Empirical Essays in Household Economics

Hao Li
hhli@smu.edu

Follow this and additional works at: https://scholar.smu.edu/hum_sci_economics_etds

Part of the Econometrics Commons, Health Economics Commons, International Economics Commons, Labor Economics Commons, and the Public Economics Commons

Recommended Citation
https://scholar.smu.edu/hum_sci_economics_etds/6

This Dissertation is brought to you for free and open access by the Economics at SMU Scholar. It has been accepted for inclusion in Economics Theses and Dissertations by an authorized administrator of SMU Scholar. For more information, please visit http://digitalrepository.smu.edu.
EMPIRICAL ESSAYS IN HOUSEHOLD ECONOMICS

Approved by:

Dr. Daniel Millimet
Professor of Economics

Dr. Wenhua Di
Senior Economist of the Federal Reserve Bank, Dallas

Dr. Thomas Fomby
Professor of Economics

Dr. Elira Kuka
Assistant Professor of Economics
EMPIRICAL ESSAYS IN HOUSEHOLD ECONOMICS

A Dissertation Presented to the Graduate Faculty of the
Dedman College
Southern Methodist University

in
Partial Fulfillment of the Requirements
for the degree of
Doctor of Philosophy

with a
Major in Economics

by
Hao Li

B.A., Economics and Law, Nankai University, Tianjin, China
Master of Economics, University of Illinois at Urbana-Champaign

May 19, 2018
ACKNOWLEDGMENTS

This work could not have been accomplished without the wisdom of my wonderful advisor, Prof. Millimet. I am thankful to Daniel Millimet, Elira Kuka and Tom Fomby for their valuable advice and continued feedback. I am also grateful to Alberto Abadie, Elizabeth Cascio, Raj Deb and the participants in 15th Texas Econometrics Camp and attendees in the Brown Bag of Department of Economics at SMU for helpful comments. The Department of Economics in SMU is a wonderful place to learn and grow, I appreciate the help from all my professors. I’m also forever grateful to my family for being so patient with me and my friends for supporting me.
This dissertation is consist of three empirical studies in economics. The first part empirically examines the effect of universal pre-k on labor force participation of fertility age women in Oklahoma. I investigate the policy effect from both theoretical and empirical perspectives. I apply the synthetic control method (SCM) to the Current Population Survey (CPS) data to identify the causal relationship between universal pre-k and female labor outcomes. I find that universal pre-k increases labor supply of women aged 25 to 45 in Oklahoma. The second empirical study focuses on how the birth outcomes of US children respond to exposure to Chinese import competition. Exploring the variation in US trade exposure driven by China’s supply-side productivity changes and falling in trade costs, I find evidence that the rise of Chinese exports to US is not harmful to newborn health and infant mortality; instead, empirical evidence shows that the percentage of low birth weight infants in US counties is largely reduced by the increase in US-China trade volume. In the last chapter, I develop a nonparametric partial identification approach to bound transition probabilities under various assumptions on the measurement error and mobility processes. This approach is applied to panel data from the United States to explore short-run mobility before and after the Great Recession.
# TABLE OF CONTENTS

LIST OF FIGURES ................................................................. ix

LIST OF TABLES ................................................................ xii

CHAPTER

1. THE EFFECT OF UNIVERSAL PRE-KINDERGARTEN POLICY ON FEMALE LABOR FORCE PARTICIPATION — A SYNTHETIC CONTROL APPROACH .................................................................. 1
   1.0.1. Main Results...................................................... 17
   1.0.2. Treatment Effect Heterogeneity......................... 27
   1.0.3. Expansion of Case Study to Georgia Universal Pre-k ....... 30

2. TRADE LIBERALIZATION AND NEWBORN HEALTH: EVIDENCE FROM US EXPOSURE TO CHINESE IMPORT COMPETITION ..................... 37
   2.0.4. US Exposure to Chinese Import Competition ............ 42
   2.0.5. US Exposure to Chinese Import Competition ............ 44
   2.0.6. Birth Outcomes and Family Characteristics............... 45
   2.0.7. Main Results...................................................... 46
   2.0.8. Potential Mechanisms ............................................. 50
       2.0.8.1. Trade Exposure and Air Pollution .................. 50
       2.0.8.2. Trade Exposure and Access to Cheaper Goods ...... 52
       2.0.8.3. Trade Exposure and Fertility ....................... 55

3. PARTIAL IDENTIFICATION OF ECONOMIC MOBILITY: WITH AN APPLICATION TO THE UNITED STATES ..................................... 63
   3.0.9. Setup ............................................................. 69
   3.0.10. Misclassification ................................................... 73
       3.0.10.1. Assumptions ............................................. 73
E. SUBSAMPLE RESULTS OF THE EFFECT OF OKLAHOMA UNIVERSAL PRE-K .................................................................. 107

F. SUPPLEMENTAL APPENDICES FOR CHAPTER 3 ......................... 115

BIBLIOGRAPHY................................................................... 144
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Maternal Labor Force Participation Rate—Oklahoma vs. the Rest of US States (excluding Georgia)</td>
<td>17</td>
</tr>
<tr>
<td>1.2 Maternal Labor Force Participation Rate (25-45 sample)—Oklahoma vs. Synthetic Oklahoma (Traditional SCM)</td>
<td>18</td>
</tr>
<tr>
<td>1.3 Maternal Labor Force Participation Rate (25-45 sample)—Oklahoma vs. Synthetic Oklahoma (SCM allowing for negative weights)</td>
<td>19</td>
</tr>
<tr>
<td>1.4 Placebo Tests on Synthetic Control Results of Oklahoma</td>
<td>20</td>
</tr>
<tr>
<td>1.5 Pre/post RMSPE Ratio of Female Labor Force Participation Rate by States</td>
<td>21</td>
</tr>
<tr>
<td>1.6 Synthetic control analysis on Percentage of Full-time Working Women Aged 25-45</td>
<td>23</td>
</tr>
<tr>
<td>1.7 Synthetic control Analysis on Working Hours of Women Aged 25-45</td>
<td>23</td>
</tr>
<tr>
<td>1.8 Synthetic Control Analysis on Employment Rate of Women Aged 25-45</td>
<td>24</td>
</tr>
<tr>
<td>1.9 Synthetic Control Analysis on Labor Force Participation Rate of Mothers with Children Younger than 5 Years Old</td>
<td>25</td>
</tr>
<tr>
<td>1.10 Synthetic Control Analysis on Weekly Working Hours of Mothers with Children Younger than 5 Years Old</td>
<td>25</td>
</tr>
<tr>
<td>1.11 Synthetic Control Analysis on Labor Force Participation Rate of Mothers with 4-year-olds Only</td>
<td>26</td>
</tr>
<tr>
<td>1.12 Synthetic Control Analysis on Weekly Working Hours of Mothers with 4-year-olds Only</td>
<td>26</td>
</tr>
<tr>
<td>1.13 Synthetic Control Analysis on Labor Force Participation Rate</td>
<td>31</td>
</tr>
<tr>
<td>1.14 Synthetic Control Analysis on Weekly Working Hours</td>
<td>31</td>
</tr>
<tr>
<td>2.1 Low Birth Weight</td>
<td>47</td>
</tr>
<tr>
<td>2.2 Mortality Rate</td>
<td>47</td>
</tr>
<tr>
<td>2.3 Marginal Effects of IPW with Confidence Interval</td>
<td>49</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>2.4</td>
<td>Marginal Effects of IPW with Confidence Interval at Different Percentile of Poverty</td>
</tr>
<tr>
<td>A.1</td>
<td>Synthetic Control Analysis Result for Women Aged 20-45</td>
</tr>
<tr>
<td>A.2</td>
<td>Synthetic Control Analysis Result for Women Aged 25-35</td>
</tr>
<tr>
<td>A.3</td>
<td>Synthetic Control Analysis Result for Women Aged 25-40</td>
</tr>
<tr>
<td>A.4</td>
<td>Synthetic Control Analysis Result for Women Aged 25-50</td>
</tr>
<tr>
<td>C.1</td>
<td>Pre-k Enrollment Rate in Oklahoma</td>
</tr>
<tr>
<td>C.2</td>
<td>Synthetic Control Analysis on Oklahoma Pre-k Enrollment Rate</td>
</tr>
<tr>
<td>C.3</td>
<td>Fertility Rate of Oklahoma Women Aged 25-45</td>
</tr>
<tr>
<td>C.4</td>
<td>Synthetic Control Analysis on Fertility Rate of Women Aged 25-45</td>
</tr>
<tr>
<td>D.1</td>
<td>Synthetic Control Analysis on Full-time Labor Force Participation Rate of Mothers with Children Younger than 5 Years Old</td>
</tr>
<tr>
<td>D.2</td>
<td>Synthetic Control Analysis on Employment Rate of Mothers with Children Younger than 5 Years Old</td>
</tr>
<tr>
<td>D.3</td>
<td>Synthetic Control Analysis on Employment Rate of Mothers with 4-year-olds Only</td>
</tr>
<tr>
<td>D.4</td>
<td>Synthetic Control Analysis on Full-time Labor Force Participation Rate (Georgia)</td>
</tr>
<tr>
<td>D.5</td>
<td>Synthetic Control Analysis on Employment Rate (Georgia)</td>
</tr>
<tr>
<td>E.1</td>
<td>Synthetic Control Analysis on Labor Force Participation Rate (Below Poverty Line)</td>
</tr>
<tr>
<td>E.2</td>
<td>Synthetic Control Analysis on Weekly Working Hours (Below Poverty Line)</td>
</tr>
<tr>
<td>E.3</td>
<td>Synthetic Control Analysis on Labor Force Participation Rate (Above Poverty Line)</td>
</tr>
<tr>
<td>E.4</td>
<td>Synthetic Control Analysis on Weekly Working Hours (Above Poverty Line)</td>
</tr>
<tr>
<td>E.5</td>
<td>Synthetic Control Analysis on Labor Force Participation Rate (Married)</td>
</tr>
<tr>
<td>E.6</td>
<td>Synthetic Control Analysis on Weekly Working Hours (Married)</td>
</tr>
<tr>
<td>E.7</td>
<td>Synthetic Control Analysis on Labor Force Participation Rate (Unmarried)</td>
</tr>
<tr>
<td>E.8</td>
<td>Synthetic Control Analysis on Weekly Working Hours (Unmarried)</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>E.9</td>
<td>Synthetic Control Analysis on Labor Force Participation Rate (Differential Education Level—Lower than Highschool, High School, College and Above College by Order)</td>
</tr>
<tr>
<td>E.10</td>
<td>Synthetic Control Analysis on Weekly Working Hours (Differential Education Level—Lower than Highschool, High School, College and Above College by Order)</td>
</tr>
<tr>
<td>E.11</td>
<td>Synthetic Control Analysis on Labor Force Participation Rate (Fewer than 2 Children)</td>
</tr>
<tr>
<td>E.12</td>
<td>Synthetic Control Analysis on Weekly Working Hours (Fewer than 2 Children)</td>
</tr>
<tr>
<td>E.13</td>
<td>Synthetic Control Analysis on Labor Force Participation Rate (More than 2 Children)</td>
</tr>
<tr>
<td>E.14</td>
<td>Synthetic Control Analysis on Weekly Working Hours (More than 2 Children)</td>
</tr>
<tr>
<td>E.15</td>
<td>Synthetic control analysis on women (25-45) with less than high school education level</td>
</tr>
<tr>
<td>E.16</td>
<td>Synthetic control analysis on women (25-45) with high school education level</td>
</tr>
<tr>
<td>E.17</td>
<td>Synthetic control analysis on women (25-45) with college and above education level</td>
</tr>
<tr>
<td>E.18</td>
<td>Synthetic control analysis on women (25-45) with family income above poverty line</td>
</tr>
<tr>
<td>E.19</td>
<td>Synthetic control analysis on women (25-45) with family income below poverty line</td>
</tr>
<tr>
<td>E.20</td>
<td>Synthetic control analysis on married mothers</td>
</tr>
<tr>
<td>E.21</td>
<td>Synthetic control analysis on unmarried mothers</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table                                                   Page
1.1 Demographic and Economic Characteristics Before and After 1998 in Oklahoma 13
1.2 Summary Statistics of Selected Variables at State Level—OK vs. rest of US ... 34
1.3 Weights in the Synthetic Control Group ................................. 35
1.4 Mean Differences between Oklahoma and “Synthetic Oklahoma” and Oklahoma and the Rest of US States in Demographic Economic Characteristics (Partially) .......................................................... 36
2.1 Summary Statistics .......................................................... 57
2.2 Effect of IPW on Aggregate birth outcomes (2SLS) ....................... 58
2.3 Trade Exposure and Air Pollution Condition................................. 59
2.4 Trade Exposure and Number of Walmart Stores by County ............... 60
2.5 Effect of IPW on Aggregate Birth Outcomes (2SLS) with Interaction with County Level Per Capita Income .............................................. 61
2.6 Effect of IPW on Aggregate Birth Outcomes (2SLS) with Interaction with County Level Poverty Rate .................................................. 62
B.1 Full Set of Explanatory Variables and Measure of Closeness ............ 102
This is dedicated to Ranjun Li and Yumin Zhang, my dear parents and my best supporters forever.
CHAPTER 1
THE EFFECT OF UNIVERSAL PRE-KINDERGARTEN POLICY ON FEMALE LABOR FORCE PARTICIPATION — A SYNTHETIC CONTROL APPROACH

1.1. Introduction

In an effort to increase the human capital development of children, universal pre-kindergarten (pre-k) has become an increasingly popular policy response. While a substantial literature shows that universal pre-k improves child outcomes (Gormley and Gayer, 2005; Berlinski et al., 2009; Drange and Telle, 2015; Chor, Andresen and Kalil, 2016; Herbst, 2017), there is little analysis of the policy’s causal impacts on other household members. A well-known fact is that time-consuming home child care and costly day care are barriers to maternal labor force participation. Cohany and Sok (2007) show that the labor force participation rate of married women with young children declined during a time of increase in the overall female labor force participation rate. High-quality child care can provide young children a good start in education while allowing their parents to work through an increase of pre-k enrollment rate. It thus potentially breaks the cycle of intergenerational poverty and improves welfare for two generations simultaneously. Understanding the effects of universal pre-k on maternal labor market participation is also essential to uncover the potential mechanisms through which universal pre-k affects child outcomes. There might be a secondary effect of universal pre-k on child outcomes through the increase of maternal labor supply, resulting in an increase in total household income and parental investment in their children. Therefore, the study of causal relationship between universal pre-k and maternal labor supply is vital for complete welfare analysis of pre-k programs.

In this paper, I examine the impact of universal pre-k on maternal labor market behavior. This paper focuses on the high quality universal pre-k in Oklahoma as a case study. The expansion of universal pre-k should not only focus on the free assess to child care, but also on the quality of child care offered. However, the quality of universal pre-k programs is
implicit and thus difficult to quantify. However, the pre-k program of Oklahoma is generally considered to be a high quality program because of its government mandated small classroom size and highly educated teachers. To further investigate the quality effect of universal pre-k, I compare the empirical results to Georgia where the program is typically considered to be of relatively lower quality.

To show the importance of pre-k program quality, I first propose a theoretical model to investigate how quality and price changes affect maternal labor outcomes. The static labor supply model suggests that price reduction and quality improvement both increase maternal labor force participation at the extensive margin. While quality improvement unambiguously increases mothers’ hours of work, the policy effect along the intensive margin is ambiguous due to opposing income and substitution effects. The empirical study investigates the effects of universal pre-k on both extensive and intensive margin, including labor force participation rate, employment rate, the percentage change in full-time job participation and working hours.

The contribution of this paper is to use the synthetic control method (SCM) to examine the causal impact of universal pre-k on multiple labor outcomes of potential mothers. The major challenge in analyzing causal-relationship between universal pre-k and maternal labor force participation is to construct a credible counterfactual. This paper employs the SCM to construct a well-fitted control group. Before this paper, three identification strategies have been applied to link the expansion of child care programs to labor supply of mothers: an instrumental variable approach (Gelbach, 2002), a regression discontinuity approach (Fitzpatrick, 2010) and a difference-in-differences approach (Cascio and Schanzenbach, 2013).

The most commonly used empirical method in the universal pre-k literature is the difference-in-differences (DID) method. Comparing to DID, which constructs a control group with equally weighted untreated, the SCM is advantageous because it allows for the construction of a linear combination of unaffected states that minimizes the distance between Oklahoma and its synthetic control group in terms of observed pre-intervention characteristics. I find that the synthetic control group has a seemingly better fit than the equally weighted control group. Moreover, since Oklahoma is a small US state with more noisy

---

1Section 1.2 briefly introduces the history and major characteristics of Oklahoma universal pre-k.
data than larger US states, using the traditional SCM developed by Abadie and Gardeazabal (2003) and Abadie et al. (2010) does not provide a good counterfactual for Oklahoma. Doudchenko and Imbens (2017) suggest that allowing for negative weights may well improve the out-of-sample prediction. Therefore, I improve the SCM by allowing for negative weights on the unaffected units in the donor pool. I also include all lags of the dependent variables as control variables in the regression procedure to emphasize the importance of lags. In this setting, the improved SCM produces a much better counterfactual than the traditional SCM. Another advantage of the SCM is that the estimation results show the deviation between Oklahoma and “Synthetic Oklahoma” in each post-intervention year, thus providing a dynamic view of the treatment effect.

The empirical analysis uses the Current Population Survey (CPS) from 1980 to 2007. Oklahoma’s universal pre-k program begun in 1998. The results of this paper are as follows. Oklahoma universal pre-k increases the labor force participation rate, employment rate, and working hours of women of childbearing age. However, universal pre-k has little effect on full-time labor force participation of women of childbearing age. I also stratify the sample of potential mothers by four socioeconomic characteristics—education level, marital status, poverty status and numbers of children in the household. I find that married mothers, mothers with lower than high school and higher than college education, and mothers with no more than two children are more likely to increase their labor supply after the implementation of the universal pre-k policy. These findings suggest that a high-quality universal pre-k impacts on the labor market decisions of mothers from disadvantaged as well as non-disadvantaged backgrounds. The synthetic control analysis yields no effect of the 1998 Oklahoma universal pre-k policy on labor outcomes of mothers with younger than 5 years old children and mothers with 4-year-olds only. My finding shows that pre-k has positive effects on labor supply for all women of fertility age, but insignificant effects when focusing only on mothers with younger children. It suggests that women may be more inclined to work before having children since they expect to continue to work after having kids, but end up with staying at home once they actually have kids. The empirical analysis on Georgia universal pre-k presents no evidence of significant effects of Georgia universal pre-k on maternal labor outcomes. It is possible that the difference between the effects of Oklahoma and Georgia
universal pre-k is caused by quality variation of universal pre-k programs.

Broadly previous studies on the relationship between child care and maternal labor supply can be classified into two categories. The first strand of literature examines the labor supply elasticity of market prices of child care. Anderson and Levine (1999) find that the response of the female labor force participation rate to the price of child care is decreasing in education levels for women with children under age 13. Connelly and Kimmel (2003) find a significant negative effect of child care price on the employment of single mothers. Lundin et al. (2008) show that the maternal labor supply did not respond to a child care reform in Sweden which set a cap on the price of child care. Another strand of literature studies the labor market effects of child care policies, including child care subsidies and pre-kindergarten programs. Michalopoulos et al. (1992) show that a refundable child care tax credit would increase the labor supply of mothers. Tekin (2005) shows that single mothers are highly responsive to child care subsidies by increasing their employment while moving away from parental and relative care toward center care in the process. Blau and Tekin (2007) find that the subsidy recipients were about 13 percentage points more likely to be employed than nonrecipients.

Since the emergence of nationwide and statewide universal pre-k programs, there has been an increasing interest in the estimation of the effects of universal pre-k on maternal labor market decisions. The empirical evidence from non-US countries generally indicates a remarkable increase of maternal labor supply with a nationwide expansion to universal pre-k. Schlosser (2005) finds that introducing free public preschool to 3- or 4-year-old children in Israel increased the labor supply of mothers by about seven percentage points. Goux and Maurin (2010) also find that early school availability to 2- and 3-year-old children in France had a significant employment effect on single mothers. Baker et al. (2008) show that the labor supply of married women rose about 14.5 percent with the highly subsidized and universally accessible child care program, “$5 per day child care” in Quebec, Canada. Simonsen (2010) shows that a price increase of 1 euro per month decreases employment by 0.08%, which corresponds to a price elasticity of -0.17 using Danish data. Koebel and Schirle (2016) show that the Canadian Universal Child Care Benefit has significant negative effects on the labor supply of legally married mothers but positively affects the labor supply of single mothers.
The US evidence, however, shows mixed results. Cascio and Schanzenbach (2013) find that universal pre-k programs had a positive impact on the employment rate of less educated mothers with 4-year-olds only. Fitzpatrick (2010) shows that the universal pre-k increased pre-k enrollment by 12 to 15 percent, but had no robust impact on maternal labor supply, she also finds increased labor supply in rural families. Fitzpatrick (2010) uses the regression discontinuity (RD) method to address the causal relationship. This method, however requires information on birth months of Oklahoma children and the sample size is much smaller than Cascio and Schanzenbach (2013). Herbst (2017) finds that maternal employment increased substantially after the US Lanham Act of 1940, a heavily subsidized and universal childcare program during World War II.

The rest of the paper is organized as follows. Section 1.2 describes the background of US pre-k programs and the Oklahoma universal pre-k policy. Section 1.3 provides a static labor supply model to examine how price and quality changes affect the consumption of market child care and the maternal labor supply. Section 1.4 describes the data used for this analysis. In section 1.5, I describe the synthetic control method. Section 1.6 presents and discusses the results from the synthetic control analysis. Section 1.7 concludes.

1.2. Background on US Pre-kindergarten Education Programs and Oklahoma Universal Pre-k

Since the President’s 2013 State of the Union address, 34 states have increased funding for their preschool programs, amounting to over $1 billion in new state resources dedicated to early education. Nowadays, expanding access and improving the quality of pre-k programs become major concerns of policymakers in early education. Before the universal pre-k, several federal and statewide pre-k programs have been established to help children from low-income families. Since 1935, the federal government has supported early care and education for poor children to promote their healthy development and give them a better chance to succeed. But the improvement in early education of the past 80 years is slow due to “a haphazard array of uncoordinated programs, shaped by outdated science and entrenched political interests,

and long driven by addressing unintended consequences of previous policies rather than core goals."  

The federal government now funds dozens of small programs providing services to children from birth through age five, but the preponderance of federal funds—$17.2 billion—is spent on three major programs: Head Start at $9.2 billion, the Child Care Development Fund (CCDF) at $5.4 billion, and child care expenditures from Temporary Assistance for Needy Families (TANF) at $2.6 billion annually. All three programs fund poor children’s participation in early care and education. Since those early child care programs are almost all highly targeted on poor families, nowadays, policymakers and economists are interested in expanding traditional public pre-k to a universal level, so that every pre-k age child (3- or 4-years-old) would have access to free public pre-k.

In the state level, several states have been on the path to funding universal pre-k during the past two decades. Georgia first established a universal pre-k program in 1995, followed by Oklahoma in 1998, Florida in 2005 and then Illinois in 2006. California and New York also started to establish universal pre-k in the 1990s, but it has not yet been implemented statewide due to budget and political issues. In 2014, of the 41 states with state-funded pre-k programs (including the District of Columbia), only nine served more than half of all 4-year-olds statewide, and eleven states served less than ten percent of all 4-year-old children. Only three states—Georgia, Oklahoma and Florida—are believed to “truly” have universal pre-k programs in terms of their pre-k enrollment rates.  

The Oklahoma universal pre-k program is in high quality and is believed to be a successful example. In the spring of 1998, House member Joe Eddins and state senator Penny Williams secured approval to amend the school formula so that four-year-olds would be included in the school funding formula. Since 1998, Oklahoma has provided universal access to public pre-kindergarten. Children in Oklahoma who turn 4 years old on or before September 1st are eligible for the public pre-kindergarten program. The Oklahoma universal pre-k program

---

offers two options: full day or half day, both of which provide high standard curriculum for young children. With 74 percent of all 4-year-olds enrolled in pre-k programs in 2014, Oklahoma maintains a high enrollment rate.

Besides the free access for every pre-k age child and the high enrollment rate, the quality of Oklahoma universal pre-k is also a remarkable feature. This is primarily based on the high quality of teachers and the small group size. All teachers must have college degrees and certificates in early childhood education and receive the same compensation as teachers in public elementary schools. In regard to the classes, the group size is set to not exceed 20 and the maximum child to staff ratio is set at 10 to 1. Additionally, National Institute for Early Education (NIEER) reports that the Oklahoma universal pre-k has a generous expenditure of almost $7,427 per child, while the spending per child in Georgia is $3,490 and the average annual pre-k expenditure per child in the US is no more than $5,000.

1.3. The Model

In this section, I present a static labor supply model to investigate how the price and quality changes of a pre-k program affect maternal labor market decisions. Assume a mother can always find a job and her working hours are perfectly flexible, the utility maximization problem is given by

\[
\max_{X,H} U(X, Q, L) \tag{1.1}
\]

\[
\text{s.t. } X + P_m H = WH + Y \tag{1.2}
\]

\[
H + L = T \tag{1.3}
\]

\[
Q = \frac{H}{T} Q_m + \frac{T - H}{T} Q_h \tag{1.4}
\]

In this setup, a mother’s utility \( U(X, Q, L) \) is derived from three sources: consumption of numeraire good \( X \), average quality of child care \( Q \), which is defined as the time weighted average of the market child care (day care) and the home child care following Michalopoulos et al. (1992), and leisure time \( L = T - H \), where \( T \) is the time endowment and \( H \) is number

---

of hours worked per day.

