-years old single white male with a high school degree. Panel A plots the implied promotion probability, where lower ranked occupations have higher probability in general. However, individual occupation can deviate from the trend severely. Panel B plots the demotion probability, and higher occupations have greater probability due to their seniority, again there exists strong heterogeneity where similarly paid occupations can have vastly different probabilities. Panel C plots the nonemployment probability, higher paid occupations have slightly lower probability in general but the strong heterogeneity is still evident. In general the plots are similar to the sample probabilities in Figure (3.4) where personal demographics are not controlled for.

**Table 3.3 : Employed Transition Regression Results**

Characteristics	I	II	III
<i>Constant</i>	0.1611	0.0098	0.0002
<i>Male</i>	1.1825	1.0985	1.8104
<i>Married</i>	0.7949	1.2106	1.2379
<i>Age</i>	0.9856	1.6816	1.7892
<i>Age</i> <sup>2</sup> /100	0.8557	0.1342	0.1630
<i>Age</i> <sup>3</sup> /1000	1.0424	1.4042	1.2831
<i>Age</i> <sup>4</sup> /100000	0.9719	0.8093	0.8806
<i>Education Category</i>			
<i>high school</i>	0.8192	1.1514	1.6297
<i>some college</i>	0.7515	1.2448	2.2963
<i>college degree</i>	0.7410	1.6233	5.7366
<i>Racial Category</i>			
<i>black</i>	1.3554	0.9350	0.7513
<i>asian</i>	1.1270	1.3028	0.9432
<i>other races</i>	1.2080	0.8477	0.8152

a) Regression I determines the probability that employed individual  $i$  will change states due to either promotion, demotion or nonemployment. The coefficients are estimated in two steps, the first step estimates  $\theta$  with a proxy for occupational effect, and the second step estimates  $\{\gamma_n\}_{n=1}^N$  while holding  $\theta$  constant from the first step. The constant term incorporates the average of  $\{\gamma_n\}_{n=1}^N$ .

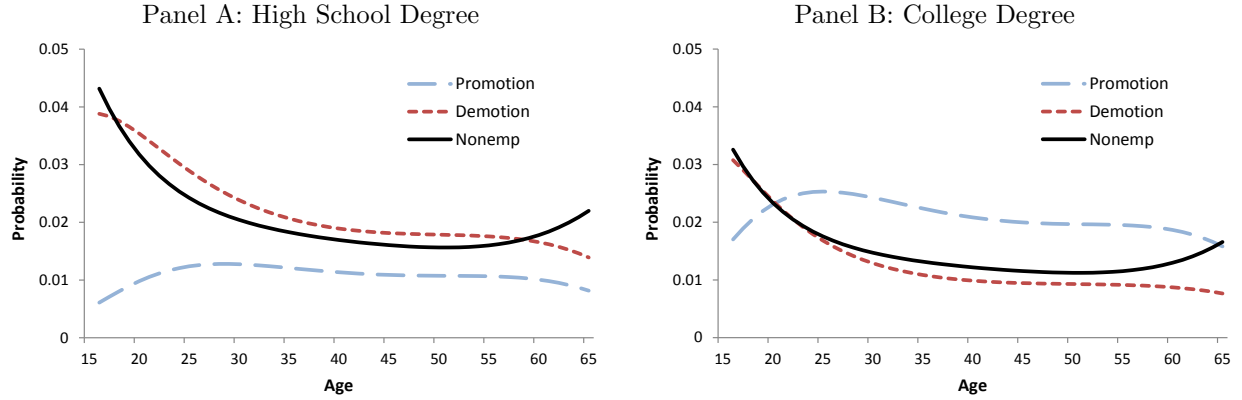
b) Regression II determines the probability that individual  $i$  will remain employed (i.e. either promotion or demotion) conditioned on change of states (i.e. promotion, demotion, or nonemployment). All coefficients are estimated jointly. The results are based on all occupations that do not perfectly predict either success or failure.

c) Regression III determines the probability that individual  $i$  will be promoted conditioned on change of states and remaining employed (i.e. either promotion or demotion). All coefficients are estimated jointly. The results are based on all occupations that do not perfectly predict either success or failure.

d) Log ratios are reported in the table. All coefficient estimates are statistically significant at 99% confidence level.



**Figure 3.5 :Regression Implied Monthly Transition Probabilities**



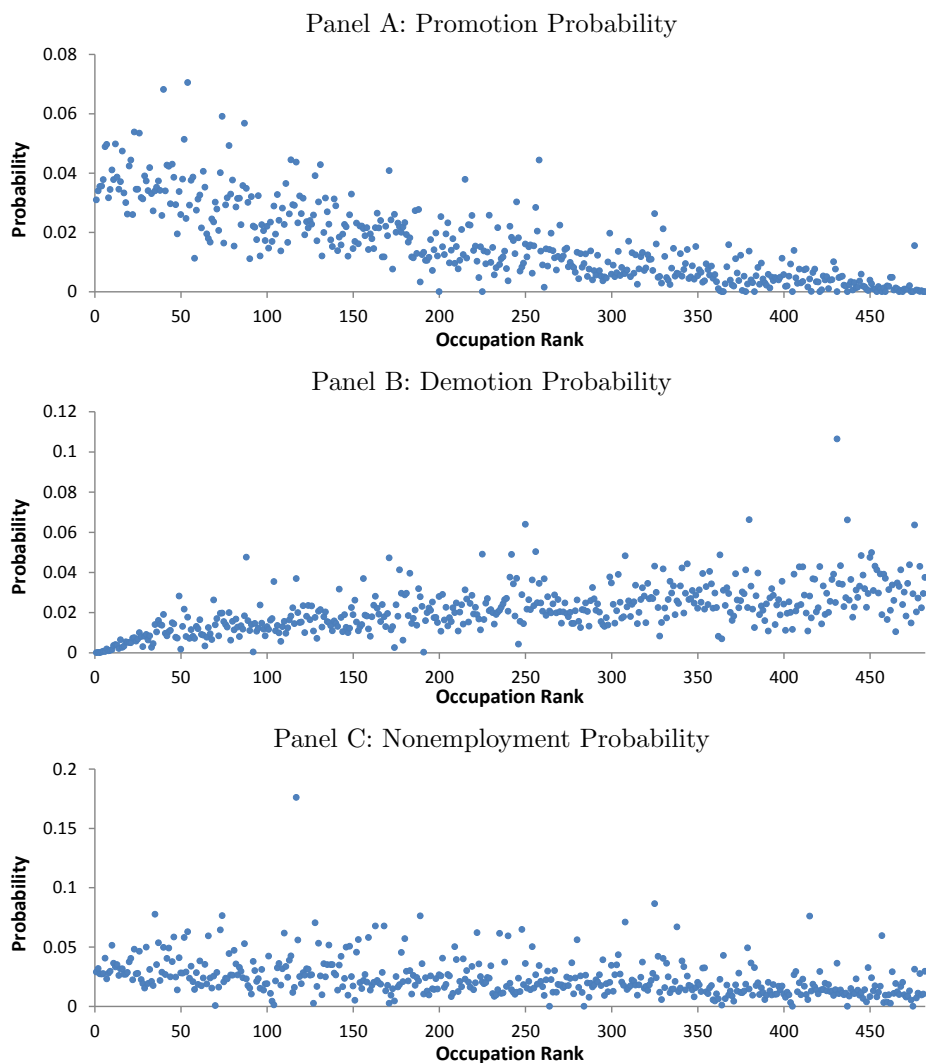
- a) Panel A is based on a white single male with a high school degree.  
b) Panel B is based on a white single male with a college degree.