Equation (1.2) is the budget constraint. \( W \) is the hourly wage rate of working mothers and \( P_m \) is the price of the market pre-k per hour. Assume that the utility of other family members is exogenous to the utility function of mothers. I also take the income of other family members as part of mother’s exogenous non-labor income \( Y \). Further, I assume that working mothers have to purchase market pre-k or participate in a public pre-k program during their work time, i.e., no one else in the family will take care of their children when the mothers are at work. Rewrite equation (1.2), we have

\[
X = (W - P_m)H + Y = \tilde{W}H + Y
\]  

(1.5)

where \( \tilde{W} \) is the hourly wage rate net of market pre-k cost.

Equation (1.3) is the time constraint. Assume that mothers spend all their leisure time on child care at home, so that the time spent on home child care is \( T - H \). Then the average quality is given by Equation (1.4), where \( Q_m \) is the quality of market child care and \( Q_h \) is the home child care quality provided by mothers. Finally, assume no market child care is better in quality than home child care provided by mothers, such that \( Q_m = \alpha Q_h \) and \( 0 < \alpha < 1 \). Therefore, we can rewrite equation (1.4) as

\[
Q = \frac{H}{T} \alpha Q_h + \frac{T - H}{T} Q_h \\
= [1 - (1 - \alpha) \frac{H}{T}] Q_h
\]

(1.6)

We can observe that mothers are affected by the universal pre-k policy through price and quality changes on both the extensive and the intensive margins.

The working decision of a mother at the extensive margin depends on whether the utility from working exceeds the utility from staying at home and taking care of her child (children). Let \( D \) denote the working decision of the mother. For the working decisions of mothers on the extensive margin, \( D = 1 \) denotes a mother chooses to work and \( D = 0 \) demotes the
mother chooses to stay at home. Therefore we have,

\[ D = \begin{cases} 
1 & \text{if } U|_{H=H^*,X=X^*,Q=Q^*} - U|_{H=0,X=0,Q=0} > 0 \\
0 & \text{otherwise}
\end{cases} \]

where \( U|_{H=0,X=0,Q=0} \) is the non-working utility and \( H^*, X^* \) and \( Q^* \) are the utility-maximizing values of each variable conditional on working.

Similarly, for the working decisions of mothers on the intensive margin, \( D = 1 \) denotes a mother chooses to work more hours and \( D = 0 \) denotes the mother chooses to work as long as before the policy implementation. Therefore we have,

\[ D = \begin{cases} 
1 & \text{if } U|_{H=H^*,X=X^*,Q=Q^*} - U|_{H=H^*,X=0,Q=0} > 0 \\
0 & \text{otherwise}
\end{cases} \]

where \( U|_{H=H',X=0,Q=0} \) is the working utilities before and after the universal pre-k policy and \( \epsilon \) is a positive number of working hours.

The threshold of entering the labor market is quantified by a reservation wage. Equation (1.2) and (1.5) indicate that the price of market pre-k \( P_m \) affects a mother’s working decision through the net wage \( \tilde{W} \). When the implementation of universal pre-k leads to a price reduction, the net wage \( \tilde{W} \) increases and it is more likely to exceed the mother’s reservation wage. Either \( \tilde{W} \) is higher than the reservation wage and mothers shift from staying at home to working, or \( \tilde{W} \) is still below the reservation wage and mothers remain at home.

Besides the large price reduction, quality improvement of a pre-k program would also have an impact on mother’s working decisions. Equation (1.4) suggests that the quality of market pre-k \( Q_m \) affects a mother’s utility function through the allocation of working hours and leisure. For a given market pre-k price, when the quality of market pre-k or public pre-k improves, the parameter \( \alpha \) in equation (1.6) increases, so that the utility loss from lower-quality child care is smaller and the gap between mothers’ net wage and reservation wage decreases. Hence, for the extensive margin, quality improvement increases the likelihood of working.
However, for working mothers, a rise in the net wage $\bar{W}$ is well known to have a theoretically ambiguous effect on working hours due to the trade-off between the negative income effect and positive substitution effect. For the effect of quality improvement, as $\alpha$ increases, the average quality of childcare $Q$ is higher, the marginal utility from home child care decreases, and mothers will allocate less time to home child care and correspondingly allocate more time to work. Thus, quality improvement increases mothers’ working hours.

In summary, universal pre-k should unambiguously increase the maternal labor force participation rate, especially when the program is in high quality. However, the model does not make a decisive prediction on mothers’ working hours, although the effect is more likely to be positive when the program quality is higher.

1.4. Empirical Strategy

The empirical method employed in this paper is the synthetic control approach, first introduced by Abadie and Gardeazabal (2003) and developed by Abadie et al. (2010). The SCM allows for estimation in settings where a single unit is exposed to an event. It provides a data-driven procedure to construct a synthetic control unit that approximates the characteristics of the treated unit. In this section, I will briefly introduce the SCM used to analyze the effect of the 1998 Oklahoma universal pre-k on female labor market decisions.

Suppose we observe $S + 1$ states, one of which is our treated state, Oklahoma, which we call state 1. Let $Y_{it}^N$ denote the outcomes of interest that would be observed for state $i$ at time $t$ in the absence of the intervention, where $i = 1, \ldots, S + 1$, $t = 1, \ldots, T$. Let $T_0$ be the number of pre-intervention periods and $1 \leq T_0 < T$. Let $Y_{it}^I$ be the outcome that would be observed for state $i$ at year $t$ if state $i$ is exposed to the universal pre-k policy in period $T_0 + 1$ to $T$. Two assumptions are needed for the synthetic control method.

**Assumption 1: No anticipation effects**

The intervention has no effect on the outcomes before the implementation period $t \in \{1, \ldots T_0\}$. Under Assumption 1, for $t \in \{1, \ldots T_0\}$ and all $i \in \{1, \ldots S + 1\}$, we have $Y_{it}^N = Y_{it}^I$. 

10
Assumption 2: No interference on untreated units

There is no interference between treated and untreated states, the outcomes of the untreated states are not affected by the intervention implemented in the treatment state.

This estimated treatment effect is $\alpha_{1t} = Y_{1t}^I - Y_{1t}^N$. Since $Y_{1t}^N$ is unobserved, we need to estimate the counterfactual $Y_{1t}^N$ with the following factor model

$$Y_{1t}^N = \delta_t + \theta_t Z_1 + \lambda_t \mu_1 + \varepsilon_{1t} \tag{1.7}$$

where $\delta_1$ is an unknown state fixed effect, $Z_1$ is a $(r \times 1)$ vector of observed covariates (not affected by the intervention), $\theta_t$ is a $(1 \times r)$ vector of unknown parameters, $\lambda_t = (\lambda_{t1}, \lambda_{t2}, \ldots, \lambda_{tF})$ is a $(1 \times F)$ vector of unobserved time fixed effect for $t = 1, \ldots, F$, $\mu_1 = (\mu_{11}, \mu_{12}, \ldots, \mu_{1F})$ is an $(F \times 1)$ vector of unobserved factor loading for the treated state Oklahoma, and the error terms $\varepsilon_{1t}$ are unobserved transitory shocks at the state level with zero mean. Note that we will obtain the difference-in-differences (fixed effect) model from equation (1.7) if $\lambda_t$ is constant over time.

The synthetic control group is obtained by assigning weights $\omega = (\omega_2, \ldots, \omega_{s+1})$ to each untreated unit in the donor pool. The value of the outcome variable for each synthetic control indexed by $\omega$ is $\sum_{s=2}^{S+1} \omega_s Y_{st} = \delta_t + \theta_t \sum_{s=2}^{S+1} \omega_s Z_s + \lambda_t \sum_{s=2}^{S+1} \omega_s \mu_s + \sum_{s=2}^{S+1} \omega_s \varepsilon_{st}$. The optimal weight vector $\omega^*$ is the one that minimizes the distance between the pre-intervention observed characteristic vector $X_i$ of the treated state and the selected control group, $||X_1 - X_0 W||$, where $X_i$ includes both covariates $Z_i$ and outcome variables $Y_i$, $X_1$ is the pre-intervention observed characteristics of the treated unit and the $X_0$ is the same observed variable set of the untreated states.\(^7\)

Abadie et al. (2010) impose three restrictions in analyzing the effect of California’s tobacco control program—no intercept, the sum of weights add up to 1, and all weights are non-negative. In this paper, to obtain a better counterfactual for the treated state, I relax the non-negativity assumption ($\omega_s \geq 0$, $s = 2, 3, \ldots, n + 1$). Doudchenko and Imbens (2017) claim that allowing for negative weights may well improve the out-of-sample prediction. First, allowing for weights on the observed characteristics of the untreated unites may be

\(^7\)See details on optimal weight vector selection on Abadie et al. (2010)
able to better fit those of the treated unit. Doudchenko and Imbens (2017) show an example when there is one treated state and two control states and the key characteristic is the share of young people. If the share of young people in the treated state is 2/3, and the shares of the control states are 1/2 and 1/3, assigning weights 2 and -1 respectively to the two control states can produce the identical share of young people to the treated states. The second reason is negative weights can help with bias-reduction as the bias goes to zero in a faster rate in settings with many covariates to be matched on.\textsuperscript{8} I still keep the no intercept and sum up to 1 restrictions in the minimization procedure because imposing these two restrictions helps produce a unique weight matrix, though Doudchenko and Imbens (2017) claims no restrictions are necessary for the SCM. The intuition behind the negative weights is that the control group is constructed by some virtual states, which have characteristics opposite to some real US states.

For inference, Abadie et al. (2010) suggest using placebo tests to measure the significance of estimates. The basic idea behind the placebo tests is to apply the synthetic control method to all the control units in the donor pool as if they were also exposed to the universal pre-k policy and test whether the treated unit behaves in a significantly different way from unexposed units. Under the null hypothesis that policy intervention has no impact on the treated unit, the estimate for the treated unit is expected to lie within the distribution of the placebo estimates. I also apply the pre/post rooted mean squared prediction error (RMSPE) ratio test to measure the significance of treatment effects. The pre/post RMSPE ratio test is an extension of the placebo tests, it measures the closeness between the observed variables of the treated unit and the synthetic control group before and after the policy intervention. Therefore, we can compare the pre/post RMSPE ratio of the treated unit to that of the untreated unit to examine whether there is a relative larger increase in post RMSPE.

1.5. Data

The primary data set is the March Current Population Survey (CPS), which provides detailed labor statistics and demographic characteristics for individuals and households in the US annually. The entire study period is from 1980 to 2007 to cover a long enough pre-

\textsuperscript{8}See Abadie and Imbens (2011).
intervention period before the intervention year 1998 and the sample ends in 2007 to avoid the effect of the financial crisis and more recent expansion of pre-k in other states. Since the enrollment of Oklahoma universal pre-k starts at the end of February each year, in case the labor outcomes of women in Oklahoma is immediately affected by the new policy in 1998, the post-intervention period is set to start from 1998.

<table>
<thead>
<tr>
<th></th>
<th>Before 1998</th>
<th>After 1998</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor force participation rate</td>
<td>0.71</td>
<td>0.74</td>
</tr>
<tr>
<td>Working hours</td>
<td>28.67</td>
<td>28.94</td>
</tr>
<tr>
<td>Family total income</td>
<td>32716.80</td>
<td>55539.84</td>
</tr>
<tr>
<td>Mother’s age: 25-30</td>
<td>0.25</td>
<td>0.21</td>
</tr>
<tr>
<td>Mother’s age: 30-35</td>
<td>0.24</td>
<td>0.23</td>
</tr>
<tr>
<td>Mother’s age: 35-40</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Mother’s age: 40-45</td>
<td>0.26</td>
<td>0.31</td>
</tr>
<tr>
<td>Mother’s education level: below high school</td>
<td>0.11</td>
<td>0.10</td>
</tr>
<tr>
<td>High school graduates</td>
<td>0.40</td>
<td>0.34</td>
</tr>
<tr>
<td>College graduates</td>
<td>0.44</td>
<td>0.50</td>
</tr>
<tr>
<td>Above college degree</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>White</td>
<td>0.86</td>
<td>0.76</td>
</tr>
<tr>
<td>Black</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>Other race</td>
<td>0.07</td>
<td>0.15</td>
</tr>
<tr>
<td>Family size</td>
<td>3.45</td>
<td>3.45</td>
</tr>
<tr>
<td>Food Stamp Recipients</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>Children enrolled in free lunch program</td>
<td>0.35</td>
<td>0.43</td>
</tr>
<tr>
<td>Below poverty line</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>Number of observations</td>
<td>6060</td>
<td>3331</td>
</tr>
</tbody>
</table>

Table 1.1. Demographic and Economic Characteristics Before and After 1998 in Oklahoma
The major outcomes of interest are labor outcomes of potential mothers (women aged 25 to 45), mothers with younger than 5 year olds and mothers with 4 year old children only after the universal pre-k policy. There are three reasons to focus on the sample of potential mothers, defined as women between 25 to 45 years old, as the primary sample of interest. First, the provision of free and high-quality pre-k shortens the duration of utilization of day care and/or home care, resulting in a higher possibility of females to stay in their current jobs before their children turn 3 or 4 years old. Therefore, there might be a spillover effect on mothers whose children is younger than 4 years old. In addition, universal pre-k may also have an effect on fertility decisions and it will thus affect the labor market outcomes of all women of childbearing age. Prior studies have shown that the increase of child care subsidies or the expansion of child care programs would increase fertility rates (Blau and Robins, 1989; Baughman and Dickert-Conlin, 2003; Haan and Worhlich, 2011; Bauernschuster et al., 2015). Third, the sample size of mothers with only 4-year-olds is small in the Current Population Survey (CPS), especially for a small US state like Oklahoma, and the sample size issue would be more severe in the study of heterogeneous effects by subgroups.

The labor outcomes includes labor force participation rate, percentage change of full-time labor force participation, hours-of-work for working mothers and employment rate. The major labor outcomes following the model in Section 1.3 are labor force participation rate as the outcome on the extensive margin and working hours as the outcomes on the intensive margin. Universal pre-k is expected to increase the likelihood that a mother is employed since mothers who have free full-time child care options are better able to work. Employment is also restricted by job availability though. It is possible that mothers who are willing to work cannot find jobs, especially after years of unemployment. Therefore, the empirical analysis is needed to investigate the effect on employment rate. The percentage change in full-time working mothers, defined as the ratio of full-time working mothers to the whole working sample of mothers, is a proxy of change in job type. When a high-quality universal pre-k policy is implemented, the marginal rate of substitution between the utility...

---

9In Appendix A, I present the synthetic control method analysis on potential mothers in different age ranges, including women aged 25-35, 25-40, 20-45, 25-50, the results are robust for slightly narrowing or expanding age groups, I choose women aged 25 to 45 mainly because this group is the commonly used sample of fertility age women.
of working and home child care is smaller and thus working mothers are more likely to switch from part-time jobs to full-time jobs. Unlike the change of working hours, percentage change in full-time working females could also indicate a change in type of jobs.

The CPS defines an individual to be ‘in the labor force’ if she is employed or unemployed, so that those who are at school, retired and staying at home are not in the labor force. The employment rate is the ratio of those who are employed to the population of females in the labor force. The ‘working hours’ variable in this paper is defined as how many hours an individual work (not including zero) the week before the survey and percentage of full-time labor force participation is directly from the indicator variable of full-time or part-time employment. The studied outcomes of interest are state level labor statistics obtained from aggregating the individual level data in CPS.

Since the dependent variables are state level labor outcomes and the universal pre-k policy takes place at the state level, it is convenient to estimate the effect in aggregate level. Meanwhile, the synthetic control method is good for the comparative case study in aggregate entities or administrative areas. Thus I use the individual CPS data to form state level controls such as the fraction of population by race and education level. CPS personal sample weights are used in the data aggregation. The sample size of the women aged 25 to 45 in the CPS is 1,478,181 for all states from 1980 to 2007. The state level data used for the synthetic analysis has a sample size of 1,428, which comes from the multiplication of the number of states and districts in the CPS (51) and the number of study years (28). The corresponding set of explanatory variables \( \{Z_i\} \) consists of demographic characteristics such as age, race and economic characteristics including total family income, personal income, family size and spouse’s working hours.\(^{10}\) To capture state level shocks to female labor outcomes and to better construct the synthetic control group, I also take the labor outcomes of women aged 45 to 60 as an additional control variable. In addition, I add state gross domestic product (GDP) into the set of explanatory variables \( \{Z_i\} \), obtained from another data source—Bureau of Economics Analysis (BEA). The CPS data also shows high serial correlations in female labor outcomes. For example, the correlation between the female labor force participation rate and its one-year lagged variable is as high as .8877. To deal

\(^{10}\)The full set of the explanatory variables is listed in Appendix B.
with this issue, I take the lags of outcome variables in each pre-intervention period into the control variable set. Therefore, lags of dependent variables are given more importance in the synthetic control analysis relative to other controls in \( \{Z_i\} \).

Table 1.1 summarizes the individual level data to show the before and after differences in the demographic and economic characteristics in Oklahoma. There are a number of things that are worth pointing out from the table. First, the statistics show that the labor force participation rate increased after 1998, but working hours seemed to remain the same. Second, family total income rose dramatically from about 32k to 55k on average.\(^{11}\) Table 1.1 also indicates that a larger percentage of females are pursuing higher level of education. Next, as expected, the demographic characteristics in the summary statistics were stable over time; whereas only the racial composition had a notable change. Finally, although the number of free lunch recipients increased after 1998, the fractions of food stamp recipients and households below the poverty line did not change much in Oklahoma.

Table 1.2 summarizes the state level statistics for Oklahoma and the rest of US before and after 1998. It suggests that, compared to the rest of US states, women aged 25 to 45 in Oklahoma were less likely to participate in the labor market and worked fewer hours per week. The GDP difference shows that the economy of Oklahoma was below the average level of other US states. Hence, it should not be surprising that the labor force participation rate of Oklahoma was lower than average. Table 1.2 also shows that Oklahoma had a higher fraction of lower-educated women, food stamp and free lunch recipients, and high-poverty households. As for the racial composition, Oklahoma had more non-whites than other states. Ultimately, when controlling for time difference, the summary statistics in Table 1.2 appear to show a non-substantial change in the labor force participation rate of fertility age women in Oklahoma relative to the simple average of the rest of US.

\(^{11}\)In the CPS data, the family income is not adjusted by CPI, for reference, $32k in 1990 is about $42k in 2000.
1.6. Empirical Estimates

1.6.1. Main Results

Now I estimate the impact of the 1998 universal pre-k policy on Oklahoma female labor force participation. The main results presented below investigate the labor outcomes of childbearing age women (women aged 25-45) and mothers with young children. In this section, I first present the effects of universal pre-k on the labor force participation rate of fertility age women to further illustrate the empirical strategy. Then I show the estimation effects on three other labor outcomes of potential mothers in Oklahoma—percentage of full-time labor force participation, average weekly working hours and employment rate. Lastly, I repeat the empirical analysis on mothers with younger than 5-years-olds and mothers with 4-year-olds only to compare with previous evidence.

![Graph](image)

Notes: The graph presents the labor force participation rate each year. The solid black line stands for the trend of labor force participation rate of Oklahoma and the dash line represents the trend of labor force participation rate of the rest of US states in average from 1980 to 2007 excluding Georgia, who had established universal pre-k policy in 1995.

Figure 1.1. Maternal Labor Force Participation Rate—Oklahoma vs. the Rest of US States (excluding Georgia)
Before applying the synthetic control method, Figure 1.1 plots the labor force participation rates of potential mothers in Oklahoma compared to a naive control constructed by assigning equal weights to all unaffected US states. Note that before 1990, the naive control group appears to be a good control, however, after 1990 and up to 1998, there is a significant gap. Therefore, using equally weighted untreated US states would not be a good strategy to investigate the policy of interest.

Note: The vertical dash line indicates the policy intervention year 1998, the starting date of the universal pre-k policy was September 1st, 1998. I treat 1998 as the first year of post-intervention period, the results are robust if 1998 is taken as a pre-intervention year.

Figure 1.2. Maternal Labor Force Participation Rate (25-45 sample)—Oklahoma vs. Synthetic Oklahoma (Traditional SCM)

Figure 1.2 presents the synthetic control without allowing for negative weights on the untreated units. Comparing to the simple average control in Figure 1.1, which is normally used in the difference-in-differences approach, the traditional SCM provides a better counterfactual in the pre-intervention (pre-1998) period. The better fit from employing the SCM is due to the fact that the SCM does not assume equal weight to each untreated unit in the control group and it takes the lags of outcome variables in minimizing the distance between the observed characteristics of treatment and control. Except for the period 1992 to 1994, labor force participation rates of the synthetic control group in most years before 1998 were generally close to labor force participation rates of Oklahoma. The huge drop in labor force
participation rate may result from measurement error and small sample bias. In that case, the low labor force participation rate in 1993 in Oklahoma bring noise to the data analysis. Another possible explanation for this big drop is that the oil price crisis in the early 1990s had a stronger effect on Oklahoma’s economy. Historical statistics show that in each oil price decline (defined as inflation-adjusted oil prices falling nearly continually by more than 30 percent and more than $20 a barrel), the employment rate in the oil and gas sector in Oklahoma alone decreased by at least seven percent.\footnote{Chad Wilkerson, “How will Oklahoma be affected be decline of oil price?”, March 11, 2015. Federal Reserve Bank of Kansas City, Denver, Oklahoma City, Omaha.}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure13.png}
\caption{Maternal Labor Force Participation Rate (25-45 sample)—Oklahoma vs. Synthetic Oklahoma (SCM allowing for negative weights)}
\end{figure}

\textit{Note:} Same as Figure 1.2

In order to obtain a better counterfactual, I allow for negative weights in constructing the control group and add labor force participation rates from 1980 to 1997 into the control variable set to emphasize on the strong correlation between the outcome variable and its lags. Though allowing for negative weights in the synthetic control method may require more extrapolation from the data, it works better in pre-trend fitting for Oklahoma. Figure 1.3 shows the results. It obviously shows that the synthetic control with negative weights and more importance on the outcome variables provides a better counterfactual than the
traditional synthetic control. Figure 1.3 suggests that the labor force participation rates of potential mothers in Oklahoma are higher than the labor force participation rates of potential mothers in the synthetic control group in the post-intervention period.

Table 1.3 provides the weights of each state used to construct the “Synthetic Oklahoma”. A state will be assigned zero weight if it is not chosen to construct the synthetic control group. Note that Georgia is eliminated from the donor pool of the unaffected states, hence the synthetic control group is ensured to be untreated before the policy intervention year 1998.

![Graph](image)

*Note:* All lines represent the distribution of estimated gaps between the treated unit and non-intervened control states. The black line stands for the estimated gaps of Oklahoma and the estimated gaps of the placebos are indicated by grey lines.

**Figure 1.4. Placebo Tests on Synthetic Control Results of Oklahoma**

Figure 1.3 shows that the trend in female labor force participation rate of “Synthetic Oklahoma” well-matches that of Oklahoma in the pre-intervention period. Table 1.4 further displays the closeness of the observable characteristics between Oklahoma and “Synthetic Oklahoma”, these variables are used to construct the synthetic control. I also list the differences in observable characteristics between Oklahoma and the simple average of unaffected US states in Table 1.4 for comparison. It shows that the observed charateristics of the synthetic control group closely match the observed charateristics of Oklahoma.
Now regarding the post-intervention period, the female labor force participation rate in Oklahoma shows no break from the pre-intervention period. However, Figure 1.3 suggests a large distance between the labor force participation rates of Oklahoma and “Synthetic Oklahoma” after the policy implementation. Moreover, the labor force participation rate of Oklahoma is larger than that of the synthetic control group in almost every post-intervention year. Abadie et al. (2010) suggest to use placebo tests to investigate the significance of empirical results obtained from the SCM. The nature of these tests is to conduct a series of placebo studies by iteratively applying the SCM to states other than Oklahoma. Figure 1.4 shows the placebo tests of the synthetic control estimation. The lines in Figure 1.4 represent the distribution of estimated gaps between the treated unit and their synthetic control group in labor force participation rate. The gap line for Oklahoma is in black and the gap lines for other states are in grey. As expected, it is found that the estimated gap between Oklahoma and “Synthetic Oklahoma” generally lies above most of the placebos. Therefore, the placebo test shows a statistically significant and positive impact of the Oklahoma universal pre-k policy on the labor force participation rate of women of childbearing age.

![Figure 1.5. Pre/post RMSPE Ratio of Female Labor Force Participation Rate by States](image)

Besides the placebo tests, I also apply the pre/post RMSPE ratio test for inference. The pre/post RMSPE ratio test is an extension of the placebo tests, but using numbers
rather than a graph to compare the pre-post difference. In the synthetic control method, the rooted mean squared prediction error (RMSPE) is used to measure the closeness between the observed variables of the treated unit and the synthetic control group, it is defined as \( \sqrt{\frac{1}{T} \sum_{t=0}^{T} e_t^2} \), where \( e = |X_{\text{treated}} - X_{\text{synth}}| \) is the distance between the treated unit and the synthetic control group in the value of variable X. The pre/post RMSPE ratio \( \sqrt{\frac{1}{T_0} \sum_{t=0}^{T_0} e_t^2} / \sqrt{\frac{1}{T-T_0+1} \sum_{t=T_0+1}^{T} e_t^2} \) is the relative measure of the pre-intervention and the post-intervention difference in observed characteristics. The pre/post RMSPE ratio of the treated unit is compared with the ratios of the placebos. A relatively smaller pre/post RMSPE means the post-intervention difference between the treated unit and the synthetic control group is larger than pre-intervention difference, thus it is likely to show a significant treatment effect. Figure 1.5 shows the pre/post RMSPE ratio of labor force participation rates of all US states. Oklahoma has a small pre/post RMSPE ratio, and the pre/post RMSPE ratio of Oklahoma is one of the smallest among all 50 US states and the District of Columbia. It suggests that the difference in the female labor force participation rate between Oklahoma and the synthetic control group before the implementation of the universal pre-k policy is smaller than the post-intervention difference. In other words, the treatment effect relative to pre-intervention difference is large.

Next, I examine the response of other three labor outcomes of women aged 25 to 45 to the universal pre-k policy: percentage change of full-time labor market participation, employment rate and mean weekly working hours. I first examine the policy effect on the percentage change of full-time job participation, which indicates not only a change in labor force participation but also a change in job type. Figure 1.6 shows that the effect of universal pre-k on the percentage of full-time female workers aged 25 to 45 in Oklahoma is statistically insignificant. There is no great difference between the full-time job participation rates of potential mothers in Oklahoma and the synthetic control group after 1998, and the placebo tests also show the gap in full-time job participation rate of fertility age women in Oklahoma is generally inside the range of placebos.

Then I examine the effect on hours of work of working mothers. Since the theoretical model has an ambiguous solution on evaluating the effect of universal pre-k on the intensive margin, it is especially useful to utilize empirics. The synthetic control estimation in Figure
1.7 shows that the post-trend of average working hours in Oklahoma is opposite to the post-trend of average working hours in the synthetic control group. The working hours of women aged 25-45 appear to decrease right after the universal pre-k policy and then increase to be much higher than the average working hours of the synthetic control group. This suggests that the universal pre-k increases working hours of women aged 25-45 in the long run.

The last labor outcome to analyze is state level employment rate of women aged 25 to 45. Figure 1.8 shows a higher employment rate of potential mothers in Oklahoma than that in the synthetic control group for most of the post-intervention years. The placebo tests also suggest the gap between the employment rate in Oklahoma and the synthetic control is generally large and above zero, but the differences in employment rates of potential mothers between Oklahoma and the synthetic control group in 2001 and 2006 are not significantly different from zero.
Figure 1.8. Synthetic Control Analysis on Employment Rate of Women Aged 25-45

At last, to compare with previous studies, I apply the synthetic control method on mothers with younger than 5 years old children and the direct policy-affected sample—mothers with 4-year-olds only. In this section, I will only show the effects of universal pre-k on the labor force participation rate and mean weekly working hours of the two samples, the estimation results of the other two labor outcomes are presented in Appendix D. Figure 1.9 and Figure 1.10 show the effects of universal pre-k on the labor force participation rate and weekly working hours of mothers with 5 years old children, respectively. Surprisingly, the results suggest no statistically significant effects of the universal pre-k policy on the extensive and intensive margin labor outcomes of mothers with younger than 5-year-olds. Figure 1.9 shows that the post-intervention labor force participation rates of mothers in Oklahoma are higher than the synthetic control group, but the distance is not significantly different from zero. The labor force participation rate of mothers with younger than 5 years old children increased in the first two years after the policy implementation and then dropped back to level almost identical to the synthetic control group. Figure 1.10 shows an increasing post-trend of mean weekly working hours of mothers with younger than 5 years old children in Oklahoma, however, the synthetic control group also follows an increasing trend close to Oklahoma. The placebo tests also show that the gap between mean weekly working hours of mothers with younger than 5 years old children in Oklahoma and its synthetic control group is not significantly different from zero.

For the sample of mothers with 4 years old children only, first note that sample size of mothers with 4 years old children is small, especially for small states. Also note that the study time period of this sample is different from the sample of women of childbearing
Figure 1.9. Synthetic Control Analysis on Labor Force Participation Rate of Mothers with Children Younger than 5 Years Old

Figure 1.10. Synthetic Control Analysis on Weekly Working Hours of Mothers with Children Younger than 5 Years Old

age. I match mothers with 4-year-olds using their household ID, which is not available in IPUMS-CPS prior to 1988. Figure 1.11 presents the synthetic control analysis on the labor force participation rate of mothers with 4-year-olds only. Except for a large decrease in 2005, the labor force participation rate in Oklahoma is not significantly different from that of the synthetic control group in the post-trend. This is also confirmed by the placebo tests in the right panel. In regard to the intensive margin effect, Figure 1.12 shows the weekly working hours of mothers with only 4 years old children in Oklahoma does not follow a clear increasing and decreasing pattern. In fact, the mean weekly working hours of mothers with only 4 years old children in Oklahoma moves up and down cyclically. Therefore, there is no clear break from the pre-trend, and there is no evidence of significant effect of universal pre-k on mean weekly working hours of mothers with 4-year-olds only. Thus, different from previous literature, this paper suggests that the effect of universal pre-k on the labor
outcomes of the policy-targeted sample is statistically insignificant.

Figure 1.11. Synthetic Control Analysis on Labor Force Participation Rate of Mothers with 4-year-olds Only

Figure 1.12. Synthetic Control Analysis on Weekly Working Hours of Mothers with 4-year-olds Only

To summarize, the main results show that the universal pre-k policy has positive effects on labor force participation rate, employment rate and weekly working hours of women of childbearing age, but the effect is statistically insignificant on full-time labor force participation of women in the fertility age. The empirical result agrees with the predictions of the theoretical model in Section 1.3. It suggests that the universal pre-k policy increases extensive margin labor outcomes of potential mothers. Though the policy effect on the intensive margins is ambiguous in theory, the empirical result shows universal pre-k positively affect mothers’ working hours. However, the positive effects of universal pre-k are neither consistent for mothers with younger than 5 years old children nor mothers with only 4 years old children. Although the labor force participation rate of mothers with younger than 5
years old children in Oklahoma is also positively affected by the universal pre-k policy, it is not statistically significant, and there is no evidence of significant and positive effect on mean working hours of mothers with younger than 5 years old children. If the study sample is further specified to mothers with 4 years old children only, the positive effect of universal pre-k vanishes. The null effect of universal pre-k on the labor outcomes of mothers with young children may be due to several possibilities.

First, the universal pre-k policy affects maternal labor supply through actual enrollment in child care programs. Albeit the total enrollment rate of Oklahoma is increasing in the post intervention period\textsuperscript{13}, mothers are not required to work to qualify for the universal pre-k program. It is possible that marginal utility from leisure and/or home child care for young children exceeds marginal utility of working, thus mothers remain to stay at home.

Second, since the universal pre-k policy may have a positive effect on the fertility decisions of women, it is also possible that the decision to have additional children would reduce a mother’s incentive to work. Thus the positive effect of the universal pre-k policy might be canceled out by the indirect effect of childbearing decisions.\textsuperscript{14}

Third, the universal pre-k policy may simply crowd out existing private day care and mothers transfer their children from private pre-k to public pre-k without changing their labor market decisions. The fourth explanation is that female labor force participation rates of highly developed countries, such as the US and Sweden (Lundin et al., 2008), are already high before a further expansion on child care policy. Moreover, studies on the wage elasticity of female labor supply show that women are no longer as responsive to wage changes as before (Blau and Kahn, 2007; Heim, 2007). Hence universal pre-k may have limited effect on the maternal labor supply.

At last, the theoretical labor supply model suggests mothers with lower family income are more likely to enter the labor market with low-cost or free child care programs, however, the existing child care programs for poor families, such as the Head Start, may already enable mothers from disadvantaged background to participate in the labor market.

\textsuperscript{13}See Appendix C.

\textsuperscript{14}In Appendix C, I show annual fertility rate in Oklahoma and estimate the effect of universal pre-k on fertility rate using the SCM.
1.6.2. Treatment Effect Heterogeneity

So far the analysis has primarily focused on all potential mothers. However, the impact of universal pre-k may vary in education level, marital status, family income and number of children mothers. Anderson and Levine (1999) find that the response of female labor force participation to the price of child care decreases in education levels for women with children under age 13. Koebel and Schirle (2016) show that the Canadian Universal Child Care Benefit has significantly negative effects on the labor supply of legally married mothers but has significantly positive effects on the labor supply of single mothers. In this section, I focus on the labor outcomes of women of childbearing age to avoid small sample size problem that maybe produced from further sample restrictions. And I will only show the effects of universal pre-k on labor force participation rate and working hours of potential mothers.

I first investigate the effects of universal pre-k on the labor force participation rate of women with family income below and above the poverty line. The theoretical model shows that mothers with lower non-labor incomes have stronger incentives to work for a given consumption level since their expected wage rates are more likely to be higher than their reservation wages. Moreover, as universal pre-k provides free child care to all pre-k age children, I also expect mothers with higher family income to respond to the policy. The results suggest that the labor outcomes of income disadvantage mothers are not so responsive to universal pre-k. The mean weekly working hours of potential mothers with family income below the poverty line in Oklahoma is higher than in the synthetic control group for a few years after the policy intervention and then drop to be close to the synthetic control group. Though the theoretical model suggests that the labor market decisions of mothers from poorer families (with a lower value of $Y$) are more likely to be affected by a child care policy, several free public pre-k and child care credit programs for poor families existed before universal pre-k. It’s possible that mothers from high-poverty families have already enrolled their children in other public pre-k programs and participated in the labor force before the availability of universal pre-k. The CPS data shows that labor force participation rate of women whose family income is below the poverty line is about 87% before 1998. Consequen-

\footnote{The synthetic control estimation and placebo tests on the labor outcomes of subsamples are presented in Appendix E.}
quently, the maternal labor supply of poorer mothers would not be significantly affected by this new child care policy. The empirical results show that the labor force participation rate of women from higher income family is also not responsive to universal pre-k. However, the mean weekly working hours of potential mothers with family income above the poverty line in Oklahoma increase after 1998 and are higher than the synthetic control group. Therefore, the results suggest that universal pre-k programs, though in high quality, may not affect the labor force participation rate of women with higher family income. This is probably because they care more about taking care of their children than working to support their families. However, for mothers who work before the implementation of universal pre-k, they are able to increase their working hours by sending their children to a free and all-day pre-k program.

The second subsample analysis is on mothers with differential marital status. The estimation results show that the treatment effect on the labor outcomes of married women is similar to the effect on all women of childbearing age. The effect of universal pre-k on the labor force participation rate of married women is positive and statistically significant. The mean working hours of the married sample present an increasing trend after 1998, and the post-trend of mean working hours of Oklahoma gradually exceed that of the synthetic control group. In the analysis of unmarried mothers, the working hours of unmarried mothers in Oklahoma are larger than the working hours of unmarried mothers in the synthetic control group during the post-intervention period, which indicates a positive effect of universal pre-k on weekly working hours of unmarried mothers. However, the labor force participation rate of unmarried mothers in Oklahoma decreases after 1998 and it is largely below the the labor force participation rate in the synthetic control group after 2001. One possible explanation is that unmarried mothers only receive financial support from the government if their income is below a threshold level. For example, children are not eligible for the free lunch if their family income is higher than the federal poverty line. The universal pre-k increases total income of poor family by reducing their child care payment, therefore, losses in labor income being compensated by welfare transfers from the government disincentivizes work. Further justification could be that the pre-intervention labor force participation rate of unmarried mothers in Oklahoma is already high. Since a great majority of unmarried mothers (82%) were already working, the total effect of the policy on maternal labor supply is limited.
Third, I analyze the labor outcomes of women of childbearing age with differential education levels. I categorize education levels of potential mothers into lower than high school, high school, and college and higher than college. The empirical results show that universal pre-k has a positive impact on women that are either low or high educated, especially those with a college or higher degree. This may be explained by the quality of Oklahoma universal pre-k because mothers with higher education level are more willing to send their children to high-quality pre-k programs. There is no evidence of statistically significant effect of universal pre-k on the working hours of women with either high or low education levels. This results may not be surprising because the effect of universal pre-k on working hours are intensive margin effect, which focuses on the sample of potential mothers who already work before the universal pre-k policy. Since education level is a proxy of working skills, it is highly possible that mothers’ working hours after the policy implementation are determined by their unchanged working abilities.

I also investigate the effect of universal pre-k on mothers with different numbers of children in the households because taking care of more children are more time-consuming. As expected, no significant effects exist on the labor force participation rate of mothers with more than two children. However, the empirical result shows that mothers with fewer than two children increase their labor force participation after universal pre-k becomes available. This suggests that the universal pre-k policy is more likely to increase the labor force participation rate of mothers with fewer children. The effect of universal pre-k on the working hours of mothers with either fewer or more than 2 children are statistically insignificant.

In summary, universal pre-k has differential effects on mothers from different socioeconomic groups, and universal pre-k, which provides free pre-kindergarten to all pre-k age children, also has an impact on mothers from non-disadvantaged backgrounds.

1.6.3. Expansion of Case Study to Georgia Universal Pre-k

In this section, I apply the synthetic control method to the case of 1995 Georgia universal pre-k. Georgia is the very first state that established universal pre-k statewide. Unlike Oklahoma, Georgia universal pre-k program is available for both 3- and 4-year-old children. As a result, the per-child expenditure on Georgia universal pre-k is much less, and Georgia
universal pre-k does not require a high standard of teacher quality in state legislation. The purpose of studying Georgia universal pre-k is to see the effect of universal pre-k on the labor outcomes of a relatively low-quality pre-k state.

Figure 1.13. Synthetic Control Analysis on Labor Force Participation Rate

Figure 1.14. Synthetic Control Analysis on Weekly Working Hours

Figure 1.13 and Figure 1.14 show the effects of Georgia universal pre-k on female labor force participation rate and working hours respectively, as well as their corresponding placebo tests. The pre-intervention period shows a good match between Georgia and its synthetic group. In the post-intervention period, both labor force participation rate and working hours of women aged between 25 to 45 are not significantly affected by Georgia universal pre-k policy. This might be explained by the possibility that maternal labor supply are more responsive to high quality universal pre-k in making labor market decisions.
1.7. Conclusion

This paper studies the impact of a high-quality universal pre-k program on the labor outcomes of mothers. I first present a theoretical labor supply model to predict the effect of universal pre-k on maternal labor supply. It suggests that price reduction and quality improvement in a child care program may increase the probability of a mother working but yields an ambiguous prediction regarding the working hours. In the empirical analysis, this paper chooses the high-quality Oklahoma universal pre-k program as the special case of interest. I apply a newly developed method—the synthetic control approach—to state level Current Population Survey data. To construct a better counterfactual for the treated state, Oklahoma, which is a small state with noisy data, I allow negative weights on the untreated state and add lags of the outcome variables in all pre-intervention years into the control variable set in the synthetic control analysis. I also use the placebo tests and pre/post RMSPE ratio test to investigate the significance of the treatment effect.

This paper examines the effects of universal pre-k on four labor market outcomes: labor force participation rates, employment rates, percentage of full-time labor force participation and working hours. The primary sample of interest is potential mothers, defined as women aged 25 to 45. The empirical findings suggest that the 1998 Oklahoma universal pre-k policy has a positive effect on the labor force participation rate, employment rate and weekly working hours of potential mothers in Oklahoma. The empirical results agree with the theoretical model predictions. And the empirics further show that the working hours of women of childbearing age are also increased by the universal pre-k policy. The empirical evidence also shows that there is little effect of universal pre-k on the labor outcomes of mothers with 4-year-olds only, though Oklahoma pre-k enrollment rate has been increased since the implementation of universal pre-k.

The analysis on heterogeneous treatment effects shows that universal pre-k has differential effects on mothers with different socioeconomic backgrounds. The universal pre-k policy increases the labor force participation rate for both low-educated (lower than high school education level) and high-educated (college or above college education level) mothers, but has no effect on high school graduates. Married women are more responsive to the universal pre-k policy than unmarried women in labor force participation, though the working hours of
both married and unmarried mothers are increased by universal pre-k policy. The working hours of potential mothers whose family income is above the poverty line is increased by universal pre-k, and mothers with fewer children (no more than 2) increase their labor force participation after the policy implementation.

At the end of the empirical study, I expand the synthetic control analysis to another pre-k state—Georgia, which is believed to have a relatively lower-quality universal pre-k program compared to Oklahoma. Georgia universal pre-k is found to insignificantly affect female labor outcomes. This may suggest that mothers are not responsive to lower-quality universal pre-k if other characteristics related to maternal labor market behavior are not significantly different between Oklahoma and Georgia.

In conclusion, the universal pre-k policy not only increases the pre-k enrollment rate and school performance of pre-k age children, but also positively affect the labor outcomes of fertility age women who live in a high-quality universal pre-k state. Hence, the universal pre-k policy may help reduce the inter-generational education and income gap by providing children good starts as well as providing mothers chances to work. The empirical results of this paper shows no discernible effect of universal pre-k on the labor market decisions of mothers with young children. Future research examining the effects of pre-k, and child policies in general, should not necessarily restrict the sample to households currently with young children. Moreover, future studies on efficiency or benefits to costs of universal pre-k programs should consider the welfare of all family members, and policy makers should pay more attention on the quality of pre-k programs.
<table>
<thead>
<tr>
<th></th>
<th>Before 1998</th>
<th>After 1998</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Oklahoma</td>
<td>rest of US</td>
</tr>
<tr>
<td>Labor force participation rate</td>
<td>0.72</td>
<td>0.74</td>
</tr>
<tr>
<td>Working hours</td>
<td>28.84</td>
<td>28.59</td>
</tr>
<tr>
<td>Family income</td>
<td>33164.04</td>
<td>35674.11</td>
</tr>
<tr>
<td>Mother’s age: 25-30</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td>Mother’s age: 30-35</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>Mother’s age: 35-40</td>
<td>0.59</td>
<td>0.60</td>
</tr>
<tr>
<td>Mother’s age: 40-45</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td>Mother’s education level: below high school</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>High school graduates</td>
<td>0.15</td>
<td>0.13</td>
</tr>
<tr>
<td>Some college, no degree</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>Associate Degree,</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Bachelors degree</td>
<td>0.05</td>
<td>0.07</td>
</tr>
<tr>
<td>Above college degree</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>White</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>Black</td>
<td>0.07</td>
<td>0.11</td>
</tr>
<tr>
<td>Other race</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>Family size</td>
<td>3.43</td>
<td>3.40</td>
</tr>
<tr>
<td>Food Stamp Recipients</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>Children enrolled in free lunch program</td>
<td>0.30</td>
<td>0.26</td>
</tr>
<tr>
<td>Missing in free lunch data</td>
<td>0.24</td>
<td>0.30</td>
</tr>
<tr>
<td>Below poverty line</td>
<td>0.16</td>
<td>0.14</td>
</tr>
<tr>
<td>GDP</td>
<td>58584.26</td>
<td>107485.4</td>
</tr>
<tr>
<td>Number of observations</td>
<td>19</td>
<td>950</td>
</tr>
</tbody>
</table>

*Sources:* State GDP data is from the Bureau of Economics Analysis (BEA), all other statistics are from the Current Population Survey (CPS).

*Notes:* The table presents the summary statistics of Oklahoma and the rest of all other U.S. states. In the state level, there is one observation each year for each state. There are 50 states and 1 district (District of Columbia) in the date set. Excluding Oklahoma, the number of the rest of the US states is 50.

Table 1.2. Summary Statistics of Selected Variables at State Level—OK vs. rest of US
<table>
<thead>
<tr>
<th>State</th>
<th>Weight</th>
<th>State</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>0.313</td>
<td>Montana</td>
<td>-0.111</td>
</tr>
<tr>
<td>Alaska</td>
<td>-0.108</td>
<td>Nebraska</td>
<td>-0.403</td>
</tr>
<tr>
<td>Arizona</td>
<td>-0.104</td>
<td>Nevada</td>
<td>0.431</td>
</tr>
<tr>
<td>Arkansas</td>
<td>-0.126</td>
<td>New Hampshire</td>
<td>0.228</td>
</tr>
<tr>
<td>California</td>
<td>0.005</td>
<td>New Jersey</td>
<td>0.028</td>
</tr>
<tr>
<td>Colorado</td>
<td>0.195</td>
<td>New Mexico</td>
<td>0.116</td>
</tr>
<tr>
<td>Connecticut</td>
<td>0.219</td>
<td>New York</td>
<td>-0.155</td>
</tr>
<tr>
<td>Delaware</td>
<td>-0.166</td>
<td>North Carolina</td>
<td>-0.036</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>0.048</td>
<td>North Dakota</td>
<td>-0.045</td>
</tr>
<tr>
<td>Florida</td>
<td>-0.167</td>
<td>Ohio</td>
<td>0.178</td>
</tr>
<tr>
<td>Hawaii</td>
<td>0.134</td>
<td>Oregon</td>
<td>-0.088</td>
</tr>
<tr>
<td>Idaho</td>
<td>-0.081</td>
<td>Pennsylvania</td>
<td>0.011</td>
</tr>
<tr>
<td>Illinois</td>
<td>-0.034</td>
<td>Rhode Island</td>
<td>-0.209</td>
</tr>
<tr>
<td>Indiana</td>
<td>-0.045</td>
<td>South Carolina</td>
<td>0.352</td>
</tr>
<tr>
<td>Iowa</td>
<td>0.158</td>
<td>South Dakota</td>
<td>-0.049</td>
</tr>
<tr>
<td>Kansas</td>
<td>0.239</td>
<td>Tennessee</td>
<td>-0.09</td>
</tr>
<tr>
<td>Kentucky</td>
<td>0.177</td>
<td>Texas</td>
<td>0.15</td>
</tr>
<tr>
<td>Louisiana</td>
<td>0.152</td>
<td>Utah</td>
<td>0.203</td>
</tr>
<tr>
<td>Maine</td>
<td>-0.151</td>
<td>Vermont</td>
<td>0.199</td>
</tr>
<tr>
<td>Maryland</td>
<td>-0.036</td>
<td>Virginia</td>
<td>-0.146</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>0.091</td>
<td>Washington</td>
<td>0.216</td>
</tr>
<tr>
<td>Michigan</td>
<td>-0.129</td>
<td>West Virginia</td>
<td>0.059</td>
</tr>
<tr>
<td>Minnesota</td>
<td>-0.025</td>
<td>Wisconsin</td>
<td>-0.23</td>
</tr>
<tr>
<td>Mississippi</td>
<td>-0.192</td>
<td>Wyoming</td>
<td>-0.026</td>
</tr>
<tr>
<td>Missouri</td>
<td>0.051</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Georgia is not included in this table of weights. Since Georgia universal pre-k starts in 1995, including Georgia in this analysis violates Assumption 2 of the synthetic control framework.

Table 1.3. Weights in the Synthetic Control Group
<table>
<thead>
<tr>
<th></th>
<th>OK vs. Synthetic OK</th>
<th>OK vs. Rest of US</th>
</tr>
</thead>
<tbody>
<tr>
<td>female education: less than high school</td>
<td>-0.001</td>
<td>0.063</td>
</tr>
<tr>
<td>female education: high school</td>
<td>0.036</td>
<td>0.265</td>
</tr>
<tr>
<td>female education: some college</td>
<td>-0.001</td>
<td>0.145</td>
</tr>
<tr>
<td>female education: associate degree</td>
<td>0.036</td>
<td>0.048</td>
</tr>
<tr>
<td>female education: college</td>
<td>-0.003</td>
<td>0.090</td>
</tr>
<tr>
<td>female education: master or doctoral</td>
<td>0.005</td>
<td>0.020</td>
</tr>
<tr>
<td>family total income</td>
<td>-5814.66</td>
<td>5829.97</td>
</tr>
<tr>
<td>married</td>
<td>0.038</td>
<td>0.090</td>
</tr>
<tr>
<td>separated, divorced, or widow</td>
<td>-0.009</td>
<td>0.024</td>
</tr>
<tr>
<td>never married</td>
<td>-0.050</td>
<td>-0.114</td>
</tr>
<tr>
<td>food stamp</td>
<td>0.015</td>
<td>0.003</td>
</tr>
<tr>
<td>free lunch</td>
<td>0.020</td>
<td>0.052</td>
</tr>
<tr>
<td>below poverty line</td>
<td>0.033</td>
<td>0.024</td>
</tr>
<tr>
<td>white</td>
<td>0.072</td>
<td>0.005</td>
</tr>
<tr>
<td>black</td>
<td>-0.049</td>
<td>-0.037</td>
</tr>
<tr>
<td>other race</td>
<td>0.018</td>
<td>0.031</td>
</tr>
</tbody>
</table>

Table 1.4. Mean Differences between Oklahoma and “Synthetic Oklahoma” and Oklahoma and the Rest of US States in Demographic Economic Characteristics (Partially)
2.1. Introduction

Trade liberalization, especially import competition from low-waged countries, is known to impact good prices, household income and labor market outcomes of high-wage countries. Import competition in turn may affect the welfare of children. Although previous studies have investigated the effects of trade liberalization on child outcomes such as child labor in developing countries (Edwards and Pavcnik, 2005) and mortality of teenagers (Levine and Rothman, 2006; Owen and Wu, 2007), little attention is paid on the impact of import competition on the birth outcomes of children in developed countries. This lack of research exists despite the well-known fact that early childhood human capital development plays a crucial role in later-life outcomes (Almond et al., 2005; Behrman and Rosenzweig, 2004; Currie et al., 2010). Prenatal disadvantaged children are at risk of worse outcomes on physical wellness in their later-lives as well as adverse socioeconomic outcomes such as poor school performance in childhood, lower future wages and higher crime rates in adulthood. As a result, understanding the effects of trade liberalization on newborn health in the US may help policymakers effectively implement public health programs for disadvantaged children, who are differentially affected by import competition. To fill this gap in the literature, I examine the effect of exposure to Chinese import competition on the birth outcomes of US children in this paper.