## EE Conditional Probabilities

The conditional probabilities,  $Q^P(O_m|O_n, X_{it})$ ,  $0 < n < m$  and  $Q^D(O_m|O_n, X_{it})$ ,  $0 < m < n$ , determine the probabilities of arriving at occupation  $m$  from  $n$  upon either a promotion or demotion. Using the whole EE transition sample regardless of demographics, the sample conditional probabilities can be constructed. Given the large number of states, it is impossible to plot the distribution of terminal states for each initial state. However, there are a few key features that are worth mentioning. First of all, not all terminal states will be visited with equal likelihood from each initial state, and some transitions are much more likely than others. These conditional probabilities reveal the possible career paths associated with each occupation. For promotional transitions with initial occupation  $O_n$ ,  $n \in \{1, 2, \dots, 481\}$ , 20 of them will transit to one of the terminal occupations with probability greater than 80%, 48 of them will transit to one of the terminal occupations with probability greater than 50%, and 116 of them will transit to one of the terminal occupations with probability as large as 30% at least. For demotional transitions with initial occupation  $O_n$ ,  $n \in \{2, 3, \dots, 482\}$ , 11 of them will transit to one of the terminal occupations with probability greater than 80%, 30 of them with one probability greater than 50%, and 84 of them with at least one probability greater than 30%. For instance, economists<sup>1</sup>, ranked at 473rd

<sup>1</sup>The definition of an economist is: someone who conducts research, prepares reports, or formulates plans to address economic problems related to the production and distribution of goods and services or monetary and fiscal policy. May collect and process economic and statistical data using sampling techniques and econometric methods. Excludes Market Research Analysts and Marketing Specialists.

Figure 3.6 :Regression Implied Transition Probabilities by Occupation Rank



- a) Occupations are ranked according to their fixed effects on one's income.
- b) All plots are based on a 30 years old, single white male with a high school degree.

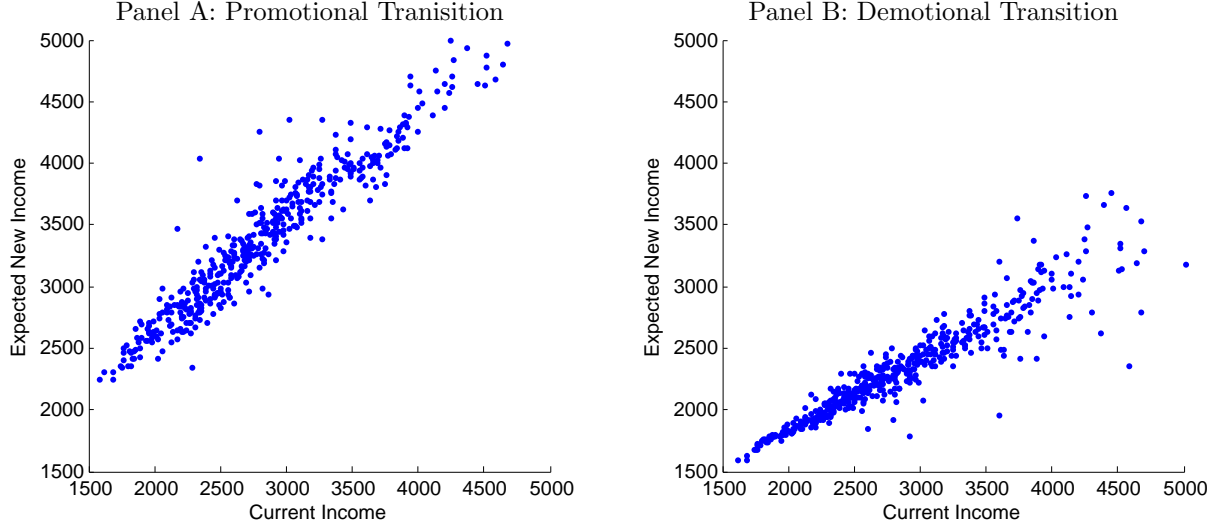
out of 482 occupations, has only two feasible promotional terminal states: lawyers with probability 76% and chief executives with probability 24%. While there are several other occupations that are ranked above economists like nuclear engineers, surgeons, oil drill operators and petroleum engineers, none of them are feasible as a career advancement for an economist. Similarly, while economists can be demoted to any of the 472 occupations ranked below, only 22 of them are actually visited in the data. The most likely terminal states are: general manager with probability 24%, management analyst with probability 10%, security sales agents with probability 7.5% and middle-school teachers with probability 7.4%. Secondly, transitions from different initial states have different expected income at the new terminal state. Figure (3.7) plots the expected new monthly income transiting from each initial states. The figure is based on a 30-years old single white male with a high school degree. Panel A plots the expected new income for promotional conditional transitions. The higher paid initial occupations are associated with higher expected new income in general, but the correlation is far from perfect. Panel B plots the expected new income for demotional conditional transitions. Again the higher paid initial occupations are roughly associated with higher expected new income, but there exists a large dispersion for similarly paid initial occupations, especially for those highly ranked ones. Of course people will not stay at the new occupation forever and will eventually transit to other states, so the figure does not imply the income for the rest of the career, but it does give an idea about how the heterogeneous income could arise from heterogeneous conditional transition probabilities.

To understand whether the conditional probabilities vary significantly in demographics, I use Chi-squared goodness of fit test to analyze whether the conditional probabilities constructed using a certain demographic subsample is statistically different from that constructed using the whole sample. The demographic subsamples are defined by gender, marital status, age, education or race. For promotional transition with initial state  $O_n$ , the test statistic is defined as:

$$T = \sum_{l \in U_n} \frac{(Q^P(O_m|O_n, X) - Q^P(O_m|O_n))^2}{Q^P(O_m|O_n)} \quad (3.7)$$

where  $Q^P(O_m|O_n)$  denote the promotional conditional probability constructed from the whole sample,  $Q^P(O_m|O_n, X)$  denote the promotional conditional probability constructed from certain subsample, and  $U_n$  denotes the set of states with  $Q^P(O_m|O_n) > 0$ ,  $m > n$ . The test statistics is compared to the Chi-square distribution with degrees of freedom equaling to the number of states

**Figure 3.7 :Expected New Income**



The figure are based on a 30-years old single white male with a high school degree.

in  $U_n$  minus one, therefore the test can only be conducted for  $O_n$  with number of states in  $U_n$  greater than 1. Similarly, the test statistic for demotional transition is defined as:

$$T = \sum_{l \in D_n} \frac{(Q^D(O_m|O_n, X) - Q^D(O_m|O_n))^2}{Q^D(O_m|O_n)} \quad (3.8)$$

where  $Q^D(O_m|O_n)$  denote the demotional conditional probability constructed from the whole sample,  $Q^D(O_m|O_n, X)$  denote the demotional conditional probability constructed from certain subsample, and  $D_n$  denotes the set of states with  $Q^D(O_m|O_n) > 0$ ,  $m < n$ . The test statistics is compared to the Chi-square distribution with degrees of freedom equaling to the number of states in  $D_n$  minus one.