Recently, existing literature is increasingly interested in the impact of import competition on health outcomes. Pierce and Scott (2016) find that counties more exposed to the change in trade policy exhibit higher rates of suicide and related causes of death that have been linked to relative loss of employment and income. McManus and Schaur (2016) find that import competition from China increases the injury rates and the injury risk in the competing US
industries. The impact of Chinese import competition is not only on workers themselves in the trade-exposed industries, Marcus (2013) and Bubonya et al. (2014) find that job losses of workers have spillover effect on the mental well-being of their spouses. Marcus (2013) finds that unemployment due to firm closure decreased mental health by 27% of a standard deviation for unemployed individuals themselves and by 18% of a standard deviation for their spouses. Bubonya et al. (2014) present evidence that the mental well-being of wives declines following their husbands’ job loss if that job loss results in a sustained period of unemployment or pre-unemployment financial hardship.

Exposure to import competition from China may impact birth outcomes of US children, either positively or negatively, by altering labor market opportunities, affecting the provision of public goods, reducing the prices of goods, mitigating pollution caused by manufacturing plants and affecting the fertility decisions of young couples. The first two channels are expected to negatively affect family welfare and newborn health, while the last three effects of trade are likely to be potential mechanisms that improve the health outcome of newborns.

First, exposure to Chinese import competition may increase unemployment in the US, in turn affects child outcomes through parental job losses. The reduction of production costs and trade barriers in less developed countries allow foreign firms with comparative advantage in labor out-compete US domestic manufacturing firms, resulting in an increase in local unemployment. A substantial literature provides theoretical models and empirical evidence for the negative relationship between Chinese import competition and employment of US workers (Egger and Kreickemeier, 2009; Autor et al., 2013; Pierce and Schott, 2012; Lake and Millimet, 2015; Acemoglu et al., 2016). The trade-induced job losses in turn may have an impact on the birth outcomes of children through maternal nutrition deprivation because parental job losses reduce family income in the short run, thus parents who lose jobs may not have enough money to buy food and nutritious supplements. Another impact of parental job losses on the health outcomes of newborns is through the mental stress of mother, even if it is their spouses rather than themselves who lose jobs (Marcus, 2013; Bubonya et al., 2014). The mothers’ mental stress may trickle down to their children in utero. Therefore, trade-induced job losses may have a negative impact on newborn health.
Second, import competition from China has an impact on the provision of public goods through its effect on government revenue. Feler and Senses (2016) find that the increasing import competition from China leads to declines in housing prices and business activities, resulting in less government revenue. As shown by Feler and Senses (2016), funding for public services such as education and health care for low-income families is highly localized in the US, with heavy reliance on property and sales tax revenues. The reduction in the provision of public goods associated with prenatal health care thus may negatively affect the birth outcome of US children.

Third, trade liberalization is known to provide households access to cheaper goods and a wide variety of consumption (Levine and Rothman, 2006; Frenstra and Weinstein, 2010; Tovar, 2012; Courtemanche et al., 2015). The idea that trade liberalization improves living standard of people is supported by the theory of comparative advantage (Frankel and Romer, 1999), rent seeking theory in the political economics (Dollar and Kraay, 2004), and substantial empirical evidence. For example, Amiti et al. (2017) find that the China trade shock reduced US manufacturing price index by 7.6 percent between 2000 to 2006, which is in principle driven by policy changes after China’s accession to WTO. Though trade liberalization may not directly affect the health care industry in the US, the availability of cheap imported materials may reduce production cost related to medical equipment and drugs. Moreover, the reduction in good prices has a direct income effect on US households, thus the accessibility to cheaper goods enables households to spend more on food and health products. As a consequence, trade liberalization may increase the health outcomes of US newborns through the reduction in good prices.

Fourth, the rise of Chinese imports may reduce pollution in trade-exposed US counties as Chinese import competition prompts closure of inefficient manufacturing plants. Domestic companies are also able to import intermediate or final products of heavily polluted industries from China instead of producing, or outsource production to low-wage countries. A number of literature is interested in the impact of air pollution on child health. Currie and Neidell (2005) and Chay and Greenstone (2005) both find improvement in air quality reduces infant mortality. The World Health Organization (WHO) conducts an assessment of research on pollution and child health in 2005. The report, entitled “Effects of air pollution on children’s
health and development”, shows that the risks posed by ambient air pollutants are related to various aspects of children’s health.

Finally, the import penetration from China may also affect the prevalence of marriage among young adults and their fertility decisions. Autor et al. (2017) explores the impact of trade shock from China on the marriage and fertility decisions. They find that import competition from China reduces marriage rate and fertility rate of young US people, and the reduction in fertility is not uniform across demographic groups. Since they find the fertility among young mothers decreases proportionally less than the fertility among old mothers, the reduction of fertility rate may positively affect birth outcomes of US children because younger women are more likely to deliver healthier babies.

To find the causal effect of the surge of Chinese imports on the health outcomes of US newborns, I use the instrumental variable method following Autor et al. (2013). The empirical results suggest that US exposure to Chinese import competition reduces the percentage of newborns with low birth weight and premature births, while other birth outcomes are not significantly affected by trade exposure. Further, in order to provide a more comprehensive understanding of health consequences of import competition, I examine two potential mechanisms that would improve newborn health—improvement in air quality and reduction in good prices. I find reduction in air pollution and availability of cheaper goods are possible channels that would improve the birth outcomes of US children. Another potential mechanism might be the increase in the share of young mothers in fertility (Autor et al., 2017).

This paper contributes to the strand of literature on the effect of Chinese trade penetration on the welfare of US people. Though a substantial literature has shown evidence that Chinese import competition causes job losses of US workers, this paper finds evidence that trade between US and China is beneficial to US children.

The rest of the paper is organized as follows. Section 2 describes the empirical strategy used to examine the effect of Chinese import competition on newborn health. Section 3 describes the data. Section 4 presents and discusses the main results. Section 5 discusses the potential mechanisms through which exposure to Chinese import competition may positively affect the birth outcomes of US children. Section 6 concludes.
2.2. Empirical Strategy

Since exposure to Chinese import competition may be endogenous to newborn health, I explore the exogenous variation of local trade exposure via county-level industry structure to examine the causal effect of Chinese trade penetration. The empirical strategy is based on Autor et al. (2013), who study the effect of Chinese import competition on US labor market using an instrumental variable approach.

The empirical model is a standard two-way fixed effects model at the US county level estimated by two-stage least squares. The regression model is

\[
Y_{ct} = \alpha + \beta_1 IPW^u_{ct} + \beta_2 IPW^{u2}_{ct} + X'_{ct}\beta + \lambda_c + \gamma_t + \varepsilon_{ct}
\] (2.1)

where \(Y_{ct}\) is the birth outcomes in county \(c\) and year \(t\), and \(X_{ct}\) is a vector of control variables including the demographic characteristics of the newborns and their mothers. \(\lambda_c\) and \(\gamma_t\) are county and year fixed effects, respectively. \(\varepsilon_{ct}\) is the error term. Since birth outcomes of children are more likely to be a result of the exposure to Chinese imports when children are in utero, the trade exposure variable \(IPW^u_{ct}\) takes the value of last time period. \(IPW^{u2}_{ct}\) is included to account for the fact that the fraction of adverse birth outcomes decline in a non-linear manner. Figure 1 shows that the predicted value of outcome variables are more likely to follow a non-linear pattern when they are regressed by the county level import per worker.

The exposure to Chinese import competition is measured by the concept of the per-capita change of Chinese import to the US (\(IPW_{ct}\)).\(^1\) The trade exposure variable \(IPW^u_{ct}\) is defined as the imports apportioned to US counties weighted by its share of national industry employment in the baseline year. Formally,

\[
IPW^u_{ct} = \frac{1}{L_{ct}^{1990}} \sum_j \frac{L_{cj}^{1990}}{L_{jt}^{1990}} M_{jt}
\] (2.2)

\(^1\)In the existing literature, there are several measures of exposure to trade competition. Jensen and Kletzer (2006) and Eliasson et al. (2012) use locational Gini coefficient to measure the industrial concentration of domestic regions. Bernard et al. (2003) measure an industrial exposure to imports from low-wage countries via the value share (VSH) of imports originating in these low-wage countries.
where $IPW_{ct}^u$ is the per-capita Chinese import to the US in county $c$ and year $t$. $L_{c}^{1990}$ is the working population by county in the baseline year 1990, $L_{j}^{1990}$ is the ratio of county level industry employment to national industry employment in 1990 and $M_{jt}$ is the import of US from China in industry $j$ and in year $t$. The key explanatory variable $IPW_{ct}^u$ is instrumented by the exposure to Chinese import competition on eight other high-income countries $IPW_{ct}^o$.

Import per worker ($IPW_{ct}$) is widely used to study the effect of the exposure to import competition from low-wage countries on labor market outcomes. It is chosen as the measure of exposure to Chinese import competition in this study for two reasons. First, import per worker ($IPW_{ct}$) is related to both trade shock and local labor market. The impact of import competition on domestic employment is a potential mechanism that affects birth outcomes of children. The second reason is that import per worker ($IPW_{ct}$) measures county level exposure to Chinese import competition by weighting industry level imports with the county-industry labor share, though the imports of each US county is not directly observed in trade data.

The identification strategy relies on the assumption that the surge of Chinese import to US is driven mainly by the productivity growth of China after its transition to a market-oriented economy and the trade costs reduction after its openness to international trade, especially after China’s accession to the WTO. Under this assumption, the rising of Chinese imports has common within-industry effects on trade sectors of the United States and other high-income countries. The exposure to Chinese import competition of eight other high-income countries $IPW_{ct}^o$ is thus used to instrument the US trade exposure variable $IPW_{ct}^u$.\footnote{Following Autor et al. (2013), the eight other high-income countries are those that have comparable trade data covering the full sample period: Australia, Denmark, Finland, Germany, Japan, New Zealand, Spain, and Switzerland.} It captures the growth in imports from China that reflect technology shocks and demand shocks common to high-income countries.

2.2.1. US Exposure to Chinese Import Competition

The key explanatory variable is county level exposure to Chinese import competition. I follow Autor et al. (2013) to use regional variation in industrial composition to construct local exposure to Chinese import competition.
In the first step, I use the 1990 Census 5% sample to construct the baseline year county level labor share by industry. The Census data has rich information on labor statistics, including occupation and associated industry. The number of industries used to construct county-industry employment share is 93. The county-industry employment share is the number of workers in one industry in each county over the total number of worker working in this industry nationwide. It is fixed in the baseline year 1990 because the contemporaneous employment share might be endogenously affected by Chinese import competition through job reallocation.

In the next step, I construct the trade exposure variable $IPW_{ct}$, I link the county-industry employment share obtained from 1990 Census to US-China trade data. The import data is collected from the World Integrated Trade Solution (WITS). WITS provides information on trade volume and the impact of protection and tariff changes. Since the industry identifier in WITS is the Standard Industry Code (SIC) while the industry identifier in the 1990 Census data is coded as “ind1990”, I map the SIC to ind1990 to obtain a common industry identifier to combine trade data with employment shares. Import per worker is obtained by summing up the products of county-industry employment and imports from China of all industries following equation (2) in Section 2. The final panel has a total number of observations of 6,034 from 1991 to 2004.  

To implement the identification strategies, I utilize a panel of trade and health data from 1992 to 2004. I first combine the US-China trade data with the county-industry employment share that is calculated from the 1990 Census to obtain county level exposure to Chinese import competition. The data is then merged with the birth outcome data. The final panel has approximately 340 counties each year because the county identifiers of the National Vital Statistics are only available for around 460 large populated counties in the US, and the number of counties that can be identified in the 1990 Census data is about 430. The county identifiers of small counties are not publicly available in these data sets due to security reason.

---

2.3. Data

2.3.1. US Exposure to Chinese Import Competition

The key explanatory variable is the county level exposure to Chinese import competition. I follow Autor et al. (2013) to use a regional variation in industrial composition to construct local exposure to Chinese import competition.

In the first step, I use the 1990 Census 5% sample to construct the baseline year county-industry labor share. The Census data has rich information on labor statistics, including occupation and associated industry. The number of industries in the trade sectors that are used to construct county-industry employment share is 93. The county-industry employment share is the number of workers in one industry in each county over the total number of worker working in this industry nationwide. It is fixed in a baseline year because the contemporaneous employment share might be endogenously affected by Chinese import competition through job reallocation. I chose 1990 as the baseline year because it is before the study time period 1992-2004 and the Census is available in 1990 to construct the county-industry employment share.

In the next step, I construct the trade exposure variable $IPW_{ct}$ by linking the county-industry employment share obtained from 1990 Census to the industry level imports from China. The import data is collected from the World Integrated Trade Solution (WITS). The WITS provides information on trade volume and the impact of protection and tariff changes. Since the industry identifier in WITS is the Standard Industry Code (SIC) while the industry identifier in the 1990 Census data is coded as “ind1990”, I map the SIC to ind1990 to obtain a common industry identifier for the data combination. Import per worker is obtained by summing up the the products of county-industry employment and imports from China of all industries following equation (2) in Section 2. The final panel has a total number of observations of 6,034 from 1991 to 2004.4

---

2.3.2. Birth Outcomes and Family Characteristics

The birth outcomes are from the National Vital Statistics System (NVSS). It is an individual level data that provides rich information on vital events such as births, deaths, marriages, divorces, and fetal deaths. The Vital Statistics also provides data on the demographic and socioeconomic characteristics of the newborns and their mothers. In this study, I collapse the birth outcomes into county level by year.

The birth outcomes include the percentage of children with low birth weight, low Apgar scores, overweight and premature birth. Low birth weight (LBW) is defined as a birth weight of a live-born infant of 2,499 grams or less, regardless of gestational age. Low birth weight infants have a greater risk of poor later-life health, which may require a longer period of hospitalization after birth, and they are also more likely to develop significant disabilities. Overweight birth is another unusual birth weight state, it is defined as a birth weight of more than 4,000 grams.\(^5\)

The Apgar score provides an accepted and convenient method for reporting the status of the newborn infant immediately after birth. A 5-minute Apgar score that is no more than 7 is considered to be a low score, and the Apgar evaluation is given every 5 minutes for the infants whose total score are less than 7. Another useful criterion of newborn health is the gestation period. Preterm birth, also known as premature birth, is defined by the National Institutes of Health as the birth of a baby at fewer than 37-week gestational age. Premature infants are at greater risk for cerebral palsy, delays in development, hearing problems, and sight problems. Preterm birth is also the most common cause of death among infants worldwide and the earlier a baby is born, the higher risk for adverse health outcomes.

The control variable set contains the demographic and family characteristics of the newborns, including a infant’s gender and birth order as well as her mother’ race, residential information, education level, age at birth and marital status. Controls are aggregated at the county level as means or percentages.

I further examine the effect of exposure to Chinese import competition on the death rate of newborns since the estimation will be biased if healthier babies are selected into live births. The infant mortality rate is the number of deaths under one year old occurring

\(^{5}\)Overweight birth is defined by using the definition of “fetal macrosomia”. 
among the live births in a given geographical area during a given year over per 1,000 live births occurring among the population of the given geographical area during the same year.

The infant mortality data is collected from the Compressed Mortality File (CMF), which is comprised of a county-level national mortality file and a corresponding national population file. The CMF is collected from the mortality files of the National Center for Health Statistics (NCHS), which record every death of a U.S. resident annually from 1968 to 2016. The detailed mortality files contain an extensive set of variables collected from hospital death certificates. In this paper, I use the county level infant mortality data. However, like the National Vital Statistics, the CMF and NCHS only collect mortality data from high population-density counties, so the county level data does not include all counties in the US. The mortality data is merged with the the birth outcome data described above by county. Table 1 shows summary statistics.

The final step of data management is to combine the health data (birth outcomes and mortality rate) and the trade data with the exposure variable. The longitudinal data thus includes the trade exposure variable, health outcomes and family characteristics of newborns in each US county over time. Figure 1 shows the fractions of mean low birth weight across all US counties in every study year. It suggests there is an increasing fraction of low birth weight newborns in the US after 1991, though the medical and health care industry was rapidly growing. The right panel of Figure 1 presents the relationship between the import per worker variable and the fraction of low birth weight by county, the OLS predicted line suggests a negative relationship between import per worker and low birth weight. Figure 2 shows the time trend of average mortality rate of newborns, which was decreasing in 1990s. The scatter plot in the right panel shows a negative relationship between import per worker and mortality rate.

2.4. Empirical Results

2.4.1. Main Results

In this section, I empirically examine the effect of exposure to Chinese import competition on health outcomes of US newborns, including the percentage of low birth weight, overweight,
low Apgar scores, early delivery and the mortality rate of US newborns. Table 2 presents the results applying the empirical model in Section 2. It shows that the trade exposure variable $IPW_{ct}$ and its quadratic form $IPW_{ct}^2$ are jointly significant in impacting low birth weight and premature birth, while only the effect on low birth weight is statistically significant at a significance level of 1%. Also note that though $IPW_{ct}$ and $IPW_{ct}^2$ do not significantly affect the outcome of low Apgar score individually, they have jointly significant impact on the fraction of low Apgar score.

Since the regression model includes the quadratic form of county level import per worker, the impact of exposure to Chinese import competition should be explained by the marginal effect of import per worker. Moreover, the sign of the marginal effect of import per worker
depends on the level of exposure to Chinese import competition. More specifically, exposure to Chinese import competition measured by import per worker \((IPW_{ct})\) reduces (increases) the incidence of low birth weight and premature births if import per worker is below (above) a certain threshold. For example, the marginal effect of \(IPW_{ct}\) on the percentage of low birth weight is negative when \(IPW_{ct} < 3.5086\), which is true for nearly 99.7% of US counties in the data. The marginal effect of \(IPW_{ct}\) on the percentage of premature births is negative if \(IPW_{ct} < 4.0969\), and nearly 99.8% of US counties are below the threshold.

However, the marginal effect of \(IPW_{ct}\) on the incidence of low Apgar scores seems to be positive for all US counties from estimation results in Table 2. This result is surprising because I find positive correlation between the percentage of low birth weigh and low Apgar score, thus \(IPW_{ct}\) and \(IPW_{ct}^2\) should have consistent impact on the fraction of newborns with low birth weight and low Apgar score. I plot the marginal effects of import per worker on the percentage of newborns with low birth weight, short gestational period, low Apgar score in Figure 2. It shows that the marginal effects of average import per worker on low birth weight and premature birth is always below zero at 5% confidence interval. However, the confidence interval of the marginal effect on the fraction of low Apgar score includes negative values at the 5% significance level. Since the confidence interval is too wide to exclude null effect, passing the joint significance test does not necessarily ensure a significant positive effect of import per worker on the percentage of newborns with low Apgar scores at the 5% confidence interval.

Table 2 also shows the test results of the first-stage F-test and the endogeneity test for the instrumental variable approach. The first-stage F-test is for testing the validation of the instrumental variables, since all IV regressions pass the F-test, import per worker of other eight high-income countries is a valid instrument for the import per worker of US. In addition, four out of five birth outcomes pass the endogeneity tests. It suggests that an instrument is needed for import per worker of US since it is endogenous to the birth outcomes in the model.

Import competition from China may have differential impacts on birth outcomes of children with different initial parental human capital endowment through its direct effect on local labor market. I analyze the effect of Chinese import competition on newborns whose
mothers are differently educated. The mothers’ education levels are classified into three categories: less than high school, high school diploma or college unfinished, and bachelor’s degree and above. The incidence of low birth weight is reduced for all sub-samples of newborns, and the percentage of children whose mothers are in lower education level decreases in a larger magnitude than those whose mothers are higher educated. It suggests that Chinese import competition might have a greater benefit on disadvantaged children.

In conclusion, the estimation results suggest exposure to Chinese import competition improve newborn health for almost all counties in the data. It is also found that the trade exposure has little effect on infant mortality. The empirical results thus suggest no evidence that higher exposure to Chinese import competition harms the health outcomes of newborns in the US; instead, the rise of Chinese imports improves the health condition of newborns on their birth weight and gestation. Moreover, the empirical evidence shows a larger impact

---

Figure 2.3. Marginal Effects of IPW with Confidence Interval

6 The estimation results are shown in Appendix A.
of Chinese import competition on the birth outcomes of disadvantaged children.

The positive effect of import competition on newborn birth weight and gestation may be explained by lower prices of goods important to prenatal health due to free trade, the reduction in air pollution due to the closure of manufacturing plants and the decline in the share of health disadvantaged mothers in fertility. It is also possible that the positive labor market effect in non-import competing and trade-beneficiary sectors exceeds the negative effect on import-competing trade sectors. Therefore, the effect of Chinese import competition is positive on average.

2.4.2. Potential Mechanisms

In this section, I discuss three potential mechanisms through which exposure to Chinese import competition may have a positive impact on newborn health: reduction in air pollution, income effect from the availability of cheaper goods and an increase in the share of young women in fertility.

2.4.2.1. Trade Exposure and Air Pollution

This section examines whether the positive effect of exposure to Chinese import competition on newborn health is through the reduction in air pollution since China’s comparative advantage in labor-abundant sectors may result in closure of heavily polluted manufacturing plants in the US. Chintrakarn and Millimet (2006) and McAusland and Millimet (2016) find evidence of a beneficial impact of expanded trade on the environmental quality. Currie and Neidell (2005) and Chay and Greenstone (2005) find that the improvement in air quality reduces infant mortality.

I use the same empirical model in Section 2 to regress air pollution outcomes on the trade exposure variable $IPW_{ct}$. Thus the regression results explain the effect of import competition from China on the air quality of differentially exposed US counties. The air quality data is collected from the United States Environmental Protection Agency (EPA), it provides air quality data collected at outdoor monitors across the United States. The Air Quality Index Report (AQI) of EPA displays annual summaries of AQI values in a county or a city. I transfer count of days in each AQI category into percentage forms because the
The total number of days under monitoring per year is not uniform across counties. I also study the effect of import per worker on the percentage of days when the AQI could be attributed to each criteria pollutant.

Table 3 shows the estimation results. The marginal effect of $IPW_{ct}$ on good days is positive when $IPW_{ct} < 2.7668$, which is true for nearly 98.8% of US counties in the data. The marginal effect of $IPW_{ct}$ on unhealthy days is negative when $IPW_{ct} < 3.4681$, and nearly 99.4% of US counties have an exposure level below 3.4681. The results suggest that $IPW_{ct}$ improves air quality by increasing the percentage of good days in air quality and decreasing the fraction of unhealthy days in AQI categories.

Table 3 also shows the trade effect on air pollution attributed to each criteria pollutant. Three out of five toxic gaseous pollutants (CO, $NO_2$, and Ozone) are not significantly affected by exposure to Chinese import competition. The import competition from China reduces the particle pollution measured by PM2.5 in the US. The marginal effect of $IPW_{ct}$ on PM2.5 is negative when $IPW_{ct} < 2.9928$, which contains nearly 99% of US counties. However, emission of $SO_2$ is increased by exposure to Chinese import competition and the marginal effect of $IPW_{ct}$ on $SO_2$ is positive for all US counties in the data. The negative effect of trade exposure on the criteria pollutant $SO_2$ may be canceled out by the reduction in other criteria pollutants. Thus we observe an overall positive effect of Chinese import competition on the percentage of days in good air quality.

Since the import data is available in the industry level, the import competitive industries can be further categorized into ‘dirty’ (heavily polluted) or ‘clean’ industries. I calculate the import per worker of the two types of industries, and then examine the effects of import per worker by different industrial air pollution levels on newborn health. I find the marginal effect of import per worker from heavily polluted industries on the fraction of low birth weight is negative for most of the US counties, however, Chinese import competition from clean industries, including the food sectors increases the fraction of low birth weight. The results suggest import competition from heavily polluted industries are more likely to improve

---

7 The industries are defined to be dirty if they belong to mining and quarrying, Manufacturing, Electricity, gas, steam and air conditioning supply, or Water supply; sewerage, waste management and remediation activities.

8 The estimation results are shown in Appendix B.
the health outcomes of US newborns.

In summary, the empirical results suggest that air quality improvement is likely to be a mechanism that leads to a positive health impact of exposure to Chinese import competition on US children. And the import competition from heavily polluted industries are more likely to improve the health outcomes of US newborns through the improve in air quality, though it is believed to increase unemployment in US labor market.

2.4.2.2. Trade Exposure and Access to Cheaper Goods

International trade is well-known to reduce prices for consumers and provide consumers a wide variety of goods and services. Thus, the rise of Chinese imports may affect the average birth outcomes of differentially trade-exposed counties through the reduction in good prices and the improvement in the living conditions of local families. Since data on household consumption or food expenditure is hard to obtain in the study time period, I use the number of Walmart stores and supercenters in each county to proxy for accessibility to cheap goods. Courtemanche et al. (2015) suggest that being close to Walmart stores can help reduce food insecurity. The number of Walmart stores in each county, however, does not consider households who live near county borders and are closer to a Walmart store in a nearby county. Therefore, the estimation using the growth of Walmart stores as a measure of accessibility to cheap goods may be downward biased.