Table (3.4) reports the test results. The null hypothesis is that the conditional probability constructed from certain demographic subsample is the same as that constructed from the whole sample. Each row summarizes all the tests with all possible initial states. Column 2 reports the size of the subsample relative to the full sample, column 3 reports the number of tests performed, column 4 reports the average p-value, column 5 the fraction of p-values greater than 0.9 and column 6 the fraction of p-values greater than 0.95. The top panel pertains to promotional transitions, where the null hypothesis holds for most part except for Asian subsample where 5.56% of p-values

are greater than 0.95, and for other racial subsample where 8.44% of p-values are greater than 0.95. However, those two subsamples are very small compared to the full sample and hence have much fewer data points than other subsamples, therefore the conditional probability constructed from them are more likely to deviate from that constructed from the full sample. The bottom panel pertains to demotional transitions, and the null hypothesis holds for most part except for, again, Asian and other racial subsamples. The Asian subsample has 5.07% of p-values greater than 0.95 and other racial subsample has 9.73% of p-values greater than 0.95. Again, the small size of these two subsamples can be the main reason. To conclude, only Asian and other racial subsamples have some evidence that the conditional transitional probability could depend on personal characteristics, and all other subsamples do not reject the null hypothesis. Therefore I assume that neither  $Q^P(O_m|O_n, X_{it})$  nor  $Q^D(O_m|O_n, X_{it})$  are variant to personal characteristics, meaning that when any two people of different  $X_{it}$  are to be promoted or demoted from the same occupation  $O_n$ , the likelihood of arriving at any occupation  $O_m$  is the same for them. Hence  $X_{it}$  can be omitted and  $Q^P(O_m|O_n)$  and  $Q^D(O_m|O_n)$  can be estimated using the whole EE transition sample regardless of personal demographics.

### 3.4.2 Nonemployed Transitions

Nonemployed people can either remain nonemployed or work in any occupation in the next month. I first describe  $\Pi^{NE}(O_0, X_{it})$ , the categorical probability of re-employment, for different demographic groups. The results imply that the likelihood of re-employment is significantly related to personal demographics. The relation is formally analyzed by using logistic regressions. Then the conditional probability of arriving at a particular occupation after re-employment,  $Q^{NE}(O_m|O_0, X_{it})$ , is constructed. The fact that different demographic groups return to very different occupations implies the re-employed people do not randomly enter each occupation with same probability.

#### NE Categorical Probabilities

The probability of returning to the workforce certainly depends on one's personal demographics. Column 2 in table (3.5) summarizes the probabilities for different demographic groups: Male is more likely to return to labor force than female, and married people are less likely to be employed

**Table 3.4 : Conditional Probability Chi-square Goodness of Fit Test**

Promotion					
Subsample	Size	No.	Mean	Fraction>0.90	Fraction>0.95
<i>Age</i>					
16 – 40	58.7 %	439	0.046	0.68 %	0.46 %
> 40	41.3 %	440	0.065	1.59 %	1.14 %
<i>Gender</i>					
Male	55.3 %	383	0.023	0.26 %	0.00 %
Female	44.7 %	444	0.097	2.03 %	1.80 %
<i>Marital Status</i>					
married	50.1 %	432	0.033	0.23 %	0.23 %
single	49.9 %	443	0.071	1.13 %	0.68 %
<i>Education Category</i>					
no high school	13.2 %	450	0.120	2.67 %	2.00 %
high school	29.8 %	445	0.107	3.15 %	1.80 %
some college	28.8 %	447	0.096	2.46 %	2.46 %
college degree	28.2 %	431	0.097	3.02 %	2.78 %
<i>Racial Category</i>					
white	78.3 %	389	0.021	0.51 %	0.51 %
black	13.8 %	450	0.136	3.56 %	3.56 %
asian	5.4 %	450	0.155	6.00 %	5.56 %
other races	2.5 %	450	0.179	8.67 %	8.44 %
Demotion					
Subsample	Size	No.	Mean	Fraction>0.90	Fraction>0.95
<i>Age</i>					
16 – 40	57.1 %	466	0.033	0.86 %	0.64 %
> 40	42.9 %	467	0.037	0.43 %	0.21 %
<i>Gender</i>					
Male	55.1 %	431	0.029	0.93 %	0.70 %
Female	44.9 %	470	0.086	2.34 %	1.91 %
<i>Marital Status</i>					
married	50.8 %	468	0.040	0.64 %	0.43 %
single	49.2 %	467	0.036	1.07 %	0.64 %
<i>Education Category</i>					
no high school	12.8 %	472	0.121	4.45 %	3.60 %
high school	29.6 %	467	0.065	1.71 %	1.50 %
some college	29.0 %	471	0.083	2.55 %	2.34 %
college degree	28.6 %	467	0.089	2.36 %	2.14 %
<i>Racial Category</i>					
white	77.7 %	436	0.012	0.23 %	0.23 %
black	14.6 %	473	0.100	1.90 %	1.48 %
asian	5.1 %	473	0.135	5.50 %	5.07 %
other races	2.6 %	473	0.175	9.94 %	9.73 %

Column 1 lists the demographic subsamples on which the test is conducted. Column 2 lists the size of the subsample relative to the full sample. Column 3 provides the number of tests that are included and there are at least two terminal states with positive probabilities based on full sample. Column 4 provides the average p-value, column 5 gives the fraction of p-values that are greater than 0.9 and column 6 gives the fraction of p-values that are greater than 0.95.

**Table 3.5 : Re-employment Summary**

Characteristics	Re-emp Probability	Avg. Occ. Rank
<i>Gender</i>		
<i>Male</i>	13.35 %	171
<i>Female</i>	8.91 %	135
<i>Marital Status</i>		
<i>married</i>	9.18 %	189
<i>single</i>	11.85 %	130
<i>Education Category</i>		
<i>no high school</i>	8.81 %	86
<i>high school</i>	11.12 %	132
<i>some college</i>	11.96 %	161
<i>college degree</i>	10.93 %	272
<i>Racial Category</i>		
<i>white</i>	10.76 %	159
<i>black</i>	10.42 %	124
<i>asian</i>	8.77 %	188
<i>other races</i>	10.90 %	131

The sample includes a total of 1,437,000 observations.

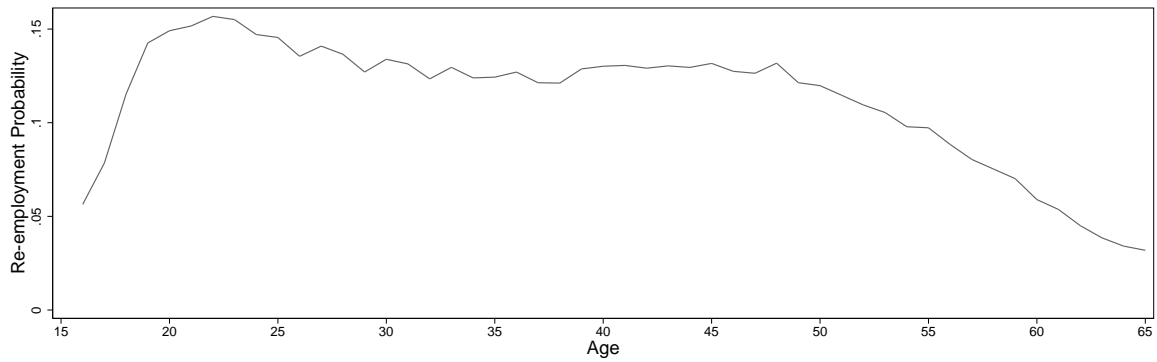
again than single people. Higher education normally leads to higher probability of re-employment except for college degree. As for different racial categories, Asian people have the lowest probability of returning to labor force and other racial category has the highest probability. Panel A in figure (3.8) plots the probability of re-employment by age. The probability starts at just above 5% at age 16 and rapidly increases to about 15% after age 20, and decreases eventually after age 50 to below 5% at age 65.

### NE Categorical Probabilities Regression

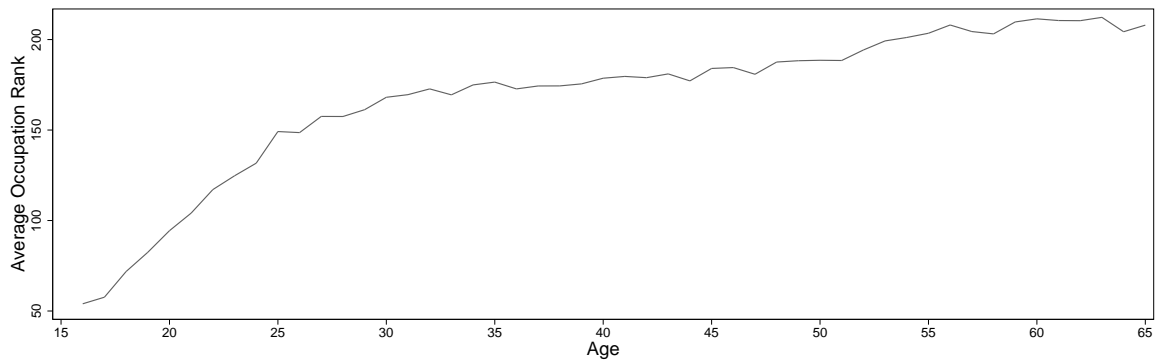
The general trend of re-employment probability is similarly studied using a logistic regression. Due to limited longitudinal dimension of monthly CPS, the last occupation  $O_n^{last}$  cannot be identified in most cases and must be ignored from regression. Hence the probability of returning to workforce

**Figure 3.8 :Nonemployed Transition**

Panel A: Re-employment Probability by Age



Panel B: Re-employed Occupation Rank by Age



- a) Figure A plots the average re-employment probability by age.
- b) Figure B plots the average occupation rank by age for those who just return to workforce.
- c) CPS final weight is used in aggregation.



**Table 3.6 : Re-employment Regression Result**

Coefficients \ Specification	I	II	III
<i>Constant</i>	6.59E-06	5.34E-06	3.82E-06
<i>Male</i>	1.671	1.697	1.697
<i>Married</i>	0.768	0.761	0.760
<i>Age</i>	2.984	3.023	3.052
<i>Age<sup>2</sup>/100</i>	0.013	0.012	0.012
<i>Age<sup>3</sup>/1000</i>	2.112	2.129	2.141
<i>Age<sup>4</sup>/100000</i>	0.622	0.619	0.617
<i>Education Category</i>			
<i>high school</i>	1.204	1.207	1.202
<i>some college</i>	1.225	1.233	1.223
<i>college degree</i>	1.269	1.270	1.254
<i>Racial Category</i>			
<i>black</i>	0.849	0.848	0.859
<i>asian</i>	0.718	0.716	0.723
<i>other races</i>	0.923	0.926	0.908
Monthly dummy	No	Yes	Yes
State dummy	No	No	Yes

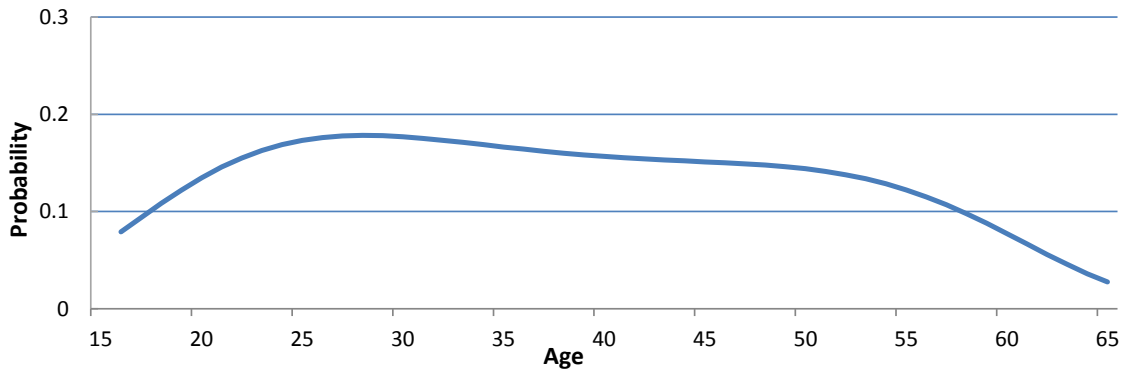
CPS final weights are used...

only depends on  $X_{it}$ . The equation is as followed:

$$\log\left(\frac{P_i}{1 - P_i}\right) = \theta X_{it} \quad (3.9)$$

where the  $p_i$  is the probability of re-employment. The regression results is summarized in Table (3.6). Three specifications are included for whether controlling for monthly and state fixed effects. Coefficient estimates are very close among all three specifications. Male is more likely to join the workforce than female, and married people are less likely than single people. As for education, higher education leads to greater probability of re-employment. For different racial groups, Asian has the lowest likelihood for returning to workforce, and white people are associated with the highest probability. To illustrate the change of probability with age, I plot the probability for a single white male with high school degree and no children in Figure (3.9). The probability increases in age before age 30 and decreases slowly afterwards, and eventually reaches a very low level at age 65 due to approaching retirement.