The locations of Walmart stores and supercenters across the US is obtained from the Remote Sensing Data, which is collected from the aerial photography and NASA satellite imagery. The location data is further collapsed into county level and merges with the trade data in Section 3.1. The empirical strategy follows Section 2. The key explanatory variable is the import per worker and the dependent variable is the growth in number of Walmart stores and supercenters in each county in the study time period.

Table 4 shows that exposure to Chinese import competition increases the number of Walmart stores in more trade-exposed counties. The marginal effect of $IPW_{ct}$ on the number of Walmart stores in each county is positive when $IPW_{ct} < 3.8388$, almost all counties

---

9The data of location of Walmart and Target stores is ready in the project of Using GIS and Remote Sensing to Teach Geoscience in the 21st Century.
(99.5%) is below this trade exposure threshold. The result suggests that exposure to Chinese import competition might improve newborn health through its effect on the prices of goods.

The increasing availability of cheap goods to local US families is more likely to have a higher impact on poorer families or poorer counties. Previous literature has shown that the increasing number of Walmart scores may reduce the poverty rate of US counties (Courtemanche et al., 2015). I then study the effects of import per worker interacted with county level poverty rates on the health outcomes of newborns. Like the county-industry employment share defined in Section 3.1, the contemporaneous poverty rate might be endogenously affected by Chinese import competition, thus I use county level poverty rates in 1990 as the baseline poverty rates. The regression model is then given by

\[
Y_{ct} = \alpha + \beta_1 IPW_{ct} + \beta_2 IPW_{ct}^2 + \beta_3 IPW_{ct}^u \times poverty_{c} + \beta_4 IPW_{ct}^u \times poverty_{c}\\ + X'_{ct}\beta + \lambda_c + \gamma_t + \varepsilon_{ct} 
\]  

(2.3)

where \(poverty_c\) is poverty rate of each county in 1990, and \(X_{ct}\) contains all the control variables in Equation (2) and Table 2.

Table 5 shows the estimation results. The marginal effect of import per worker on health outcomes of newborns in Equation (3) now depends not only on county level exposure to Chinese import competition, it depends on county level poverty rate as well. For example, the marginal effect of import per worker on the percentage of newborns with low birth weight is now given by

\[
ME_{IPW} = 0.0028 - 2 \times 0.0001 IPW - 0.0651 poverty + 2 \times 0.0055 IPW \times poverty 
\]  

(2.4)

Equation (4) shows that both the sign and the magnitude of the marginal effect of IPW depend on IPW and the poverty rate of counties, and the marginal effect is likely to be positive if there is one county that is more exposure to Chinese import competition and its poverty rate is high. It is easier to show the marginal effect of import per worker interacted with poverty rate of each county in a graph. Figure 3 shows the marginal effects of average import per worker over years on the fraction of low birth weight at different percentiles of
poverty rate. It suggests that the marginal effect of aggregated IPW on the percentage of low birth weight is generally negative, and as the poverty rate increases from a lower percentile to a higher percentile, the negative effect becomes larger.

Figure 2.4. Marginal Effects of IPW with Confidence Interval at Different Percentile of Poverty

I can also examine the effect of exposure to Chinese import competition interacted with time-variant county wealth measured by per capita income (in 10,000 dollars) of each county over time. Similarly, the regression model is given by

\[
Y_{ct} = \alpha + \beta_1 IPW_{ct} + \beta_2 IPW_{ct}^2 + \beta_3 IPW_{ct} \times pcinc_{ct} + \beta_4 IPW_{ct}^2 \times pcinc_{ct} + X'_{ct}\beta + \lambda_c + \gamma_t + \varepsilon_{ct}
\]  

(2.5)

where \( pcinc_{ct} \) is per capita income of counties each year, and \( X_{ct} \) contains all the control variables in Equation (2) and Table 2.

Table 6 shows the estimation results. The marginal effect of import per worker on the percentage of low birth weights \( ME_{IPW} \) is now given by

\[
ME_{IPW} = 0.0190 - 2 \times 0.0079 IPW - 0.0047 pcinc + 2 \times 0.0018 IPW \times pcinc
\]  

(2.6)
Moreover, the marginal effect of per capita income also depends on the level of trade exposure, so that

\[ ME_{pcinc} = 0.0014 - 0.0047IPW_{ct} + 2 \times 0.0018IPW_{ct} \times pcinc_{ct} \] (2.7)

The marginal effect of import per worker and per capita income is similar to the marginal effect of import per worker and poverty rate. Equation (6) shows that the sign of the marginal effect of import per worker depends on both the trade exposure and the per capita income of a county. Suppose the \(1.2 \leq IPW \leq 1.36\), the first part of the marginal effect \(0.019 - 2 \times 0.0079IPW \leq 0\), and as \(4.71 - 2 \times 1.76IPW > 0\), the second part of Equation (6) is also negative. In this case, the marginal effect of IPW on the percentage of low birth weight is negative and decreases as the per capita income increases. However, only 3% of counties has an import per worker between 1.2 and 1.36. For the rest of 97% counties whose import per worker is below 1.2 or above 1.36 import per worker, the sign of the marginal effect of import per worker is determined simultaneously by the interaction between county level trade exposure and per capita income.

2.4.2.3. Trade Exposure and Fertility

Autor et al. (2017) study the effect of trade shock from China on the marriage rate and fertility rate of young US people. Following Autor et al. (2013), the marriage and fertility decisions of young US people are affected through the adverse labor market effect of import competition. They show evidence that one-unit import shock lowers births per thousand women of ages 20-39 by 4%. But this decline is not uniform across demographic groups. Fertility among younger (including teens) and unmarried women falls proportionately less than fertility among older and married women.

Though the increase in out-of-wedlock birth share may have a long-run adverse effect on birth outcomes, the decline in the share of fertility among older women may result in an increase in average birth weight. This is because pregnancy is often physically easier for women in their 20s and there’s a lower risk of health complications since young women are also less likely to have gynecological problems. As a result, younger women are less likely to
have premature or low-birth-weight babies than older women, and thus China penetration in trade may potentially reduce the incidents of poor outcomes at birth by increasing the share of young women in fertility.

### 2.5. Conclusion

This paper uses an instrumental variable fixed effect model to examine the causal effect of an increasing exposure to Chinese import competition on newborn health in the US. Despite the strong evidence that US exposure to Chinese imports has had negative impacts on labor market outcomes and adult health, I find no evidence of an adverse effect on newborn health. In fact, there is evidence of a reduction in the incidence of low birth weight and premature birth. This paper thus makes a contribution to trade literature by suggesting the benefits of free trade on the newborns.

I further investigate three potential mechanisms through which trade shock from China may positively affect birth outcomes of US children. I find that exposure to Chinese import competition improves the average air quality of more trade-exposed US counties and the benefit of trade on air quality may lead to an improvement in newborn health. The empirical evidence also shows that reduction in good prices induced by the rise of Chinese imports is another mechanism that may drive down the incidence of low birth weight. Moreover, Chinese import competition has a larger impact on the average birth outcomes of US counties that have a higher poverty rate or a lower per-capita income level. At last, previous studies show that exposure to Chinese import competition reduces the fertility rate of US young couples while increases the share of young women in fertility. In turn, early age childbearing improves the average birth outcomes of US children.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Birth Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Birth Weight</td>
<td>0.075</td>
<td>0</td>
<td>0.202</td>
<td>4426</td>
</tr>
<tr>
<td><em>(Birth Weight &lt; 2500 grams)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overweight Birth</td>
<td>0.024</td>
<td>0</td>
<td>0.222</td>
<td>4426</td>
</tr>
<tr>
<td><em>(Birth Weight &gt; 4000 grams)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Apgar Score</td>
<td>0.033</td>
<td>0</td>
<td>0.6</td>
<td>4304</td>
</tr>
<tr>
<td><em>(5 minutes Apgar Score ≤ 7)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Premature Birth</td>
<td>0.192</td>
<td>0</td>
<td>0.397</td>
<td>4426</td>
</tr>
<tr>
<td><em>(Gestation &lt; 37)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mortality Rate</td>
<td>0.008</td>
<td>0.002</td>
<td>0.022</td>
<td>4426</td>
</tr>
<tr>
<td><strong>Trade Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Import per worker (IPW)</td>
<td>0.549</td>
<td>0.018</td>
<td>5.866</td>
<td>4426</td>
</tr>
<tr>
<td><strong>Mother Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother’s Age</td>
<td>27</td>
<td>24</td>
<td>34</td>
<td>4426</td>
</tr>
<tr>
<td>Married</td>
<td>0.677</td>
<td>0.261</td>
<td>1</td>
<td>4426</td>
</tr>
<tr>
<td><strong>Mother’s Education Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ High School</td>
<td>0.191</td>
<td>0</td>
<td>0.583</td>
<td>4426</td>
</tr>
<tr>
<td>High School</td>
<td>0.540</td>
<td>0</td>
<td>1</td>
<td>4426</td>
</tr>
<tr>
<td>College Degree</td>
<td>0.268</td>
<td>0</td>
<td>1</td>
<td>4426</td>
</tr>
<tr>
<td><strong>Mother’s Race</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>0.779</td>
<td>0</td>
<td>1</td>
<td>4426</td>
</tr>
<tr>
<td>Black</td>
<td>0.152</td>
<td>0</td>
<td>0.802</td>
<td>4426</td>
</tr>
<tr>
<td>Other Race</td>
<td>0.069</td>
<td>0</td>
<td>1</td>
<td>4426</td>
</tr>
<tr>
<td><strong>Child Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boy Ratio</td>
<td>0.512</td>
<td>0.222</td>
<td>0.600</td>
<td>5208</td>
</tr>
<tr>
<td>Birth Order</td>
<td>2.026</td>
<td>1.709</td>
<td>3.667</td>
<td>5208</td>
</tr>
</tbody>
</table>

*Note: The summary statistics are from the county level data. The table describes the birth outcomes and family characteristics for each surveyed county from 1990 to 2004. Except for birth order and mother’s age, which are in level, variables are in percentage form.*

Table 2.1. Summary Statistics
<table>
<thead>
<tr>
<th>Birth Outcome</th>
<th>Low Birth Weight</th>
<th>Overweight</th>
<th>Low Apgar Score</th>
<th>Premature</th>
<th>Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>$IPW_{ct}$</td>
<td>-0.0044***</td>
<td>0.0001</td>
<td>0.0110</td>
<td>-0.0034*</td>
<td>-0.0001</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0004)</td>
<td>(0.0090)</td>
<td>(0.0018)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>$IPW_{ct}^2$</td>
<td>0.0006***</td>
<td>0.0004</td>
<td>-0.0006</td>
<td>0.0004</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0001)</td>
<td>(0.0028)</td>
<td>(0.0003)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>Joint Significance</td>
<td>0.0000</td>
<td>0.1828</td>
<td>0.0172</td>
<td>0.0630</td>
<td>0.3416</td>
</tr>
<tr>
<td>First-stage F-test</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Over Identification Test</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Test of Endogeneity</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0180</td>
<td>0.1319</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

*Note:* The control set includes county level average mother’s age at birth; percentage of mothers with less than high school, high school and college education; percentage of married mothers; percentage of mothers who are white, back and other race; ratio of male newborns by county.

The US trade exposure variable $IPW_{ct}$ is instrumented by the exposure to Chinese import competition of eight other high-income countries.

The table presents test results in p-value.

The regression is also weighted by the inverse of county population squared.

*** Indicates significant at the 1% level.

** Indicates significant at the 5% level.

* Indicates significant at the 10% level.

Table 2.2. Effect of IPW on Aggregate birth outcomes (2SLS)
<table>
<thead>
<tr>
<th>Pollution Outcomes</th>
<th>Good (Days)</th>
<th>Moderate</th>
<th>Unhealthy</th>
<th>Very Unhealthy</th>
<th>Hazardous</th>
</tr>
</thead>
<tbody>
<tr>
<td>$IPW_{ct}$</td>
<td>0.0296**</td>
<td>-0.0165</td>
<td>-0.0051**</td>
<td>-0.0004</td>
<td>0.00002</td>
</tr>
<tr>
<td></td>
<td>(0.0127)</td>
<td>(0.0106)</td>
<td>(0.0022)</td>
<td>(0.0012)</td>
<td>(0.00004)</td>
</tr>
<tr>
<td>$IPW_{ct}^2$</td>
<td>-0.0054**</td>
<td>0.0032</td>
<td>0.0007*</td>
<td>0.0001</td>
<td>-0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0022)</td>
<td>(0.0018)</td>
<td>(0.0004)</td>
<td>(0.0002)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>Joint Significance</td>
<td>0.0442</td>
<td>0.2008</td>
<td>0.0685</td>
<td>0.6368</td>
<td>0.5713</td>
</tr>
<tr>
<td>First-stage F-test</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Over Identification Test</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Test of Endogeneity</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.2300</td>
<td>0.8926</td>
<td>0.5214</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CO</th>
<th>NO$_2$</th>
<th>Ozone</th>
<th>SO$_2$</th>
<th>PM2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$IPW_{ct}$</td>
<td>0.0053</td>
<td>0.0095</td>
<td>-0.0180</td>
<td>0.0185</td>
<td>-0.0439**</td>
</tr>
<tr>
<td></td>
<td>(0.0115)</td>
<td>(0.0102)</td>
<td>(0.0178)</td>
<td>(0.0121)</td>
<td>(0.0186)</td>
</tr>
<tr>
<td>$IPW_{ct}^2$</td>
<td>-0.0017</td>
<td>0.0005</td>
<td>0.0028</td>
<td>-0.0008</td>
<td>0.0073**</td>
</tr>
<tr>
<td></td>
<td>(0.0020)</td>
<td>(0.0017)</td>
<td>(0.0030)</td>
<td>(0.0021)</td>
<td>(0.0032)</td>
</tr>
<tr>
<td>Joint Significance</td>
<td>0.4881</td>
<td>0.6092</td>
<td>0.5994</td>
<td>0.0216</td>
<td>0.0569</td>
</tr>
<tr>
<td>First-stage F-test</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Over Identification Test</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Test of Endogeneity</td>
<td>0.4129</td>
<td>0.0418</td>
<td>0.0076</td>
<td>0.5352</td>
<td>0.0000</td>
</tr>
<tr>
<td>Number of Obs.</td>
<td>4,820</td>
<td>4,820</td>
<td>4,820</td>
<td>4,820</td>
<td>4,820</td>
</tr>
</tbody>
</table>

*Data Source:* United States Environmental Protection Agency.
The results are from using fixed effect model and instrumental fixed effect model.
The US trade exposure variable $IPW_{ct}$ is instrumented by the exposure to Chinese import competition of eight other high-income countries.
No other control variables are used in the estimation.
The table presents test results in p-value.
The regression is also weighted by the inverse of county population squared.
*** Indicates significant at the 1% level.
** Indicates significant at the 5% level.
* Indicates significant at the 10% level.

Table 2.3. Trade Exposure and Air Pollution Condition
<table>
<thead>
<tr>
<th></th>
<th>Number of Walmart Stores</th>
</tr>
</thead>
<tbody>
<tr>
<td>$IPW_{ct}$</td>
<td>0.4579***</td>
</tr>
<tr>
<td></td>
<td>(0.0764)</td>
</tr>
<tr>
<td>$IPW_{ct}^2$</td>
<td>-0.0596***</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
</tr>
<tr>
<td>Joint Significance</td>
<td>0.0000</td>
</tr>
<tr>
<td>First-stage F-test</td>
<td>0.0000</td>
</tr>
<tr>
<td>Over Identification Test</td>
<td>0.0000</td>
</tr>
<tr>
<td>Test of Endogeneity</td>
<td>0.0003</td>
</tr>
<tr>
<td>Number of Obs.</td>
<td>4,411</td>
</tr>
</tbody>
</table>

*Data Source: Remote Sensing Data.*

*Note: The results are from using fixed effect model and instrumental fixed effect model.*

The US trade exposure variable $IPW_{ct}$ is instrumented by the exposure to Chinese import competition of eight other high-income countries.

No other control variables are used in the estimation.

The table presents test results in p-value.

The regression is also weighted by the inverse of county population squared.

*** Indicates significant at the 1% level.

** Indicates significant at the 5% level.

* Indicates significant at the 10% level.

Table 2.4. Trade Exposure and Number of Walmart Stores by County
<table>
<thead>
<tr>
<th>Birth Outcome</th>
<th>Low Birth Weight</th>
<th>Overweight</th>
<th>Low Apgar Score</th>
<th>Premature</th>
<th>Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>( IPW_{ct} )</td>
<td>0.0028***</td>
<td>-0.0027***</td>
<td>-0.0253**</td>
<td>-0.0048*</td>
<td>0.0007***</td>
</tr>
<tr>
<td></td>
<td>(0.0012)</td>
<td>(0.0006)</td>
<td>(0.0122)</td>
<td>(0.0027)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>( IPW_{ct}^2 )</td>
<td>-0.0001</td>
<td>0.0008***</td>
<td>0.0248***</td>
<td>0.0031**</td>
<td>-0.0064</td>
</tr>
<tr>
<td></td>
<td>(0.0006)</td>
<td>(0.0003)</td>
<td>(0.0064)</td>
<td>(0.0013)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>( IPW_{ct} \times poverty_{c} )</td>
<td>-0.0651***</td>
<td>0.0299***</td>
<td>0.3478***</td>
<td>0.0334</td>
<td>-0.0064**</td>
</tr>
<tr>
<td></td>
<td>(0.0123)</td>
<td>(0.0055)</td>
<td>(0.0936)</td>
<td>(0.0264)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>( IPW_{ct}^2 \times poverty_{c} )</td>
<td>0.0055</td>
<td>-0.0081</td>
<td>-0.2268***</td>
<td>-0.0304**</td>
<td>-0.0014</td>
</tr>
<tr>
<td></td>
<td>(0.0064)</td>
<td>(0.0029)</td>
<td>(0.0561)</td>
<td>(0.0138)</td>
<td>(0.0013)</td>
</tr>
</tbody>
</table>

Joint Significance: 0.0000 0.0000 0.0156 0.0085 0.0000
First-stage F-test: 0.0000 0.0000 0.0000 0.0000 0.0000
Over Identification Test: 0.0000 0.0000 0.0000 0.0000 0.0000
Test of Endogeneity: 0.0000 0.0000 0.0118 0.0000 0.0000

Note: The control set includes county level average mother’s age at birth; percentage of mothers with less than high school, high school and college education; percentage of married mothers; percentage of mothers who are white, back and other race; ratio of male newborns by county.
The US trade exposure variable \( IPW_{ct} \) is instrumented by the exposure to Chinese import competition of eight other high-income countries.
The regression is also weighted by the inverse of county population squared.
The presented coefficients are rounded to 4 decimal digits, 0.000 does not mean null effect, the significance level is shown by the symbol of *.
The table presents test results in p-value.
*** Indicates significant at the 1% level.
** Indicates significant at the 5% level.
* Indicates significant at the 10% level.

Table 2.5. Effect of IPW on Aggregate Birth Outcomes (2SLS) with Interaction with County Level Per Capita Income
<table>
<thead>
<tr>
<th>Birth Outcome</th>
<th>Low Birth Weight</th>
<th>Overweight</th>
<th>Low Apgar Score</th>
<th>Premature</th>
<th>Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>$IPW_{ct}$</td>
<td>0.0190***</td>
<td>-0.0002</td>
<td>0.0421*</td>
<td>0.0474***</td>
<td>0.0047***</td>
</tr>
<tr>
<td></td>
<td>(0.0030)</td>
<td>(0.0008)</td>
<td>(0.0222)</td>
<td>(0.0065)</td>
<td>(0.0006)</td>
</tr>
<tr>
<td>$IPW_{ct}^2$</td>
<td>-0.0079***</td>
<td>0.0007</td>
<td>-0.0232***</td>
<td>-0.0118***</td>
<td>-0.0016***</td>
</tr>
<tr>
<td></td>
<td>(0.0011)</td>
<td>(0.0002)</td>
<td>(0.0005)</td>
<td>(0.0023)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>$IPW_{ct} \times pcinc_{ct}$</td>
<td>-0.0047***</td>
<td>-0.0001</td>
<td>-0.0101*</td>
<td>-0.0118***</td>
<td>-0.0010***</td>
</tr>
<tr>
<td></td>
<td>(0.0006)</td>
<td>(0.0003)</td>
<td>(0.0056)</td>
<td>(0.0013)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>$IPW_{ct}^2 \times pcinc_{ct}$</td>
<td>0.0018***</td>
<td>-0.0001</td>
<td>0.0065***</td>
<td>0.0028***</td>
<td>0.0003***</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0001)</td>
<td>(0.0022)</td>
<td>(0.0005)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>$pcinc_{ct}$</td>
<td>0.0014*</td>
<td>-0.0001</td>
<td>0.0187***</td>
<td>0.0068***</td>
<td>0.0007***</td>
</tr>
<tr>
<td></td>
<td>(0.0007)</td>
<td>(0.0003)</td>
<td>(0.0056)</td>
<td>(0.0015)</td>
<td>(0.0002)</td>
</tr>
</tbody>
</table>

Note: The control set includes county level average mother’s age at birth; percentage of mothers with less than high school, high school and college education; percentage of married mothers; percentage of mothers who are white, back and other race; ratio of male newborns by county.

The US trade exposure variable $IPW_{ct}$ is instrumented by the exposure to Chinese import competition of eight other high-income countries.

The regression is also weighted by the inverse of county population squared.

The presented coefficients are rounded to 4 decimal digits, 0.000 does not mean null effect, the significance level is shown by the symbol of *.

The table presents test results in p-value.

*** Indicates significant at the 1% level.

** Indicates significant at the 5% level.

* Indicates significant at the 10% level.

Table 2.6. Effect of IPW on Aggregate Birth Outcomes (2SLS) with Interaction with County Level Poverty Rate
CHAPTER 3
PARTIAL IDENTIFICATION OF ECONOMIC MOBILITY: WITH AN APPLICATION TO THE UNITED STATES

3.1. Introduction

There has been substantial interest of late in intra- and inter-generational mobility. Dang et al. (2014, p. 112) state that mobility “is currently at the forefront of policy debates around the world.” Within the popular press, it has been noted that “social mobility ... has become a major focus of political discussion, academic research and popular outrage in the years since the global financial crisis.”¹ In this paper, we study the analysis of economic mobility while accounting for measurement error in income data. Specifically, we offer a new approach to addressing measurement error in the estimation of transition matrices.

Measurement error in income data is known to be pervasive, even in administrative data. In survey data, measurement error arises for two main reasons: misreporting (particularly with retrospective data) and imputation of missing data (Jäntti and Jenkins 2015). It is now taken as given that self-reported income in survey data contain significant measurement error, and that the measurement error is nonclassical in the sense that it is mean-reverting and serially correlated (Duncan and Hill 1985; Bound and Krueger 1991; Bound et al. 1994; Pischke 1995; Pedace and Bates 2000; Bound et al. 2001; Kapteyn and Ypma 2007; Gottschalk and Huynh 2010). Compounding matters, Meyer et al. (2015) find that both problems – nonresponse and accuracy conditional on answering – are worsening over time. In administrative data, measurement error arises for three main reasons: misreporting (tax evasion or filing errors), conceptual differences between the desired and available income measures, and processing errors (Bound et al. 1994; Bound et al. 2001; Kapteyn and Ypma 2007; Pavlopooulos et al. 2012; Abowd and Stinson 2013; Meyer et al. 2015; Obserski et al. 2016).

Even if administrative data are entirely accurate, they are only available in a handful of developed countries.

However, existing studies of mobility either ignore the issue or utilize complex solutions that invoke strong (and often non-transparent) identification assumptions and have data requirements that are quite limiting. The most frequent response to measurement error in the empirical literature on mobility is to mention it as a caveat (Dragoset and Fields 2006). While the usual assumption is that measurement error will lead to upward bias in measures of mobility, the complexity of various mobility measures along with the nonclassical nature of the measurement error makes the direction of any bias uncertain. Glewwe (2012, p. 239) states that “all indices of relative mobility tend to exaggerate mobility if income is measured with error,” yet others offer a different opinion. Dragoset and Fields (2006, p. 1) contend that “very little is known about the degree to which earnings mobility estimates are affected by measurement error.” Gottschalk and Huynh (2010, p. 302) note that “the impact of nonclassical measurement error on mobility is less clear since mobility measures are based on the joint distribution of reported earnings in two periods.”

Our approach to the analysis of mobility given measurement error in income data concentrates on the partial identification of transition matrices. We provide informative bounds on the transition probabilities under minimal assumptions concerning the measurement error process and a variety of nonparametric assumptions on income dynamics. To our knowledge, this is the first study to extend the literature on partial identification to the study of transition matrices (see, e.g., Horowitz and Manski 1995; Manski and Pepper 2000; Kreider and Pepper 2007, 2008; Gundersen and Kreider 2008, 2009; Kreider et al. 2012). Within this environment, we first derive sharp bounds on transition probabilities under minimal assumptions on the measurement error process. We then show how the bounds may be narrowed by imposing more structure via shape restrictions, level set restrictions that relate transition probabilities across observations with different attributes (Manski 1990; Lechner 1999), and monotonicity restrictions that assume monotonic relationships between the true income and certain observed covariates (Manski and Pepper 2000).