**Figure 3.9 :Regression Implied Monthly Re-employment Probabilities**



The plot is based on a single white male with a high school degree.

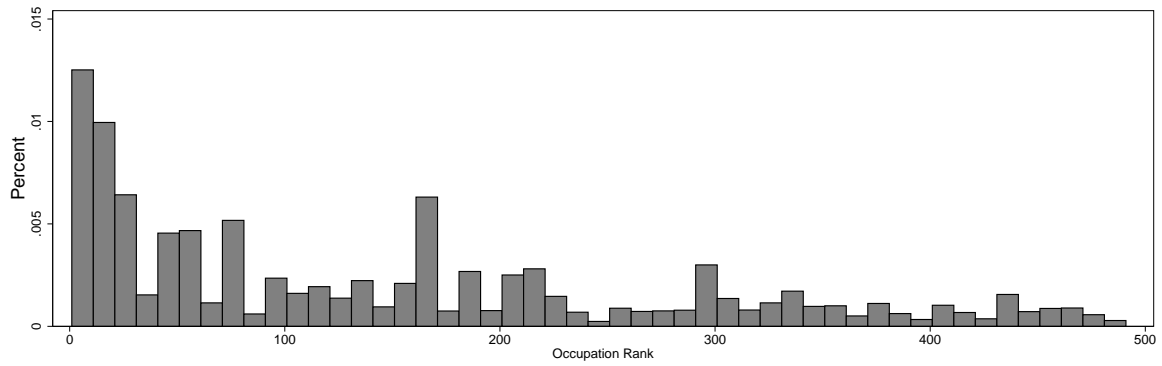
### NE Conditional Probabilities

Figure (3.10) describes the occupation distribution upon re-employment. Panel A plots the histogram of sample occupation distribution upon re-employment, where the majority of people return to lower ranked occupations, especially those with ranking below 100. Panel B compares the cumulative density function of occupation rank upon re-employment to that of the full sample. The black solid line represents the re-employment sample and the red dash line represents the full sample. The CDF for the re-employed sample is significantly higher than that for the full sample, meaning a larger fraction of the re-employed sample are concentrated in lower paid occupations. The plot also suggests people are likely to return to a lower paid occupation when returning to workforce, so not all people are able to return to the same occupation as before.

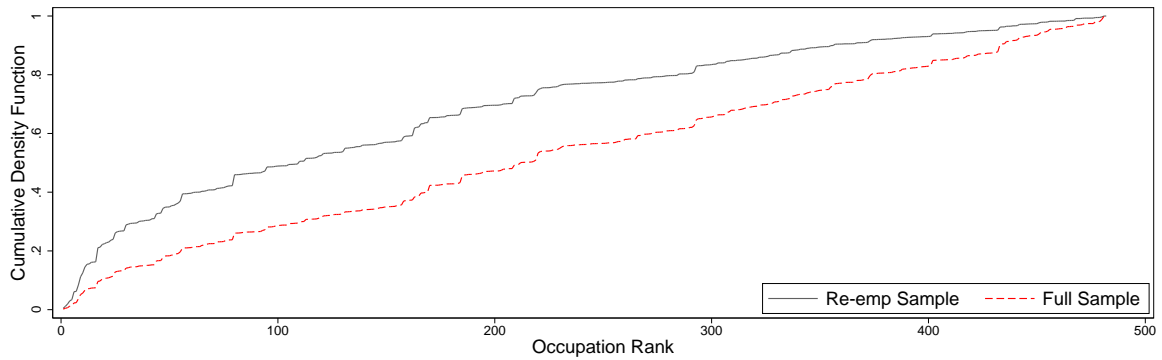
Unlike the EE conditional probabilities, the NE conditional probability of arriving at a particular occupation after re-employment is significantly related to one's personal demographics. Column 3 in Table (3.5) lists the average occupation rank upon returning to the workforce for different demographical groups. Male returns to a better paid occupation on average than female, so do married people compared to single people. As for educational category, higher educational attainment significantly increases the seniority of the occupation. For different racial groups, Asian people are associated with the highest average occupation rank upon re-employment, whereas black people are associated with the lowest average occupation rank. Panel B in Figure (3.8) plots the average occupation rank upon re-employment by age. The seniority of occupation increases

**Figure 3.10 :Re-employment Occupation Choice**

Panel A: Re-employment Occupation Sample Distribution



Panel B: Re-employment Occupation CDF



- a) Figure A plots the sample occupation distribution for the re-employed sample.
- b) Figure B plots the CDF for the re-employed sample and that for the whole sample. The black solid line represents the re-employed sample and the red dash line represents the full sample.

significantly in age over the life-cycle. This is mostly likely due to the fact that re-employment people are likely to return to their previous occupation than simply randomly sampling a new occupation. For example, previous results show that people with higher education attainment have significantly higher probability of promotion, and this means they are more likely to work in a higher ranked occupation before becoming nonemployed. And by the time they return to the workforce, they may very well return to the same highly ranked occupation as before, which could explain the difference in average occupation rank upon re-employment. Therefore I assume that a person will return to his previous occupation with probability  $S^{NE}$ , or randomly choose an occupation according to the re-employment occupation distribution as in Panel A of Figure (3.10) with probability  $1 - S^{NE}$ . Equation (3.5) can be rewritten as:

$$P(O_m|O_0, O_n^{last}, X_{it}) = \Pi^{NE}(O_0, X_{it}) \left( S^{NE} I(O_m = O_n^{last}) + (1 - S^{NE}) Q^{NE}(O_m|O_0) \right) \quad \text{if } m > 0 \quad (3.10)$$

where  $O_n^{last}$  denotes the last occupation before becoming nonemployed and  $Q^{NE}(O_m|O_0)$  is the sample distribution of occupation upon re-employment regardless of personal demographics as in Figure (3.10). This transition probability from nonemployment is used for constructing the probability space of life-cycle income stream.

### 3.5 Heterogeneous Income Profile

In this section I illustrate that the different initial occupation and personal demographics can generate heterogeneous income profile. Given the transition probabilities from previous section, the probability space over all possible life-cycle income streams can be constructed using forward induction. The income profile is the times series of life-cycle income stream expectations given initial condition  $O_{ni1}$  and  $X_{it}$ . The first part illustrates the heterogeneous income profile caused by either different initial occupation or different personal demographics. The second part measures the fraction of income growth rate that can be explained by heterogeneous income profile. The rest of parameters are listed in Table (3.7). To compute the present value of one's life-time income, a discounting factor of  $0.98^{0.083}$  is used. The probability of returning to the same occupation after re-employment,  $S^{NE}$ , is assumed to be 80%, and the monthly income while nonemployed is 500. The monthly income while employed is evaluated by Equation (3.1).