In contrast to existing approaches to handle measurement error in studies of mobility (discussed in Section 3.2), our approach has several distinct advantages. First, the assump-
tions invoked to obtain a given set of the bounds are transparent, easily understood by a wide audience, and easy to impose or not impose depending on the particular context. Moreover, bounds on the elements of transition matrices extend naturally to bounds on mobility measures derived from transition matrices. Second, our approach only requires data at two points in time. Third, our approach is easy to implement (through our creation of a generic Stata command). Fourth, our approach extends easily to applications other than income, such as dynamics related to consumption, wealth, occupational status, labor force status, health, student achievement, etc.

The primary drawback to our approach is the lack of point identification. Two responses are in order. First, our approach should be viewed as a complement to, not a replacement for, existing approaches. Indeed, one usefulness of our approach is to provide bounds with which point estimates derived via alternative estimation techniques may be compared. Second, many existing approaches to deal with measurement error in mobility studies end up producing bounds even though the solutions are not couched as a partial identification approach (e.g., Dang et al. 2014; Lee et al. 2017). This arises due to an inability to identify all parameters in some structural model of observed and actual incomes.

Perhaps a secondary drawback of our approach is the focus on transition matrices to capture mobility. Such matrices have the disadvantage of not providing a scalar measure of mobility, simplifying spatial and temporal comparisons of mobility. While there is merit to this critique, there are several responses. First, transition matrices are an obvious starting point in the measurement of mobility. Jäntti and Jenkins (2015, p. 822) argue that, when measuring mobility across two points in time, “the bivariate joint distribution of income contains all the information there is about mobility, so a natural way to begin is by summarizing the joint distribution in tabular or graphical form.” Second, transition matrices are easily understood by policymakers and the general public and thus are frequently referenced within these domains. The importance of this cannot be oversold. For example, a recent article in The New Yorker (March 26, 2014) argued that an “essential part” of the work by Piketty and others as it relates to inequality is the presentation of the data in a manner that is “easier to understand” through the avoidance of “clever but complicated statistics ... which

---

2 Available at [http://faculty.smu.edu/millimet/code.html](http://faculty.smu.edu/millimet/code.html).
attempted to reduce the entire income distribution to a single number.\textsuperscript{3} Third, transition matrices allow one to examine mobility at different parts of the income distribution (Lee et al. 2017). Finally, bounds on (scalar) measures of mobility derived from the elements of the transition matrix are easily obtained from our approach.

We illustrate our approach examining intra-generational mobility in the United States using data from the Survey of Income and Program Participation (SIPP) for the United States. Specifically, we examine mobility over two four-year periods, 2004 to 2008 and then 2008 to 2012. Understanding mobility patterns in the US is important as there is convincing evidence that income inequality has been increasing in the US. However, the welfare impact of this rise depends crucially on the level of economic mobility.\textsuperscript{4} Shorrocks (1978, p. 1013) argues that “evidence on inequality of incomes or wealth cannot be satisfactorily evaluated without knowing, for example, how many of the less affluent will move up the distribution later in life.” More recently, Kopczuk et al. (2010, p. 91-2) conclude that “a comprehensive analysis of disparity requires studying both inequality and mobility” as “annual earnings inequality might substantially exaggerate the extent of true economic disparity among individuals.”

Our analysis of US mobility yields some striking results. First, we show that relatively small amounts of measurement error leads to bounds that can be quite wide in the absence of other information or restrictions. Second, the restrictions considered contain significant identifying power as the bounds can be severely narrowed. Finally, allowing for misclassification errors in up to 10% of the sample, we find that the probability of being in poverty (out of poverty) four years later conditional on being in poverty (out of poverty) in the initial period is at least 33% (87%) under our most restrictive set of assumptions.

The rest of the paper is organized as follows. Section 3.2 provides a brief review of existing approaches to address measurement error in studies of mobility. Section 3.3 presents our partial identification approach. Section 3.4 contains the empirical application. Section 3.5 concludes.


\textsuperscript{4}The level of income inequality in the US has followed a U-shaped pattern over the past century (Picketty and Saez 2003; Kopczuk et al. 2010; Atkinson and Bourguignon 2015).
3.2. Literature Review

Burkhauser and Couch (2009) and Jäntti and Jenkins (2015) provide excellent reviews of the numerous mobility measures. Bound et al. (2001) and Meyer et al. (2015) offer excellent surveys regarding measurement error in microeconomic data. Here, we focus on approaches that have been taken to address (or not address) measurement error in analyses of economic mobility. We identify three general approaches in the existing literature: (i) ignore it, (ii) \textit{ad hoc} data approaches, and (iii) \textit{structural} approaches. In the interest of brevity, we relegate much of the discussion of the prior literature to the appendix. Here, we discuss only those methods most comparable to our approach. These methods fall within the third category and utilize structural models to simulate error-free income. Armed with the simulated data, any mobility measure may be computed, including transition matrices. Clearly, the validity of this approach rests on the quality of the simulated error-free data. Obtaining simulated values of error-free data is not trivial and typically relies on complex models invoking a number of fairly opaque assumptions.

Studies pursuing this strategy include McGarry (1995), Glewwe and Dang (2011), Pavlopoulos et al. (2012), Dang et al. (2014), and Lee et al. (2017). McGarry (1995) posits a variance components model to isolate the portion of observed income that represents measurement error. Upon simulating error-free income, conditional staying probabilities for the poor are examined. The results indicate substantially less mobility in the simulated data. However, the model defines measurement error as the individual-level, time-varying, serially uncorrelated component of income. Thus, all time-varying idiosyncratic sources of income variation are removed. Moreover, the individual-level, time-varying, serially correlated component of income is not considered measurement error. Finally, parametric distributional assumptions are required for identification in practice.

Glewwe and Dang (2011) begin with the assumption that log income follows an AR(1) process. The authors then combine OLS and IV estimates of the forward and reverse regressions, along with assumptions about the variance components of the model, to simulate error-free income. The simulated data are then used to assess income growth across the distribution. As in McGarry (1995), the results suggest substantial bias from measurement error. However, as in McGarry (1995), identification of error-free income relies on strong
assumptions for identification, such as serially uncorrelated measurement error, particular functional forms, and valid instrumental variables.

Pavlopoulos et al. (2012) build on Rendtel et al. (1998) and specify a mixed latent Markov model to examine error-free transitions between low pay, high pay, and non-employment. The model requires data from at least three periods, as well as requires perhaps strong assumptions concerning unobserved heterogeneity and initial conditions. In addition, serial correlation in measurement error is difficult to address and extending the model to more than three states is problematic. Nonetheless, the results align with the preceding studies in that mobility is dampened once measurement error is addressed.

Dang et al. (2014) consider the measurement of mobility using pseudo-panel data. Since the same individuals are not observed in multiple periods, the authors posit a static model of income using only time invariant covariates available in all periods. The model estimates, along with various assumptions concerning how unobserved determinants of income are correlated over time, are used to bound measures of a two-by-two poverty transition matrix. This approach implicitly addresses measurement error through the imputation process as missing data can be considered an extreme form of measurement error. However, measurement error in observed incomes used to estimate the static model and compute the poverty transition matrix is not addressed. Moreover, it is not clear how one could extend the method to estimate more disaggregate transition matrices.

Finally, Lee et al. (2017) estimates a complex model based on an AR(1) model of consumption dynamics with time invariant and time-varying sources of measurement error to simulate error-free consumption and estimate transition matrices. Consistent with the preceding studies, significantly less mobility is found in the simulated data. While the authors’ model has some advantages compared to earlier attempts to simulate error-free outcomes, these advantages come at a cost of increased complexity, decreased transparency of the identifying assumptions, and a need for four periods of data. In addition, bounds are obtained as not all parameters required for the simulations are identified.

In summary, the literature on addressing measurement error in studies of mobility has witnessed significant recent growth. However, there remains much scope for additional work. While simulation-based methods allow for estimation of transition matrices, these methods
are complex, lack transparency, rely on strong functional form and distributional assumptions, and often require more than two years of data. Moreover, the common reliance in the majority of the simulation approaches on an AR(1) model of income or consumption dynamics is worrisome. Lee et al. (2017, p. 38) acknowledge that “this model is not so much derived from a well-developed theory, but it is a convenient reduced-form model.” Finally, the reliance on precise assumptions concerning the nature of the variance components is unappealing in light of Kapteyn and Ympa’s (2007, p. 535) finding that “substantive conclusions may be affected quite a bit by changes in assumptions on the nature of error in survey and administrative data.”

Our proposed approach complements these existing approaches. However, in contrast to simulation approaches, which often end up with bounds on transition probabilities, we set out to estimate bounds from the beginning, making it transparent exactly how the bounds are affected by each assumption one may wish to impose. Furthermore, the assumptions imposed to narrow the bounds are much easier for non-experts to comprehend.

3.3. Model

3.3.1. Setup

Let $y^*_{it}$ denote the true income for observation $i$, $i = 1, ..., N$, in period $t$, $t = 0, 1$. An observation may refer to an individual or household observed at two points in time in the case of intragenerational mobility or a parent-child pair observed at two points in time in the case of intergenerational mobility. Further, let $F_{0,1}(y^*_{0}, y^*_{1})$ denote the joint (bivariate) cumulative distribution function (CDF), where $y^*_{t} \equiv [y^*_{1t} \cdots y^*_{Nt}]$.

While movement through the distribution from an initial period, 0, to a subsequent period, 1, is completely captured by $F_{0,1}(y^*_{0}, y^*_{1})$, this is not practical. Moreover, policymakers and the media often focus on more easily understood transition matrices. A $K \times K$ transition
matrix, $P^*_{0,1}$, summarizes this joint distribution and is given by

$$P^*_{0,1} = \begin{bmatrix}
  p^*_{11} & \cdots & \cdots & p^*_{1K} \\
  \vdots & \ddots & \ddots & \vdots \\
  \vdots & \ddots & \ddots & \vdots \\
  p^*_{K1} & \cdots & \cdots & p^*_{KK}
\end{bmatrix}. \quad (3.1)$$

Elements of this matrix have the following form

$$p^*_{kl} = \frac{\Pr(y^*_0 \in k, y^*_1 \in l)}{\Pr(y^*_0 \in k)} \quad k, l = 1, ..., K,$$

where the $\zeta$s are cutoff points between the $K$ partitions such that $0 = \zeta^t_0 < \zeta^1_1 < \zeta^2_2 < \cdots < \zeta^t_{K-1} < \zeta^t_K < \infty, t = 0, 1$.\(^5\) Thus, $p^*_{kl}$ is a conditional probability. A complete lack of mobility implies $p^*_{kl}$ equals unity if $k = l$ and zero otherwise.\(^6\) Finally, we can define conditional transition matrices, conditioned upon $X = x$, where $X$ denotes a vector of observation attributes. Denote the conditional transition matrix as $P^*_{0,1}(x)$, with elements given by

$$p^*_{kl}(x) = \frac{\Pr(y^*_0 \in k, y^*_1 \in l|X = x)}{\Pr(y^*_0 \in k|X = x)} \quad k, l = 1, ..., K.$$ 

\(^5\)For example, if $K = 5$, then the cutoff points might correspond to quintiles within the two marginal distributions of $y^*_0$ and $y^*_1$.

\(^6\)In contrast, ‘perfect’ mobility may be characterized by origin-destination independence, implying $p^*_{kl} = 1/K$ for all $k, l$, or by complete rank reversal, implying $p^*_{kl} = 1$ if $k + l = K + 1$ and zero otherwise. See Jäntti and Jenkins (2015) for discussion.
Implicit in this definition is the assumption that $X$ includes only time invariant attributes.\footnote{Note, while the probabilities are conditional on $X$, the cutoff points $\zeta$ are not. Thus, we are capturing movements within the \textit{overall} distribution among those with $X = x$.}

For clarity, throughout the paper we consider two types of transition matrices: (i) those with equal-sized partitions and (ii) those with unequal-sized partitions. With equal-sized partitions, the $\zeta$s are chosen such that each partition contains $1/K$ of the population. For example, equal-sized partitions with $K = 5$ correspond to a quintile transition matrix. In this case, the rows and columns of $P_{0,1}^*$ sum to one and mobility is necessarily zero-sum (i.e., if an observation is misclassified in the upward direction, there must be at least one observation misclassified in the downward direction). With unequal-sized partitions, only the rows of $P_{0,1}^*$ sum to one and mobility is not zero-sum. For example, we shall consider the case of a $2 \times 2$ poverty transition matrix, where $\zeta_t^1$ is the poverty line in period $t$.

Given the definition of $P_{0,1}^*$ or $P_{0,1}^*(x)$, our objective is to learn something about its elements. With a random sample $\{y_{it}, x_i\}$ and a choice of $K$ and the $\zeta$s, the transition probabilities are point identified as they are functions of nonparametrically estimable quantities. The corresponding plug-in estimator is consistent. However, as stated previously, ample evidence indicates that income is measured with error. Let $y_{it}$ denote the observed income for observation $i$ in period $t$. With data $\{y_{it}, x_i\}$ and a choice of $K$ and the $\zeta$s, the empirical transition probabilities are inconsistent for $p_{kl}^*$ and $p_{kl}^*(x)$.

With access only to data containing measurement error, our goal is to bound the probabilities given in (3.2) and (3.3). The relationships between the true partitions of $\{y_{it}\}_{t=0}^1$ and the observed partitions of $\{y_{it}\}_{t=0}^1$ are characterized by the following joint probabilities:

$$\theta_{(k,l)}^{(k'-k, l'-l)} = \Pr(y_0 \in k', \ y_1 \in l', \ y_0^* \in k, \ y_1^* \in l)$$

(3.4)

where the subscript $(k,l)$ indexes the true partitions in period 0 and 1 and the superscript $(k'-k, l'-l)$ indicates the degree of misclassification given by the differences between the observed partitions $k'$ and $l'$ and true partitions $k$ and $l$. If $k' - k, l' - l > 0$, then there is \textit{upward} misclassification in both periods. If $k' - k, l' - l < 0$, then there is \textit{downward} misclassification in both periods. If $k' - k$ and $l' - l$ are of different signs, then the di-
rection of misclassification changes across periods. \( \theta_{(k,l)}^{(0,0)} \) represents the probability of no misclassification in either period for an observation with true income in partitions \( k \) and \( l \).8

With this notation, we can now rewrite the elements of \( P_{0,1}^* \) as

\[
\begin{align*}
    p_{kl}^* &= \frac{\Pr(y_0^* \in k, y_1^* \in l)}{\Pr(y_0^* \in k)} \\
    &= \frac{\Pr(y_0 \in k, y_1 \in l) + \sum_{k',l'=1,2,...,K} \theta_{(k',l')}^{(k' \neq k, l \neq l)} - \sum_{k',l'=1,2,...,K} \theta_{(k',l')}^{(k' \neq k, l \neq l)}}{\Pr(y_0 \in k) + \sum_{k',l'=1,2,...,K} \theta_{(k',l')}^{(k' \neq k, l \neq l)} - \sum_{k',l'=1,2,...,K} \theta_{(k',l')}^{(k' \neq k, l \neq l)}} \\
    &\equiv r_{kl} + Q_{1,kl} - Q_{2,kl} \\
    &= p_k + Q_{3,k} - Q_{4,k} \\
    &\equiv K(r_{kl} + Q_{1,kl} - Q_{2,kl}),
\end{align*}
\]

where the final line holds only in the case of equal-sized partitions.9 \( Q_{1,kl} \) measures the proportion of false negatives associated with partition \( kl \) (i.e., the probability of being misclassified conditional on \( kl \) being the true partition). \( Q_{2,kl} \) measures the proportion of false positives associated with partition \( kl \) (i.e., the probability of being misclassified conditional on \( kl \) being the observed partition). Similarly, \( Q_{3,k} \) and \( Q_{4,k} \) measure the proportion of false negatives and positives associated with partition \( k \), respectively.

The transition probabilities in (3.5) and (3.6) are not identified from the data alone. The data identify \( r_{kl} \) and \( p_k \), but not the misclassification parameters, \( \theta \). One can compute sharp bounds by searching across the unknown misclassification parameters. There are \( K^2(K^2-1) \) misclassification parameters in \( P_{0,1}^* \). However, several constraints must hold (see Appendix G). Even with these constraints, obtaining informative bounds on the transition probabilities is not possible without further restrictions. Section 3.2 considers assumptions

---

8\( \theta_{(k,l)}^{(0,0)} \) may be strictly positive even though income is misreported in either or both periods (i.e., \( y_{it} \neq y_{it}^* \) for at least some \( i \) and \( t \)) as long as the misreporting is not so severe as to invalidate the observed partitions (i.e., \( k' = k \) and \( l' = l \) regardless). Throughout the paper, we use the term measurement error to refer to errors in observed income (\( y_{it} \neq y_{it}^* \)) and misclassification to refer to errors in the observed partition (\( k' \neq k \) or \( l' \neq l \)).

9The expression in (3.5) is identical to that in Gundersen and Kreider (2008, p. 368) when \( K = 2 \).
on the $\theta$s. Section 3.3 considers restrictions on the underlying mobility process.

3.3.2. Misclassification

3.3.2.1. Assumptions

Given the presence of measurement error, we obtain bounds on the elements of $P_{0,1}^*$, given in (3.5). We consider the following misclassification assumptions.

**Assumption 1** (Rank Preserving Measurement Error). *Misreporting does not alter an observation’s rank in the income distribution in either period.* Formally, defining $F_t(y_{it})$ and $F_t^*(y_{it}^*)$, $t = 0, 1$, as the marginal CDFs of observed and true income in each period, then

$$F_t(y_{it}) = F_t^*(y_{it}^*) \forall i, t$$

$$\implies Q_{j,kl} = 0, \ j = 1, ..., 4, \ \forall k, l.$$ 

**Assumption 2** (Maximum Misclassification Rate).

(i) (Arbitrary Misclassification) *The total misclassification rate in the data is bounded from above by $Q \in (0, 1)$.* Formally,

$$\sum_{k=1}^{K} \sum_{l=1}^{K} \sum_{j=1}^{4} Q_{j,kl} \leq Q.$$ $\quad (3.7)$

(ii) (Uniform Misclassification) *The total misclassification rate in the data is bounded from above by $Q \in (0, 1)$ and is uniformly distributed across partitions.* Formally,

---

$^{10}$In the interest of brevity, we focus attention from here primarily on the unconditional transition matrix. We return to the conditional transition matrix in Section 3.3.
\[ \sum_{l=1}^{K} Q_{1,kl} + \sum_{l=1}^{K} Q_{2,kl} + \sum_{l=1}^{K} Q_{3,k} + \sum_{l=1}^{K} Q_{4,k} \leq \frac{Q}{K} \quad \forall k \quad (3.8) \]

\[ \sum_{k=1}^{K} Q_{1,kt} + \sum_{k=1}^{K} Q_{2,kl} \leq \frac{Q}{K} \quad \forall l. \quad (3.9) \]

Assumption 1 is similar to Heckman et al.’s (1997) rank invariance assumption in the context of the distribution of potential outcomes in a treatment effects framework. Although this assumption is quite strong, it is a useful benchmark. Assumption 2 places restrictions on the total amount of misclassification allowed in the data. As we discuss below, the amount of misclassification is dependent on the choice of \( K \). As such, one could express \( Q \) as \( Q(K) \); we dispense with this notation for expositional purposes.

For the case of equal-sized partitions, misclassification is necessarily zero-sum; upward misclassification of some observations necessarily implies downward misclassification of others. Thus, even if measurement error in income is uni-directional, misclassification errors must be bi-directional. However, for the case of unequal-sized partitions, this need not be the case. In such cases, we also consider adding the following assumption.

**Assumption 3** (Uni-Directional Misclassification). *Misclassification occurs strictly in the upward direction.* Formally,

\[ Q_{2,11} = Q_{4,1} = Q_{1,KK} = Q_{3,K} = 0. \]

Assumption 3 rules out the possibility of any false positives (negatives) occurring in the worst (best) partition. Note, this assumption is consistent with mean-reverting measurement error as long as the negative measurement errors for observations with high income are not sufficient to lead to misclassification. For example, if \( P_{0,1}^* \) is a \( 2 \times 2 \) poverty transition matrix, Assumption 3 permits observations with true incomes exceeding the poverty threshold to underreport income, but not to a degree whereby they are misclassified as in poverty. Such an assumption may not hold, for instance, if some households above the poverty threshold report
incomes below the poverty threshold in an attempt to qualify for means-tested transfers.

3.3.2.2. Bounds

3.3.2.2.1 Rank Preserving Measurement Error (Assumption 1) Under Assumption 1 the sampling process identifies the transition probabilities despite the presence of measurement error, yielding the following proposition.

**Proposition 1.** Under Assumption 1, the transition probabilities are identified by

\[ p_{kl}^* = \frac{\Pr(y_0 \in k, y_1 \in l)}{\Pr(y_0 \in k)} = \frac{\left[ I(y_0 \in k, y_1 \in l) \right]}{I(y_0 \in k)} \, . \]  \hspace{1cm} (3.10)

The transition probabilities are nonparametrically identified. Estimation proceeds by replacing the terms with their sample analogs, given by

\[ \hat{p}_{kl} = \frac{\sum_i I(y_{0i} \in k, y_{1i} \in l)}{\sum_i I(y_{0i} \in k)} \, \frac{1}{K} \sum_i I(y_{0i} \in k, y_{1i} \in l) \]  \hspace{1.1cm} (3.11)

where the last line follows in the case of equal-sized partitions. Proof: See Appendix G.

3.3.2.2.2 Maximum Misclassification Rate (Assumption 2) Under Assumption 2 if \( Q > 0 \), the transition probabilities are no longer nonparametrically identified. We have the following propositions.

**Proposition 2.** Consider a transition matrix, \( P_{0,1}^* \), with equal-sized partitions. Under Assumption 2(i), the transition probabilities are bounded sharply by

\[ \max\{K(r_{kl} - Q), 0\} \leq p_{kl}^* \leq \min\{K(r_{kl} + Q), 1\} \]  \hspace{1cm} (3.13)

Under Assumption 2(ii), the transition probabilities are bounded sharply by

\[ \max\{K(r_{kl} - Q/K), 0\} \leq p_{kl}^* \leq \min\{K(r_{kl} + Q/K), 1\} \]  \hspace{1cm} (3.14)

Proof: See Kreider and Pepper (2008, p. 335) and Horowitz and Manski (1995, Corollary 1.2).
Proposition 3. Consider a transition matrix, $P_{0,1}^*$, with unequal-sized partitions. Under Assumption 2, the transition probabilities are bounded sharply by

$$\max\left\{ \frac{r_{kl} - Q_{2,kl}}{p_k + Q_{3,k}}, 0 \right\} \leq p_{kl}^* \leq \min\left\{ \frac{r_{kl} + Q_{1,kl}}{p_k - Q_{4,k}}, 1 \right\}$$

(3.15)

where the following constraints must hold

\begin{align*}
0 \leq Q_{1,kl} & \leq 1 - r_{kl} \\
0 \leq Q_{2,kl} & \leq r_{kl} \\
0 \leq Q_{3,k} & \leq 1 - p_k \\
0 \leq Q_{4,k} & \leq p_k
\end{align*}

in addition to

\begin{align*}
Q_{1,kl} + Q_{4,k} & \leq Q \\
Q_{2,kl} + Q_{3,k} & \leq Q
\end{align*}

under Assumption 2(i) or

\begin{align*}
Q_{1,kl} + Q_{4,k} & \leq Q/K \\
Q_{2,kl} + Q_{3,k} & \leq Q/K
\end{align*}

under Assumption 2(ii). Proof: See Appendix G.

Estimation of the bounds in Propositions 2 and 3 proceeds by replacing $r_{kl}$ and $p_k$ with their sample analogs. Estimation of the lower bound in Proposition 3 requires the additional step of minimizing the lower bound with respect to $Q_2$ and $Q_3$ subject to the appropriate constraints. Similarly, estimation of the upper bound Proposition 3 requires maximizing the upper bound with respect to $Q_1$ and $Q_4$ subject to the appropriate constraints.

3.3.2.2.3 Uni-Directional Misclassification (Assumption 3) For simplicity, we only consider Assumption 3 in the case of a $2 \times 2$ transition matrix. We have the following proposition.
Proposition 4. Under Assumptions 2(i) and 3, the four elements of a $2 \times 2$ transition matrix with unequal-sized partitions are bounded sharply by

$$\frac{r_{11}}{\min\{p_1 + Q, 1\}} \leq p_{11}^* \leq \min\left\{\frac{r_{11} + Q}{p_1}, 1\right\}$$  \hspace{1cm} \text{(3.16)}$$

$$\max\left\{\frac{r_{12} - Q}{p_1}, 0\right\} \leq p_{12}^* \leq \min\left\{\frac{r_{12} + Q}{p_1 + Q}, 1\right\}$$  \hspace{1cm} \text{(3.17)}$$

$$\max\left\{\frac{r_{21} - Q}{p_2 - Q}, 0\right\} \leq p_{21}^* \leq \min\left\{\frac{r_{21} + Q}{\max\{p_2 - Q, 0\}}, 1\right\}$$  \hspace{1cm} \text{(3.18)}$$

$$\max\left\{\frac{r_{22} - Q}{p_2}, 0\right\} \leq p_{22}^* \leq \min\left\{\frac{r_{22}}{\min\{p_2 - Q, 0\}}, 1\right\}$$  \hspace{1cm} \text{(3.19)}$$

where the following constraints must hold

$$Q_1, Q_4 \geq 0$$ \hspace{1cm} \text{(3.20)}$$

$$Q_1 + Q_4 \leq Q.$$

Under Assumptions 2(ii) and 3, sharp bounds are identified by replacing $Q$ with $Q/2$ in (3.16)-(3.20). Proof: See Appendix G.

Estimation of the bounds are straightforward using the appropriate sample analogs with the exception of the upper bound for $p_{21}^*$.

3.3.3. Restrictions

Propositions 2-4 provide bounds on transition probabilities considering only restrictions on the misclassification process. Here, we explore the identifying power of incorporating restrictions on the mobility process. The restrictions may be imposed alone or in combination.

3.3.3.1. Shape Restrictions

Shape restrictions place inequality constraints on the population transition probabilities. Here, we consider imposing shape restrictions assuming that large transitions are less likely than smaller ones.
Assumption 4 (Shape Restrictions). The transition probabilities are weakly decreasing in the size of the transition. Formally, \( p_{kl}^* \) is weakly decreasing in \( |k - l| \), the absolute difference between \( k \) and \( l \).

This assumption implies that within each row or each column of the transition matrix, the diagonal element (i.e., the conditional staying probability) is the largest. The remaining elements decline weakly monotonically moving away from the diagonal element. This leads to the following proposition.

**Proposition 5.** Denote the bounds on \( p_{kl}^* \) under some combination of Assumptions 2-3 as
\[
LB_{kl} \leq p_{kl}^* \leq UB_{kl}.
\]
Adding Assumption 4 implies the following sharp bounds:

\[
\begin{align*}
\max \left\{ \sup_{l' = 1, \ldots, K} LB_{kl'}, \sup_{k' = 1, \ldots, K} LB_{k'l} \right\} & \leq p_{kl}^* \leq UB_{kl} \quad \text{if } k = l \\
\max \left\{ \sup_{l' \geq l} LB_{kl'}, \sup_{k' \leq k} LB_{k'l} \right\} & \leq p_{kl}^* \leq \min \left\{ \inf_{k \leq k' \leq l} UB_{kl'}, \inf_{k \leq k' \leq l} UB_{k'l} \right\} \quad \text{if } k < l \\
\max \left\{ \sup_{l' \leq l} LB_{kl'}, \sup_{k' \geq k} LB_{k'l} \right\} & \leq p_{kl}^* \leq \min \left\{ \inf_{l \leq l' \leq k} UB_{kl'}, \inf_{l \leq l' \leq k} UB_{k'l} \right\} \quad \text{if } k > l
\end{align*}
\]

**Proof:** See Appendix G.

Estimation is straightforward given estimates of the preliminary bounds, \( LB_{kl} \) and \( UB_{kl} \).

3.3.3.2. Level Set Restrictions

Level set restrictions place equality constraints on population transition probabilities across observations with different observed attributes (Manski 1990; Lechner 1999).

Assumption 5 (Level Set Restrictions). The conditional transition probabilities, given in (3.3), are constant across a range of conditioning values. Formally, \( p_{kl}^*(x) \) is constant for all \( x \in \mathcal{A}_x \subset \mathcal{R}_m \), where \( x \) is an \( m \)-dimensional vector.

For instance, if \( x \) denotes the age of an individual in years, one might wish to assume that \( p_{kl}^*(x) \) is constant for all \( x \) within a five-year window around \( x \).
From (3.3) and (3.5), we have

\[
p_{kl}^*(x) = \frac{\Pr(y_0 \in k, y_1 \in l | X = x) + \sum_{k',l',\tilde{m},\tilde{l}=1,2,...,K, k' \neq k, l' \neq l} \theta_{(k,l)}^{(k'-k,l'-\tilde{l})}(x) - \sum_{k',l',\tilde{m},\tilde{l}=1,2,...,K, k' \neq k, l' \neq l} \theta_{(k',\tilde{m},\tilde{l})}^{(k'-k,\tilde{m},\tilde{l})}(x)}{\Pr(y_0 \in k | X = x) + \sum_{k',\tilde{m},\tilde{l}=1,2,...,K, k' \neq k} \theta_{(k',\tilde{m},\tilde{l})}^{(k'-k,\tilde{m},\tilde{l})}(x) - \sum_{k',\tilde{m},\tilde{l}=1,2,...,K, k' \neq k} \theta_{(k',\tilde{m},\tilde{l})}^{(k'-k,\tilde{m},\tilde{l})}(x)}
\]

\[
\equiv \frac{r_{kl}(x) + Q_{1,kl}(x) - Q_{2,kl}(x)}{p_k(x) + Q_{3,k}(x) - Q_{4,k}(x)}
\]

(3.24)

where now \(Q_j(x), j = 1, ..., 4\), represent the proportions of false positives and negatives conditional on \(x\). As such, we also consider the following assumption regarding the conditional misclassification probabilities.

**Assumption 6** (Independence). Misclassification rates are independent of the observed attributes of observations. Formally,

\[
Q_{j}(x) = Q_{j}, j = 1, ..., 4.
\]

This leads to the following proposition.

**Proposition 6.** Denote the bounds for \(p_{kl}^*(x)\) under some combination of Assumptions 2-4 and 6 as

\[
LB(x) \leq p_{kl}^*(x) \leq UB(x).
\]

(3.25)

Adding Assumption 5 implies the following sharp bounds on the conditional transition probabilities:

\[
\sup_{z \in A_x} LB(z) \leq p_{kl}^*(x) \leq \inf_{z \in A_x} UB(z).
\]

(3.26)

Assuming \(X\) is discrete, sharp bounds on the unconditional transition probabilities are given as

\[
\sum_x \Pr(X = x) \left( \sup_{z \in A_x} LB(z) \right) \leq p_{kl}^* \leq \sum_x \Pr(X = x) \left( \inf_{z \in A_x} UB(z) \right).
\]

(3.27)

**Proof:** See Manski and Pepper (2000).

To operationalize Proposition 6, bounds on the conditional transition probabilities, shown in (3.25), must be obtained. This is done in the following corollaries.
Corollary 6.1. Consider a transition matrix, $P_{0,1}^*$, with equal-sized partitions. Under Assumptions 5, 2(i) and 6, $p_{kl}^*(x)$ is bounded sharply by

$$\max \left\{ \frac{1}{p_k(x)} (r_{kl}(x) - Q), 0 \right\} \leq p_{kl}^*(x) \leq \min \left\{ \frac{1}{p_k(x)} (r_{kl}(x) + Q), 1 \right\}.$$  \hspace{1cm} (3.28)

Under Assumptions 2(ii) and 6, $p_{kl}^*(x)$ is bounded sharply by

$$\max \left\{ \frac{1}{p_k(x)} (r_{kl}(x) - Q/K), 0 \right\} \leq p_{kl}^*(x) \leq \min \left\{ \frac{1}{p_k(x)} (r_{kl}(x) + Q/K), 1 \right\}.$$  \hspace{1cm} (3.29)

Under Assumption 2(i) or 2(ii) without Assumption 6, $p_{kl}^*(x)$ is bounded sharply by

$$\max \left\{ \frac{r_{kl}(x) - Q_{2,kl}(x)}{p_k(x) + Q_{3,k}(x)}, 0 \right\} \leq p_{kl}^*(x) \leq \min \left\{ \frac{r_{kl}(x) + Q_{1,kl}(x)}{p_k(x) - Q_{4,k}(x)}, 1 \right\},$$  \hspace{1cm} (3.30)

where the following constraints must hold

$$0 \leq Q_{1,kl}(x) \leq 1 - r_{kl}(x)$$
$$0 \leq Q_{2,kl}(x) \leq r_{kl}(x)$$
$$0 \leq Q_{3,k}(x) \leq \min \left\{ 1 - p_k(x), \frac{1}{K \cdot \Pr(X = x)} - p_k(x) \right\}$$
$$0 \leq Q_{4,k}(x) \leq \min \left\{ p_k(x), p_k(x) - \frac{1}{K - [1 - \Pr(X = x)] \Pr(X = x)} \right\}$$

in addition to

$$Q_{1,kl}(x) + Q_{4,k}(x) \leq \min \{Q/\Pr(X = x), 1\}$$
$$Q_{2,kl}(x) + Q_{3,k}(x) \leq \min \{Q/\Pr(X = x), 1\}$$

under Assumption 2(i) or

$$Q_{1,kl}(x) + Q_{4,k}(x) \leq \min \{Q/\lceil K \cdot \Pr(X = x) \rceil, 1\}$$
$$Q_{2,kl}(x) + Q_{3,k}(x) \leq \min \{Q/\lceil K \cdot \Pr(X = x) \rceil, 1\}$$

under Assumption 2(ii). Proof: See Appendix G.
Corollary 6.2. Consider a transition matrix, $P_{0,1}^*$, with unequal-sized partitions. Under Assumption 5, 2(i) or 2(ii) and 6, $p_{kl}^*(x)$ is bounded sharply by

$$\max \left\{ \frac{r_{kl}(x) - Q_{2,kl}}{p_k(x) + Q_{3,k}}, 0 \right\} \leq p_{kl}^*(x) \leq \min \left\{ \frac{r_{kl}(x) + Q_{1,kl}}{p_k(x) - Q_{4,k}}, 1 \right\}$$

(3.31)

where the following constraints must hold

\begin{align*}
0 \leq Q_{1,kl} & \leq 1 - r_{kl}(x) \\
0 \leq Q_{2,kl} & \leq r_{kl}(x) \\
0 \leq Q_{3,l} & \leq 1 - p_k(x) \\
0 \leq Q_{4,l} & \leq p_k(x)
\end{align*}

in addition to

\begin{align*}
Q_{1,kl} + Q_{4,l} & \leq Q \\
Q_{2,kl} + Q_{3,l} & \leq Q
\end{align*}

under Assumption 2(i) or

\begin{align*}
Q_{1,kl} + Q_{4,l} & \leq Q/K \\
Q_{2,kl} + Q_{3,l} & \leq Q/K
\end{align*}

under Assumption 2(ii). Under Assumption 2(i) or 2(ii) without 6, sharp bounds for $p_{kl}^*(x)$ are given in (3.30), where the following constraints must hold

\begin{align*}
0 \leq Q_{1,kl}(x) & \leq 1 - r_{kl}(x) \\
0 \leq Q_{2,kl}(x) & \leq r_{kl}(x) \\
0 \leq Q_{3,k}(x) & \leq 1 - p_k(x) \\
0 \leq Q_{4,k}(x) & \leq p_k(x)
\end{align*}

in addition to

\begin{align*}
Q_{1,kl}(x) + Q_{4,k}(x) & \leq \min\{Q/\Pr(X = x), 1\} \\
Q_{2,kl}(x) + Q_{3,k}(x) & \leq \min\{Q/\Pr(X = x), 1\}
\end{align*}
under Assumption 2(i) or

\[
Q_{1,kl}(x) + Q_{4,k}(x) \leq \min\{Q/[K \cdot \Pr(X = x)], 1\}
\]
\[
Q_{2,kl}(x) + Q_{3,k}(x) \leq \min\{Q/[K \cdot \Pr(X = x)], 1\}
\]

under Assumption 2(ii). Proof: See Appendix G.

**Corollary 6.3.** Consider a \(2 \times 2\) transition matrix, \(P^*_{0,1}\), with unequal-sized partitions. Under Assumption 3, 5, \(p^*_{kl}(x)\) is bounded sharply by

\[
\frac{r_{11}(x)}{\min\{p_1(x) + Q, 1\}} \leq p^*_{11}(x) \leq \min\left\{\frac{r_{11}(x) + Q}{p_1(x)}, 1\right\}
\]
(3.32)

\[
\max\left\{\frac{r_{12}(x) - Q}{p_1(x)}, 0\right\} \leq p^*_{12}(x) \leq \min\left\{\frac{r_{12}(x) + Q}{p_1(x) + Q}, 1\right\}
\]
(3.33)

\[
\max\left\{\frac{r_{21}(x) - Q}{p_2(x) - Q}, 0\right\} \leq p^*_{21}(x) \leq \min\left\{\frac{r_{21}(x) + Q\tilde{Q}_1}{\max\{p_2(x) - Q\tilde{Q}_4, 0\}}, 1\right\}
\]
(3.34)

\[
\max\left\{\frac{r_{22}(x) - Q}{p_2(x)}, 0\right\} \leq p^*_{22}(x) \leq \max\left\{\frac{r_{22}(x)}{\min\{p_2(x) - Q, 0\}}, 1\right\}
\]
(3.35)

where the following constraints must hold

\[
\tilde{Q}_1, \tilde{Q}_4 \geq 0
\]
(3.36)

\[
\tilde{Q}_1 + \tilde{Q}_4 \leq Q
\]

under Assumptions 2(i) and 6 or

\[
\tilde{Q}_1, \tilde{Q}_4 \geq 0
\]
(3.37)

\[
\tilde{Q}_1 + \tilde{Q}_4 \leq Q/2
\]

under Assumptions 2(ii)

Without Assumption 6, with Assumption 3 and 5, \(p^*_{kl}(x)\) is bounded sharply by
\[
\begin{align*}
\frac{r_{11}(x)}{\min\{p_1(x) + \frac{Q}{\text{Pr}(X = x)}, 1\}} & \leq p_{11}^*(x) & \leq \min\left\{ \frac{r_{11}(x) + \frac{Q}{\text{Pr}(X = x)}}{p_1(x)}, 1 \right\} \\
\max\left\{ \frac{r_{12}(x) - \frac{Q}{\text{Pr}(X = x)}}{p_1(x)}, 0 \right\} & \leq p_{12}^*(x) & \leq \min\left\{ \frac{r_{12}(x) + \frac{Q}{\text{Pr}(X = x)}}{p_1(x) + \frac{Q}{\text{Pr}(X = x)}}, 1 \right\} \\
\max\left\{ \frac{r_{21}(x) - \frac{Q}{\text{Pr}(X = x)}}{p_2(x) - \frac{Q}{\text{Pr}(X = x)}}, 0 \right\} & \leq p_{21}^*(x) & \leq \min\left\{ \frac{r_{21}(x) + \frac{Q\tilde{Q}_1(x)}{\max\{p_2(x) - \frac{Q\tilde{Q}_4(x)}{\text{Pr}(X = x)}\}, 0}}{1} \right\} \\
\max\left\{ \frac{r_{22}(x) - \frac{Q}{\text{Pr}(X = x)}}{p_2(x)}, 0 \right\} & \leq p_{22}^*(x) & \leq \max\left\{ \frac{r_{22}(x)}{\min\{p_2(x) - \frac{Q}{\text{Pr}(X = x)}\}, 0} \right\}
\end{align*}
\] (3.38) (3.39) (3.40) (3.41)

under Assumptions 2(i) or

\[
\begin{align*}
Q\tilde{Q}_1(x), Q\tilde{Q}_4(x) & \geq 0 \\
Q\tilde{Q}_1(x) + Q\tilde{Q}_4(x) & \leq \min\{Q/[K \cdot \text{Pr}(X = x)], 1\}
\end{align*}
\] (3.42)

under Assumptions 2(ii). Proof: See Appendix G.

Under Corollaries 6.1-6.3, estimation of the bounds for \(p_{kl}^*(x)\) are straightforward using the appropriate sample analogs and minimizing (maximizing) the lower (upper) bound subject to the appropriate constraints. Upon obtaining bounds for \(p_{kl}^*(x)\), sharp bounds for the conditional and unconditional transition probabilities are given in (3.26) and (3.27).\(^{11}\)

Before continuing, it is worth pointing out a special case of level set restrictions when the conditioning variable, \(x\), represents time. For example, one might separately bound transition matrices from \(t = 0 \rightarrow 1\) and \(t = 1 \rightarrow 2\) and then impose the restriction that mobility is constant across the two time periods. Here, the level set restriction is identical to a stationarity assumption about the Markov process governing the outcome variable. This is formalized in the following assumption and proposition.

\(^{11}\) Note, there is no assurance that the bounds under Assumption 5, but without Assumption 6, will be narrower than the corresponding bounds without Assumption 5.
Assumption 7 (Stationarity). The transition matrix is constant across two consecutive periods. Formally,

\[ P_{t,t+1}^* = P_{t+1,t+2}^*. \]

Proposition 7. Let \( p_{kl}(t, t + 1) \) represent the elements of \( P_{t,t+1}^* \). Denote the bounds for \( p_{kl}(t, t + 1) \) under some combination of Assumptions 2-6 as

\[ LB(t, t + 1) \leq p_{kl}^*(t, t + 1) \leq UB(t, t + 1). \] (3.43)

Define the elements and corresponding bounds similarly for \( P_{t+1,t+2}^* \). Adding Assumption 7 implies the following sharp bounds on the elements of \( P^* = P_{t,t+1}^* = P_{t+1,t+2}^* \):

\[ \max\{LB(t, t + 1), LB(t + 1, t + 2)\} \leq p_{kl}^* \leq \min\{UB(t, t + 1), UB(t + 1, t + 2)\}, \] (3.44)

where \( p_{kl}^* \) refers to the elements of \( P^* \). Proof: Follows directly from Proposition 6.

3.3.3.3. Monotonicity Assumptions

Monotonicity restrictions place inequality constraints on population transition probabilities across observations with different observed attributes (Manski and Pepper 2000).

Assumption 8 (Monotonicity). The conditional probability of upward mobility is weakly increasing in a vector of attributes, \( u \), and the conditional probability of downward mobility is weakly decreasing in the same vector of attributes. Formally, if \( u_2 \geq u_1 \), then

\[
\begin{align*}
  p_{kl}^*(u_1) & \leq p_{kl}^*(u_2) & \forall l > k \\
  p_{kl}^*(u_1) & \geq p_{kl}^*(u_2) & \forall l < k \\
  p_{11}^*(u_1) & \geq p_{11}^*(u_2) \\
  p_{KK}^*(u_1) & \leq p_{KK}^*(u_2).
\end{align*}
\]

For instance, if \( u \) denotes the education of an individual, one might wish to assume that the probability of upward (downward) mobility is no lower (higher) for individuals with more education. Note, the monotonicity assumption provides no information on the conditional staying probabilities, \( p_{kk}^*(u) \), for \( k = 2, ..., K - 1 \).
This leads to the following proposition.

**Proposition 8.** Denote the bounds for \( p_{kl}(u) \) under some combination of Assumptions 2-5 as

\[
LB(u) \leq p_{kl}^*(u) \leq UB(u).
\]

(3.45)

Adding Assumption 8 implies the following sharp bounds on the conditional transition probabilities

\[
\sup_{u_1 \leq u} LB(u_1) \leq p_{kl}^*(u) \leq \inf_{u_2 \leq u} UB(u_2) \quad \forall l > k
\]

(3.46)

\[
\sup_{u \leq u_1} LB(u) \leq p_{kl}^*(u) \leq \inf_{u_2 \leq u} UB(u_2) \quad \forall l < k
\]

(3.47)

\[
\sup_{u \leq u_1} LB(u) \leq p_{11}^*(u) \leq \inf_{u_2 \leq u} UB(u_2)
\]

(3.48)

\[
\sup_{u_1 \leq u} LB(u_1) \leq p_{KK}^*(u) \leq \inf_{u_2 \leq u} UB(u_2)
\]

(3.49)

Assuming \( U \) is discrete, sharp bounds on the unconditional transition probabilities are given as

\[
\sum_u \Pr(U = u) \left( \sup_{u_1 \leq u} LB(u_1) \right) \leq p_{kl}^* \leq \sum_u \Pr(U = u) \left( \inf_{u_2 \geq u} UB(u_1) \right).
\]

(3.50)

**Proof:** This is a simple extension of Manski and Pepper (2000, Proposition 1 and Corollary 1).

### 3.3.4. Summary Mobility Measures

Upon obtaining bounds on the elements of the transition matrix, bounds on various measures derived from these elements follow automatically. The Prais (1955) measure of mobility is based on the mean exit time from partition \( k \), given by

\[
\frac{1}{1 - p_{kk}^*}, \quad k = 1, \ldots, K.
\]

(3.51)

Shorrocks (1978) defines the Immobility Ratio measure as

\[
IR = \frac{K - \text{tr}(P_{0,1}^*)}{K - 1}
\]

(3.52)

where \( \text{tr}(\cdot) \) is the trace of a matrix. Finally, Bradbury (2016) defines measures of upward
and downward mobility that account for the size of the partitions. The upward mobility measure is given by

$$UM = \frac{K}{K-1} (1 - p^*_1);$$

(3.53)
downward mobility is given by

$$DM = \frac{K}{K-1} (1 - p^*_{K1}).$$

(3.54)

Mobility is decreasing in the value of the Prais measure; increasing in the remaining three measures.

The measures in in (3.51) and (3.53)-(3.54) can be sharply bounded in a straightforward manner using sharp bounds on the individual conditional staying probabilities since each measure depends on only one element from the transition matrix. This is not the case for $IR$ under the assumption of arbitrary misclassification errors (Assumption 2(i)). Under uniform misclassification errors (Assumption 2(ii)), however, there is no additional information one can use to tighten bounds on $IR$ as the measure depends on only one element from each row and column of the transition matrix. With arbitrary errors, there is additional information and the bounds can be tightened. Here, the bounds are given by

$$\frac{K - \sum_k UB_{kk}}{K-1} + Q \leq IR \leq \frac{K - \sum_k LB_{kk}}{K-1} - Q,$$

(3.55)

where $LB_{kk}$ ($UB_{kk}$) is the lower (upper) bounds for $p^*_{kk}$ under some set of assumptions. The proof is given in Appendix G.

3.3.5. Properties

3.3.5.1. Bias Correction

In most of the cases considered here, estimates of the bounds are obtained via plug-in estimators relying on infima and suprema. Such estimators are biased in finite samples, producing bounds that are too narrow (Kreider and Pepper 2008). To circumvent this issue, a bootstrap bias correction is typically used in the literature on partial identification.
Denote the plug-in estimators of the lower and upper bounds under some set of the preceding assumptions as $\hat{LB}$ and $\hat{UB}$, respectively. The bootstrap bias corrected estimates are given by

\[
\hat{LB}^{c} = 2\hat{LB} - E^{*}[\hat{LB}], \\
\hat{UB}^{c} = 2\hat{UB} - E^{*}[\hat{UB}],
\]

where $\hat{LB}^{c}$ and $\hat{UB}^{c}$ denote the bootstrap bias corrected estimates and $E^{*}[\cdot]$ denotes the expectation operator with respect to the bootstrap distribution. See Kreider and Pepper (2008) and the references therein. However, there is an added complication here. Because we are estimating bounds on probabilities, the upper (lower) bound is constrained by one (zero). It is well known that the traditional bootstrap does not work for parameters at or near the boundary of the parameter space (Andrews 2000). Instead, we employ subsampling, using replicate samples with $N/2$ observations (Andrews and Guggenberger 2009; Martínez-Muñoz and Suáreza 2010).\(^\text{12}\)

\subsection*{3.3.5.2. Inference}

A substantial body of literature exists on inference in partial identification models. Horowitz and Manski (2000) use nonparametric analysis to provide bounds for data with missing covariates and outcomes. They develop confidence intervals (CIs) that cover the entire identification region using the bootstrap. Chernozhukov, Hong and Tamer (2007) extended Horowitz and Manski (2000) to models where the criterion function is minimized on a identified set. Imbens and Manski (2004) propose CIs that asymptotically cover the “true” value of the parameter with a fixed probability rather than cover the entire region with such

\textsuperscript{12}We employ sub-sampling (without replacement) rather than an $m$-bootstrap (with replacement), where $m < N$, as sub-sampling is valid under weaker assumptions (Horowitz 2001). Noneless, our Stata code allows for both options. Moreover, we set $m = N/2$ as it is unlikely that an optimal, data-driven choice of $m$ is available (or computationally feasible in the present context). Politis et al. (1999, p. 61) state that “subsampling has some asymptotic validity across a broad range of choices for the subsample size” as long as $m/N \to 0$ and $m \to \infty$ as $N \to \infty$. Martínez-Muñoz and Suáreza (2010, p. 143) note that setting $m = N/2$ is “typical.”
a probability.

In this paper, inference is handled via subsampling and the Imbens-Manski (2004) correction to obtain 90% confidence intervals (CIs). As with the bias correction, we set the size of the replicate samples to $N/2$. Some comments on this choice is necessary as there has been much recent work on inference in partially identified models; Canay and Shaikh (2016) provide an excellent review. For instance, the random set theory, (conditional) moment inequality and intersection bounds are widely used to provide estimations and inference for partial identification models.

Moment Inequality Method works for models characterized by moment inequalities and equalities. Chernozhukov, Hong and Tamer (2007) conduct the finite-sample inference by obtaining asymptotic estimates of quantiles of the critical value using either a generic subsampling method or the asymptotic approximation to the distribution for it in the moment condition problems. Andrews and Shi (2013) use instrument functions to transfer conditional moment inequalities to unconditional moment inequalities. They use "Plug-in Asymptotic" (PA) critical values to construct confidence sets, they also show the confidence sets constructed by subsampling, but the former one is preferred. To apply the (conditional) moment inequality method, the economic model is required to be characterized by moment function. In our paper, the parameter of interest $Pr_{kl}^*$ is determined by a given value of $Q$ and the properties of probability, rather than a population criterion function. Therefore, it is more tractable to directly use subsampling bootstrap.