**Table 3.7 : Personal Income Process Parameters**

Parameter:	$S^{NE}$	b	Age
Value:	80%	500	30 - 65

### 3.5.1 Different initial conditions

#### Different Initial Occupation

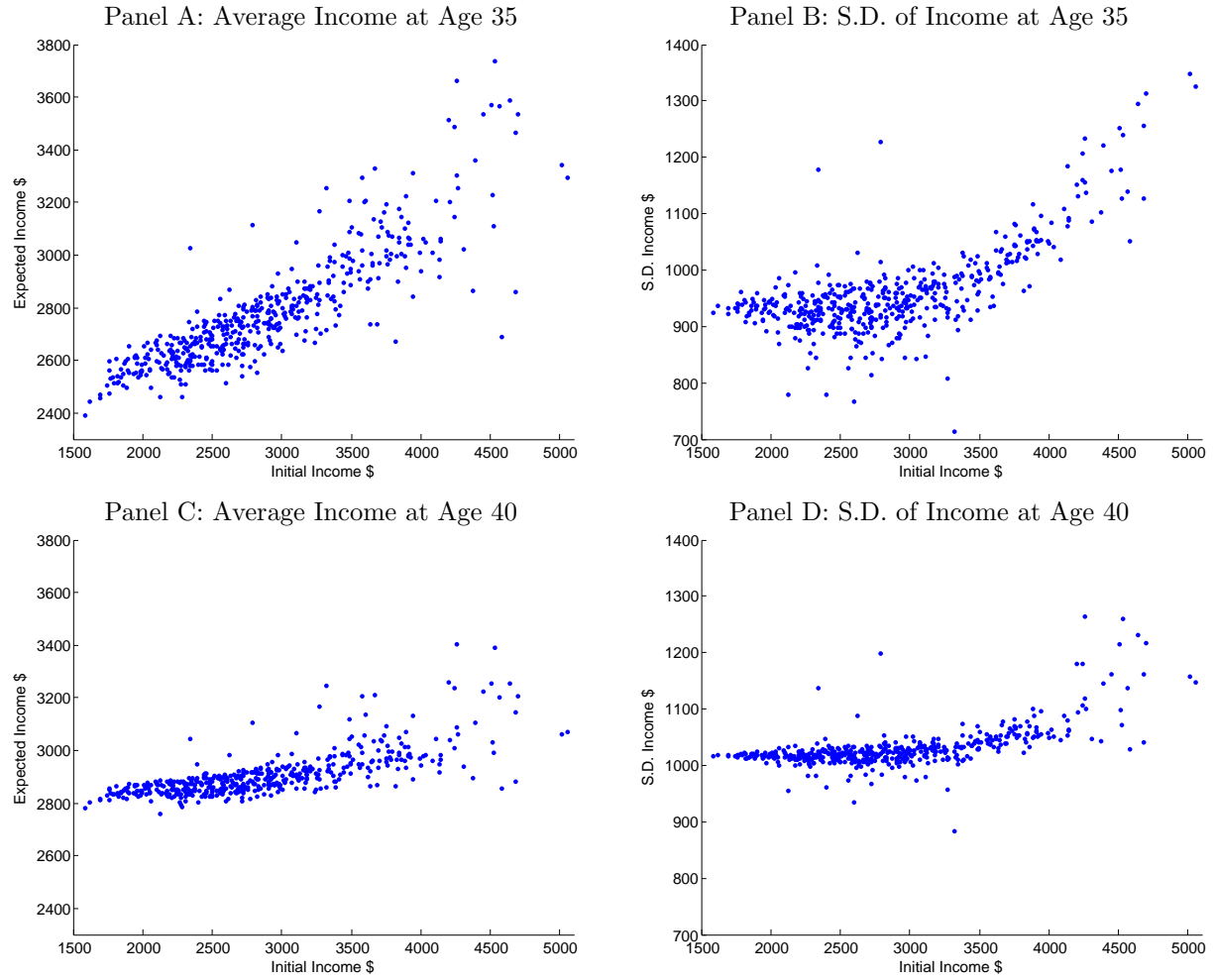
A 30-years old white single male with a high school degree is used as the example in this example, and each person stays in the economy for 35 years until age 65. Figure (3.11) plot the income process for different starting occupation at different age. Panel A and B plot the mean and s.d. of income at age 35 by initial occupations. Both the mean and the s.d. of income are positively correlated with the initial income, but the correlations are far from perfect as similarly paid starting occupation can have very different income after 10 years due to heterogeneous transition probabilities. Panel C and D plot the mean and s.d. of income at age 40 by initial occupations. Similarly, both mean and s.d. are positively correlated with the initial income, but the correlation is much less significant than that at age 40. This is due to the Markov property of the stochastic income process where the initial state starts to matter less and less and everyone tends to follow the long-run stationary distribution over all states. The plots illustrate the heterogeneous transition probabilities can generate heterogeneous income profiles.

#### Different Personal Demographics

This part uses two personal profiles to illustrate the differences that personal demographics can make. In addition to the same personal profile used in the previous part, three more personal profiles added: a) a 30-years old white single male with less than a high school degree , b) a 30-years old white single male with some college education, c) a 30-years old white single male with a college degree. So the only difference between those four personal profiles is the education level.

Figure (3.12) compares the life-cycle income profile of those four personal profiles by initial income. Each profile has 482 dots representing each starting occupation. The black dots represent

**Figure 3.11 : Heterogenous Income Profile by Occupation**



The figure are based on a 30-years old single white male with a high school degree.

less than high school degree, green cross the high school, red circle the some college degree and blue asterisks the college degree. For the same initial occupation, the person with higher education level has higher initial income because of the educational effect as in Equation (3.1). Panel A and B plot the mean and s.d. of the income at age 35, where both are greater for higher educations. Panel C and D plot the mean and s.d. of the income at age 40, and again both the mean and s.d. are greater for higher educations, and the difference is more pronounced than those at age 40. The comparison between those four personal profiles highlights the variation in career development associated with different personal demographics. For the same initial income, the person with lower degree needs to work in a higher ranked occupation than the person with a higher degree, and higher ranked occupation generally has greater potential for demotion and lower potential for promotion. Furthermore, the person with a lower degree is more likely to be demoted and less likely to be promoted than the person with a higher degree in the same occupation. Both factors cause the income for the person with a higher degree to grow faster than that for the person with a lower degree. This finding is consistent with Guvenen (2009) where the author found that individuals with college degree have significantly larger dispersion of income profiles and higher wage growth rate than individuals with high school degree. Of course, education is only one of the personal characteristics that affect the growth rate of income, other personal demographics like gender, marital status or race can also play a role as the regression results in Table (3.3) suggest.

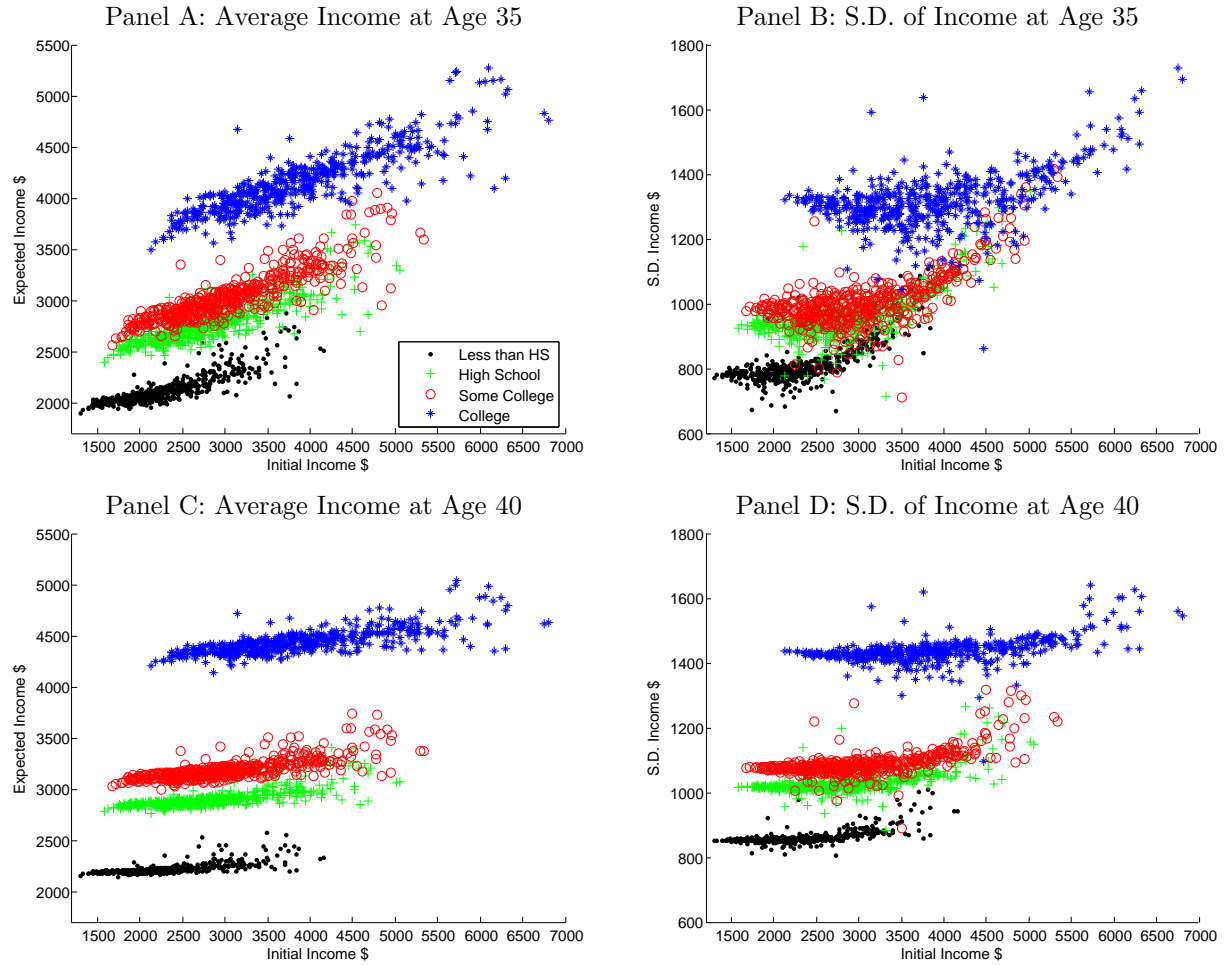
### 3.5.2 Heterogeneity vs Inequality

In this part I will measure how much the income growth inequality can be explained by the heterogeneity generated by initial occupation and personal demographics. The income growth at month  $t$  is defined as  $g_{it} = \log y_{it} - \log y_{i1}$ , where the subscript  $i$  denotes one's initial state  $O_{ni1}$  and  $X_{i1}$  with weight  $w_i$ . Using the law of total variance, the population variance of  $g_{it}$  can be decomposed as:

$$Var(g_{it}) = E_i[Var(g_{it}|i)] + Var_i[E(g_{it}|i)] \quad (3.11)$$

where the expectations are taken over the weight  $w_i$ . The first part on the right hand side,  $E_i[Var(g_{it}|i)]$ , is the average uncertainty faced by each individual. The second part,  $Var_i[E(g_{it}|i)]$ , measures the income growth inequality due to heterogeneity and is known to the individuals. Therefore the fraction of income inequality explained by heterogeneity is  $\frac{Var_i[E(g_{it}|i)]}{Var(g_{it})}$ . The weight

**Figure 3.12 : Different Personal Profiles**



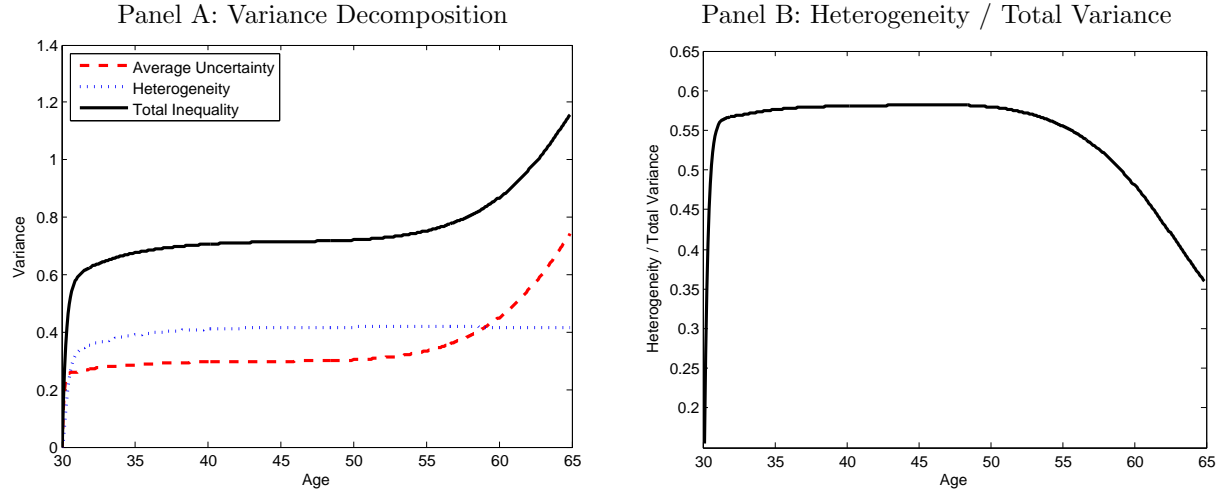
The black dots represent less than high school degree, green cross the high school, red circle the some college degree and blue asterisks the college degree.

Panel A plots the average income at age 35 for each initial occupation and personal profile. Panel B plots the s.d. of income at age 35 for each initial occupation and personal profile. Panel C plots the average income at age 40 for each initial occupation and personal profile. Panel D plots the s.d. of income at age 40 for each initial occupation and personal profile.



$w_i$  is over all 483 states (including nonemployed) and constructed from all monthly CPS where only white single male at age 30 are included. Figure (3.13) plots the time series of these two components. Panel A plots the two components of total variance and the total variance. The red dash line represents the average uncertainty which remains relatively stable before taking off after age 55. This is because the fraction of nonemployed people increases significantly after age 55, and this brings up the income uncertainty due to vast income difference between employed and nonemployed. Blue dotted line represents the variance of income profile, which remains very stable throughout the career. The black solid line is the sum of these two and represents the unconditional income growth variance. Panel B plots the fraction of income growth variance explained by heterogeneous income profile. 55-57% of the total variance is due to heterogeneous income profile before age 55, and the fraction decreases eventually afterwards due to increasing average uncertainty.

**Figure 3.13 : Heterogeneity vs Inequality**



The weights are constructed from all monthly CPS where only white single male at age 30 are included.

### 3.6 Conclusion

In this paper I propose a new way to model personal income process. The log income process during employment is composed into two components, the first one is determined by personal characteristics and evolves deterministically, and the second component is determined by the occupation which changes over time stochastically. Occupations are ranked according to their

effects in the cross-sectional regression in which log monthly income is regressed on personal characteristics and occupational dummies. The rank is used to determine whether an EE transition is promotion or demotion. For employed people, the transitions are classified into three categories: promotion, demotion and nonemployment. These three categorical probabilities are studied and evaluated using a series of logistic regressions, and the results are able to capture the heterogeneous transition probabilities due to both occupation and personal demographics. The EE conditional probabilities, which determines the probability of arriving at a specific occupation conditioned on either promotion or demotion, are computed using the whole EE transition sample. By combining the regression results and EE conditional transition probabilities, I am able to construct any transition probabilities for employed people. For nonemployed people, the probability of nonemployment is studied and evaluated using a logistic regression which captures the effects of both personal characteristics and occupation. The probability of arriving at a particular occupation after re-employment is determined by a combination of returning to the same occupation with a fixed probability or a random draw from the empirical occupation distribution upon re-employment. Combining the re-employment probability and NE conditional transition probability, I am able to construct the transition probabilities for nonemployed people.

Given the transition probabilities, I am able to compute the distribution of income at any time for any given initial occupation and personal demographics. The results show that both factors play significant roles in determining income growth and uncertainty. The first example uses the same personal profile with different starting occupations, and the results show the starting occupation has a long lasting effect on one's income. The second example uses four personal profiles which differ in education level. People with higher degree have greater dispersion in income profiles and higher wage growth rate which is consistent with findings in Guvenen (2009). The results highlight that heterogeneous people with the same initial income can have very different income growth rate due to different starting occupation and personal demographics. The initial occupation becomes less and less important as time goes by due to the Markov property of the stochastic income process, but the effect of personal demographics lasts for the whole career. Lastly, by decomposing the variance of income growth rate into expected individual uncertainty and variance of heterogeneous income profile, the model predicts that 55-57% of the income growth variance is due to heterogeneous income profile before age 55.

## References

- Browning, Ejrnaes, Alvarez, 2010, Modelling income processes with lots of heterogeneity, *Review of Economic Studies* 77, 1353-1381
- Chamberlain, Hirano, 1999, Predictive distribution based on longitudinal earnings data, *Panel Data Econometrics* 55, 211-242
- Gourinchas, Parker, 2002, Consumption over the life cycle, *Econometrica* 70, 47-89
- Gustavsson, Osterholm, 2014, Does the labor-income process contain a unit root? Evidence from individual-specific time series, *Journal of Economic Dynamics and Control* 47, 152-167
- Guvonen, 2007, Learning your earning: are labor income shocks really very persistent?, *American Economic Review* 97(3), 687-712
- Guvonen, 2009, An empirical investigation of labor income processes, *Review of Economic Dynamics* 12, 58-79
- Kambourov, Manovskii, 2009, Occupational specificity of human capital, *International Economic Review* 50(1), 63-115
- Kambourov, Manovskii, 2013, A cautionary note on using march CPS and PSID data to study worker mobility, working paper.
- Madrian, Lefgren, 1999, A note on longitudinal matching current population survey respondents, NBER technical working paper No. 247
- Meghir, Pistaferri, 2004, Income variance dynamics and heterogeneity, *Econometrica* 72(1), 1-32
- Moscarini, Thomsson, 2007, Occupational and job mobility in the US, *Scandinavian Journal of Economics* 109(4), 807-836
- Moscarini, Vella, 2008, 2008, Occupational mobility and the business cycle, NBER working paper No. 13819
- Storesletten, Telmer, Yaron, 2004, Consumption and risk sharing over the life cycle, *Journal of Monetary Economics* 51, 609-633

# APPENDIX

## 3.A Occupation Choice

This section describes how to use the CPS sample to construct the empirical distribution of occupation switch probabilities either for continuously employed people or people who just join the workforce.

### 3.A.1 Occupation Choice upon EE Promotion/Demotion

For continuously employed people, whenever an promotion or demotion arises, the new occupation is determined by the following algorithm. There are 482 occupations  $\{O_n\}_{n=1}^{482}$  included, and all occupations are ordered by their fixed occupational effect in Equation (3.1) where rank 1 denotes the lowest. For each occupation  $O_n$ , Let  $S_{n,l}$  denote the set of all observed occupation switch from  $O_n$  to  $O_l$  each with individual  $i$ 's weight  $w_i^{nl}$ , then the empirical probability of transiting from  $O_n$  to  $O_m$  ( $m > n$ ) upon promotion is defined as:

$$P(O_m|O_n) = \frac{\sum_{S_{n,m}} w_i^{nm}}{\sum_{l=n+1}^{482} \sum_{S_{n,l}} w_i^{nl}} \quad (3.12)$$

Similarly, the empirical probability of transiting from  $O_n$  to  $O_m$  ( $m < n$ ) upon demotion is defined as:

$$P(O_m|O_n) = \frac{\sum_{S_{n,m}} w_i^{nm}}{\sum_{l=1}^{n-1} \sum_{S_{n,l}} w_i^{nl}} \quad (3.13)$$

### 3.A.2 Occupation Choice upon Re-employment

For nonemployed people who will become employed again in the following month, I assume the person with return to the same occupation at last employed with probability  $S^{NE}$ , otherwise the new occupation is determined by the following re-employment empirical distribution. Let  $R_n$  denote the set of all observed re-employment with new occupation being  $O_n$  and each has weight  $w_i^n$ , then

the empirical probability of working in occupation  $O_n$  is:

$$P(O_n) = \frac{\sum_{R_n} w_i^n}{\sum_{n=1}^{482} \sum_{R_n} w_i^n} \quad (3.14)$$