The random set variable approach is used for estimation and inference for partial identified population which has a compact and convex identification region that is equal to a transformation of the Aumann expectation. Beresteau and Molinari (2008) estimate set valued random variables (SVRVs) instead of having parameters of interest as real numbers or real vectors in point estimation. The inferential approach is based on the Hausdorff Distance between the population identification region and the estimator. The critical value is obtained either by subsampling bootstrap or a simulation method. In the random set theory, the key point is to have set valued random variable (SVRV), so that the parameter of interest is interval- or set-identified. However, in our paper, the sum of measurement errors ($Q$) is pre-determined and each conditional probability in the transition matrix is thus not mapped
to a random set because the bounds depend on the value of Q.

Chernozhukov, Lee and Rosen (2013) develop a partial identification method for intersection bounds estimation and inference. Intersection bounds are bounds defined by either the infimum or supremum of a parametric or non-parametric function. They develop theory for large sample inference based on the strong approximation of a sequence of series or kernel-based empirical processes by a sequence of penultimate Gaussian processes. The parameter of interest in the intersection bounds method is characterized by lower bound and an upper bound functions conditional on a vector of observed variables. The subsampling method is still preferred in our paper because the lower and upper bounds can be directly obtained under each assumption or restriction, it is not necessary to apply intersection bounds.

3.4. U.S. Mobility

3.4.1. Data

To assess U.S. intragenerational mobility, we use panel data from the Survey of Income and Program Participation (SIPP). Collected by the U.S. Census Bureau, SIPP is a rotating, nationally representative longitudinal survey of households. Begun in 1984, SIPP collects detailed income data as well as data on a host of other economic and demographic attributes. Households in the SIPP are surveyed over a multi-year period ranging from two and a half years to four years. Then, a new sample of households are drawn. The sample sizes range from approximately 14,000 to 52,000 households. Here, we use the 2004 and 2008 panels to examine mobility leading up to the Great Recession and during the early recovery period. For the 2004 panel, the initial period is November 2003 and the terminal period is October 2007. For the 2008 panel, the initial period is June 2008 and the terminal period is September 2012. Thus, we investigate household-level income dynamics over two separate four-year windows. We also assess mobility pooling the two panels.

For the analysis, the outcome variable is derived from total monthly household income (variable THTOTINC). This includes income from all household members and sources: labor market earnings, pensions, social security income, interest dividends, and other income sources. When analyzing the $2 \times 2$ poverty matrix, we determine poverty status for each
household in each period by comparing income with the SIPP-reported poverty threshold for the household (variable RHPOV). When analyzing general mobility, we estimate $3 \times 3$ matrices based on terciles of the income distribution in each period. However, to adjust for household composition, we construct three different measures of so-called equivalized household income.\textsuperscript{13} Adjusting income for household size when drawing welfare or policy conclusions is known to be crucial (e.g., Chiappori 2016). In our baseline analysis, we use OECD equivalized household income (OECD 1982).\textsuperscript{14} As alternatives, we also construct OECD-modified equivalized household income (Haagenars et al. 1994) and per capita household income.\textsuperscript{15} Specifically, the OECD (OECD-modified) equivalence scale assigns a value of one to the first household member, 0.7 (0.5) to each additional adult, and of 0.5 (0.3) to each child. In contrast, the per capita measure assigns a value of one to all household members. In the interest of brevity, results based on these alternative equivalence scales are relegated to the appendix.

When assessing the two panels separately and imposing level set restrictions, we use age of the household head in the initial period. Specifically, we group households into ten-year age bins (25-34, ..., 55-65) and impose the restriction that mobility is constant across adjacent bins. For example, we tighten the bounds on mobility for households where the head is, say, 30-34 by assuming that mobility is constant across households where the head is 25-29, 30-34, and 35-39. When pooling the two panels and imposing level set restrictions, we combine the age of household head restriction used in the case of separate panels with a stationarity assumption that mobility is constant across the two panels. For example, we tighten the bounds on mobility for households where the head is, say, 30-34 in the initial period of the 2004 panel by assuming that mobility is constant across households where the head is 25-29, 30-34, and 35-39 in the 2004 and 2008 panels.

\textsuperscript{13}There is no need to adjust income for household size when estimating the poverty transition matrix since the poverty threshold already accounts for differences in household composition.

\textsuperscript{14}OECD equivalized household income for an individual household is defined as $Y/N$, where $Y$ is total household income, $N = 1 + 0.7(A - 1) + 0.5C$, and $A$ ($C$) is the total number of adults (children) in the household.

\textsuperscript{15}OECD-modified equivalized household income for an individual household is defined as $Y/N$, where $Y$ is total household income, $N = 1 + 0.5(A - 1) + 0.3C$, and $A$ ($C$) is the total number of adults (children) in the household.
When imposing the monotonicity restrictions, we use the education of the household head in the initial period. Here, households are grouped into three bins (high school graduate and below, some college but less than a four-year degree, and at least a four-year college degree).

In constructing our estimation sample, we use only the initial and terminal wave for each panel. The sample, by necessity, must be balanced. Households with any invalid or missing information on the relevant variables are excluded. Finally, we restrict the sample to households where the head is between 25 and 65 years old in the initial period. The sample size for the 2004 panel is 7,834 and for the 2008 panel is 16,006. Summary statistics are presented in Table 1.

3.4.2. Results

3.4.2.1. Poverty Transition Matrix

Results for the $2 \times 2$ poverty transition matrix are presented in Tables 2-4. Overall, the observed poverty rate declined from 11.8% to 10.7% in the first panel (November 2003 to October 2007) and held constant at 12.6% in the second panel (June 2008 to September 2012); see Table 1. Turning to mobility, under the baseline assumption of Rank-Preserving Measurement Error (Table 2, Panel I) the probability of a household remaining in poverty across the initial and terminal periods in the first (second) SIPP panel is 0.448 (0.462), while the probability of remaining out of poverty is 0.939 (0.923). Thus, observed transitions out of (into) poverty are higher in the first (second) SIPP panel (transition out of poverty: 0.552 versus 0.538; transitions into poverty: 0.061 versus 0.077). This is not surprising since

---

16 The 2004 panel contains 10,503 households observed in the initial and terminal periods. Two observations are dropped due to negative household income. The remainder are dropped because the household head is outside the 25-65 year old age range. The 2008 panel contains 21,616 households observed in the initial and terminal periods. 88 observations are dropped due to negative or missing household income. The remainder are dropped because the household head is outside the 25-65 year old age range.

17 In all cases, we use 25 replicate samples for the subsampling bias correction and 100 replicate samples to construct 90% Imbens-Manski (2004) confidence intervals via subsampling using $m = N/2$ without replacement. For brevity, we do not report bounds based on all possible combinations of restrictions. Unreported results are available upon request.

18 Throughout the analysis, poverty status is measured only at the initial and terminal period. Thus, for example, “remaining in poverty” does not mean a household is necessarily continuously in poverty over the four-year period. For expositional purposes, however, we describe the results in terms of remaining in or out of poverty.
the second SIPP panel spans the end of the Great Recession and the early part of the recovery.

**Misclassification Assumptions**

Panels II and III in Table 2 allow for misclassification, but impose arbitrary (Assumption 2(i)) and uniform (Assumption 2(ii)) errors, respectively. The assumed maximum misclassification rate is 10% ($Q = 0.10$). The rationale for this choice is discussed in Supplemental Appendix G; we also explore sensitivity to this choice below. In Panel II the bounds are nearly uninformative on the mobility of households in poverty in the initial period in both SIPP panels. Thus, a relatively small amount of arbitrary misclassification results, in the absence of other information, in an inability to say anything about the four-year mobility rates of households initially in poverty. For households initially above the poverty line, the probability of remaining out of poverty four years later is at least 0.825 (0.808) in the first (second) SIPP panel.\(^\text{19}\) Moreover, in the second SIPP panel, we are able to rule out the possibility (at the 90% confidence level) that no households move into poverty over the four years spanned by the second SIPP panel.

In Panel III the bounds are more informative. Thus, the assumption of uniform errors has some identifying power. Under this assumption, the probability of escaping poverty is at least 0.130 (0.142) in the first (second) SIPP panel. The probability of remaining out of poverty is at least 0.882 (0.865) in the first (second) SIPP panel.

Panels IV and V in Table 2 add the assumption that misclassification is only in the upward direction (Assumption 3). This assumption has no identifying power on the transition probabilities for households above the poverty line in the initial period. However, it is useful in tightening the bounds on the transition probabilities for households in poverty in the initial period. With arbitrary and uni-directional misclassification (Assumptions 2(i) and 3), bounds on the probability of remaining in poverty four years later are [0.243, 1.000] in the first SIPP panel and [0.258, 1.000] in the second SIPP panel. Under uniform and uni-directional misclassification (Assumptions 2(ii) and 3), bounds on the probability of remaining in poverty four years later are further tightened to [0.315, 0.870] in the first SIPP panel.

\(^{19}\)Throughout the discussion of the results, we focus on the point estimates for simplicity. The confidence intervals are generally not much wider than the point estimates of the bounds.
panel and [0.331, 0.858] in the second SIPP panel. While the assumptions of uniformity and uni-directional misclassification certainly tighten the bounds, the width of the bounds under the assumption of 10% misclassification makes it clear than even relatively small amounts of misclassification adds considerable uncertainty to estimates of income mobility.

**Level Set Restrictions**

Table 3 allows for misclassification, but imposes different combinations of Assumptions 2–7. For the separate SIPP panels, level set restrictions are based on the age of the household head in the initial period. For the pooled panels, level set restrictions based on the age of the household head are imposed within each panel and stationarity (Assumption 7) is imposed across the panels. In Panel I, the level set restrictions are not combined with shape restrictions (Assumption 4). In Panel II, Assumption 4 is added to the level set restrictions. Assumption 4 corresponds to the restriction that households are more likely to maintain the same poverty status over the four-year period than change status. With each panel, we present results based on different types of misclassification errors based on Assumptions 2-3.

Several findings stand out. First, under arbitrary and independent misclassification errors (Assumptions 2(i) and 6), Panels IA and IIA reveal that the level set and shape restrictions have little identifying power. There is some tightening of the lower bounds relative to Panel II in Table 2, but it is modest. Second, under uniform and independent misclassification errors (Assumptions 2(ii) and 6), Panels IB and IIB reveal that the level set and shape restrictions have substantial identifying power. For example, bounds on the probability of remaining in poverty over the four-year period in the first SIPP panel under uniform errors alone are [0.026, 0.870] (Table 2, Panel III), under level set restrictions with independent errors are [0.100, 0.810] (Table 3, Panel IB), and under level set and shape restrictions with independent errors is [0.183, 0.810] (Table 3, Panel IIB). In addition, if we utilize the pooled panels and impose the stationarity assumption, the bounds are further tightened to [0.198, 0.823] (Table 3, Panel IIB). Under these assumptions, which seem plausible, at least 1 in 5 impoverished households in the initial period remain in poverty four years later. Third,

---

20For brevity, not all combinations are presented. Full results are available upon request.
adding the assumption of uni-directional misclassification errors has considerable identifying power on the transition probabilities for households below the poverty line in the initial period. Now the bounds on the probability of remaining in poverty over the four-year period in the first SIPP Panel are $[0.315, 0.870]$, implying that at least 3 in 10 impoverished households in the initial period remain in poverty four years later. Finally, adding the stationarity assumption modestly tightens the bounds further; bounds on the probability of remaining in poverty over the four-year period under uniform, independent, and uni-directional errors are $[0.326, 0.862]$ (Table 3, Panels IC and IIC).

**Monotonicity Restriction**

Table 4 is similar to Table 3, but adds Assumption 8. The monotonicity restriction requires upward mobility to be weakly increasing in the household head’s education level in the initial period. In general, the monotonicity assumption has little identifying power in this application as the bounds are only modestly tightened, if at all. For instance, assuming uniform, independent, and uni-directional misclassification and imposing shape, level set, and stationarity restrictions, adding the monotonicity restriction fails to tighten the bounds on the probability of remaining in poverty across the initial and terminal periods (see Panel IIC in Table 3 and 4).

### 3.4.2.2. Tercile Transition Matrix

Results for the $3 \times 3$ tercile transition matrix based on OECD equivalized household income are presented in Tables 5-7. These tables are analogous to Tables 2-4 except we no longer consider the assumption of uni-directional misclassification since now any upward misclassification must induce downward misclassification as well. Results based on alternative equivalence scales are reported in the appendix, Tables G.1-G.8.

Under the baseline assumption of Rank-Preserving Measurement Error (Table 5, Panel I) the conditional staying probabilities in the first (second) SIPP panel are 0.683, 0.533, and 0.692 (0.685, 0.538, and 0.685) for terciles 1, 2, and 3, respectively. Thus, the observed four-year conditional staying probabilities do not vary much across the two panels. Further-

---

21 For brevity, not all combinations are presented. Full results are available upon request.
more, we find that the probability of observing larger movements in the income distribution are less likely than smaller movements. For example, pooling the two panels together, the probability of moving from the first to second tercile is 0.245 and the first to third tercile is 0.071. Similarly, the probability of moving from the third to second tercile is 0.217 and the third to first tercile is 0.095.

**Misclassification Assumptions**

Panels II and III in Table 5 allow for misclassification, but impose Assumption 2(i) and 2(ii), respectively. The assumed maximum misclassification rate is at most 20% \((Q = 0.20)\). The rationale for this choice is discussed in Supplemental Appendix G; we also explore sensitivity to this choice below. Under arbitrary misclassification (Assumption 2(i)), the width of the bounds is one \((= 2KQ)\) unless the bounds hit one of the boundaries. Under uniform misclassification (Assumption 2(ii)), the width is 0.4 \((= 2Q)\) unless the bounds hit one of the boundaries. Thus, the bounds are guaranteed to be at least somewhat informative in both cases, but the assumption of uniform misclassification has significant identifying power. This assumption is reasonable if misclassification is equally likely in the upward and downward direction. With mean-reverting measurement error in income, this is plausible.

Focusing on the pooled results, as these differ very little from the individual panel results, we find that the bounds on the conditional staying probabilities are \([0.385, 0.985]\), \([0.238, 0.838]\), and \([0.388, 0.988]\) across terciles 1, 2, and 3 under arbitrary misclassification. The bounds tighten to \([0.585, 0.785]\), \([0.438, 0.638]\), and \([0.588, 0.788]\) under uniform misclassification. Bounds on the off-diagonal elements, while generally lower as one moves further from the diagonal, cannot rule out the possibility that large movements in the income distribution are more likely than smaller movements (conditional on changing terciles).

**Level Set Restrictions**

Table 6 allows for misclassification, but imposes different combinations of Assumptions 2–7.\(^{22}\) Because of the similarity of the results across the two SIPP panels in Table 5, we focus on the results for the pooled sample where the stationarity restriction (Assumption 7)

\(^{22}\)For brevity, not all combinations are presented. Full results are available upon request.
is imposed. In Panel I, the level set restrictions are not combined with shape restrictions (Assumption 4). In Panel II, shape restrictions are imposed on top of the level set restrictions. This assumption corresponds to the restriction that households are more likely to make smaller movements in the income distribution than larger movements.

Several findings stand out. First, under arbitrary and independent misclassification errors (Assumptions 2(i) and 6), Panels IA and IIA reveal that the level set restrictions have some identifying power. The shape restrictions do not add new information. As stated previously, the bounds under arbitrary errors in Table 5 have a width of 0.6 unless the boundary comes into play. After imposing the level set restrictions, the width of the bounds on the conditional staying probabilities falls to around 0.5. Thus, while still wide, there is some information in the level set restrictions. Second, under uniform and independent misclassification errors (Assumptions 2(ii) and 6), Panels IB and IIB reveal that the level set restrictions continue to have some identifying power. The shape restrictions continue to add no new information. The bounds under uniform errors in Table 5 have a width of 0.2 unless the boundary comes into play. After imposing the level set restrictions, the width of the bounds on the conditional staying probabilities falls to around 0.12. For example, bounds on the probability of remaining in the bottom tercile over the four-year period in the pooled sample under uniform errors alone are \([0.585, 0.785]\) (Table 5, Panel III), but under level set restrictions with independent errors are \([0.623, 0.732]\) (Table 6, Panel IB). Finally, under uniform and independent errors with the level set restrictions (including the stationarity assumption), the bounds on the probabilities of extreme income mobility – both upward and downward – exclude zero. However, under arbitrary and independent errors, the bounds include zero. Thus, we can rule out the possibility that there is no movement from the first to third or third to first tercile over the two four-year periods (at the 90% confidence level) only under the assumption of uniform and independent errors.

**Monotonicity Restriction**

Table 7 adds the monotonicity assumption. Two findings emerge. First, the monotonicity assumption has only modest identifying power under arbitrary, independent errors (Panels IA and IIA). For instance, the bounds on the probability of remaining in the bottom tercile...
across the initial and terminal periods in the pooled sample tighten from $[0.449, 0.907]$ to $[0.449, 0.888]$ (Panel IA in Table 6 and 7). Second, the monotonicity assumption has more identifying power, in relative terms, under uniform, independent errors (Panels IB and IIB). Here, the bounds on the probability of remaining in the bottom tercile across the initial and terminal periods in the pooled sample tighten from $[0.623, 0.732]$ to $[0.623, 0.706]$ (Panel IB in Table 6 and 7). The bounds on the probability of remaining in the top tercile across the initial and terminal periods in the pooled sample tighten from $[0.635, 0.764]$ to $[0.635, 0.719]$ (Panel IB in Table 6 and 7). In both cases, the bounds are fairly tight around the observed conditional staying probabilities of 0.685 and 0.688, respectively (Table 5, Panel I).

**Summary Mobility Measures**

Bounds on the summary mobility measures are reported in Table 8. Generally speaking, three conclusions can be drawn by this exercise. First, relative to the baseline assumption of Rank-Preserving Measurement Error, one can assess the dramatic increase in uncertainty once misclassification rates of up to 10% are allowed. For example, the 90% confidence interval for Shorrocks (1978) Immobility Ratio measure based on the first SIPP panel is $[0.529, 0.563]$ under rank-preserving measurement error. Under the assumption of arbitrary errors (with $Q = 0.10$), the confidence interval is $[0.084, 1.009]$. Second, our strictest set of assumptions – uniform, independent errors under level set, shape, and monotonicity restrictions – can tighten these bounds. Under these assumptions, the 90% confidence interval for Shorrocks (1978) Immobility Ratio measure based on the first SIPP panel is $[0.435, 0.659]$. Finally, the bounds differ very little across the two SIPP panels. Thus, allowing for misclassification, there is no evidence that mobility changed across the two panels.

3.4.2.3. Sensitivity to $Q$

To explore the sensitivity of the bounds to the choice of $Q$, we re-estimate the bounds varying $Q$ from 0 to 0.20 (0.40) in the case of poverty (tercile) transition matrices. For the sake of computational time, we focus on the point estimates of the bounds, not the confidence

---

23 For brevity, Table 8 displays only the 90% confidence intervals and not the point estimates of the bounds. In addition, only the results for the individual panels are provided. All results are available upon request.
3.5. Conclusion

That self-reported income contains complex, nonclassical measurement error is a well-established fact. That administrative data on income is imperfect is also relatively uncontroversial. As such, addressing measurement error in the study of income mobility should no longer be optional. To that end, several recent attempts to address measurement error have been put forth. Here, we offer a new and complementary approach based on the partial identification of transition matrices.

Among others, our approach has the advantage of transparency, as the assumptions used to tighten the bounds are easily understood and may be imposed in any combination depending on the particular context and the beliefs of the researcher. Moreover, our approach only requires data at two points in time. Finally, our approach extends easily to applications other than income. The primary drawback to our approach is the lack of point identification. Consequently, our approach should be viewed as a complement to existing approaches that produce point estimates under more stringent (or, at least, alternative) identifying assumptions. Using data from the SIPP, we show that relatively small amounts of measurement error leads to bounds that can be quite wide in the absence of other information or restrictions. However, the restrictions we consider contain significant identifying power. We are hopeful that future work will consider additional restrictions that may be used to further tighten the bounds on transition probabilities.
APPENDIX A

ROBUSTNESS CHECK ON DIFFERENT AGE GROUPS OF CHILDBEARING AGE WOMEN

Figure A.1. Synthetic Control Analysis Result for Women Aged 20-45

Figure A.2. Synthetic Control Analysis Result for Women Aged 25-35
Figure A.3. Synthetic Control Analysis Result for Women Aged 25-40

Figure A.4. Synthetic Control Analysis Result for Women Aged 25-50
## APPENDIX B

### FULL SET OF CONTROL VARIABLES

<table>
<thead>
<tr>
<th>Treated Synthetic</th>
<th>Treated Synthetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of children younger than 5</td>
<td>female lfpr(1991) 0.704 0.733</td>
</tr>
<tr>
<td>1 child</td>
<td>female lfpr(1992) 0.718 0.749</td>
</tr>
<tr>
<td>2 children</td>
<td>female lfpr(1993) 0.644 0.717</td>
</tr>
<tr>
<td>3 children</td>
<td>female lfpr(1994) 0.701 0.712</td>
</tr>
<tr>
<td>more than 4</td>
<td>female lfpr(1995) 0.717 0.736</td>
</tr>
<tr>
<td>female lfpr(1996)</td>
<td>female lfpr(1996) 0.725 0.738</td>
</tr>
<tr>
<td>female lfpr(1997)</td>
<td>female lfpr(1997) 0.759 0.749</td>
</tr>
<tr>
<td>female lfpr 45-65 years old(1985-1989)</td>
<td>female lfpr(1990-1994) 0.918 0.916</td>
</tr>
<tr>
<td>female lfpr 45-65 years old(1990-1994)</td>
<td>female lfpr(1995-1997) 0.908 0.908</td>
</tr>
<tr>
<td>female lfpr 45-65 years old(1995-1997)</td>
<td>female lfpr 45-65 years old(80-84) 0.772 0.769</td>
</tr>
<tr>
<td>above poverty line</td>
<td>female lfpr 45-65 years old(85-89) 0.753 0.736</td>
</tr>
<tr>
<td>food stamp</td>
<td>male lfpr 45-65 years old(90-90) 0.793 0.740</td>
</tr>
<tr>
<td>free lunch</td>
<td>male lfpr 45-65 years old(95-97) 0.800 0.748</td>
</tr>
<tr>
<td>age&lt;4</td>
<td>GDP 57309.33 94868.06</td>
</tr>
<tr>
<td>age4-18</td>
<td>white 0.853 0.823</td>
</tr>
<tr>
<td>age18-45</td>
<td>black 0.073 0.146</td>
</tr>
<tr>
<td>age&gt;45</td>
<td>other race 0.073 0.032</td>
</tr>
</tbody>
</table>

101
<table>
<thead>
<tr>
<th></th>
<th>Treated Synthetic</th>
<th>Treated Synthetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>average male education (1980-1991)</td>
<td>famsize1</td>
<td>0.133</td>
</tr>
<tr>
<td>less than high school</td>
<td>0.141</td>
<td>0.158</td>
</tr>
<tr>
<td>high school</td>
<td>0.344</td>
<td>0.356</td>
</tr>
<tr>
<td>college unfinished</td>
<td>0.228</td>
<td>0.211</td>
</tr>
<tr>
<td>college and above</td>
<td>0.254</td>
<td>0.232</td>
</tr>
<tr>
<td>average male education (1992-1997)</td>
<td>famsize-5 members</td>
<td>0.134</td>
</tr>
<tr>
<td>less than high school</td>
<td>0.115</td>
<td>0.155</td>
</tr>
<tr>
<td>high school</td>
<td>0.378</td>
<td>0.367</td>
</tr>
<tr>
<td>some college</td>
<td>0.210</td>
<td>0.201</td>
</tr>
<tr>
<td>2 years in college</td>
<td>0.055</td>
<td>0.063</td>
</tr>
<tr>
<td>3 years in college</td>
<td>0.176</td>
<td>0.154</td>
</tr>
<tr>
<td>college degree</td>
<td>0.066</td>
<td>0.061</td>
</tr>
<tr>
<td>average female education (1980-1991)</td>
<td>good health condition</td>
<td>0.368</td>
</tr>
<tr>
<td>less than high school</td>
<td>0.147</td>
<td>0.161</td>
</tr>
<tr>
<td>high school</td>
<td>0.416</td>
<td>0.423</td>
</tr>
<tr>
<td>college unfinished</td>
<td>0.218</td>
<td>0.208</td>
</tr>
<tr>
<td>college and above</td>
<td>0.181</td>
<td>0.178</td>
</tr>
<tr>
<td>less than high school</td>
<td>0.109</td>
<td>0.144</td>
</tr>
<tr>
<td>high school</td>
<td>0.392</td>
<td>0.362</td>
</tr>
<tr>
<td>some college</td>
<td>0.223</td>
<td>0.223</td>
</tr>
<tr>
<td>3 years in college</td>
<td>0.151</td>
<td>0.150</td>
</tr>
<tr>
<td>college degree</td>
<td>0.050</td>
<td>0.047</td>
</tr>
<tr>
<td>family total income</td>
<td>32673.94</td>
<td>33921.59</td>
</tr>
<tr>
<td>number of children</td>
<td>female lfpr(1983)</td>
<td>0.736</td>
</tr>
<tr>
<td>no children</td>
<td>0.309</td>
<td>0.357</td>
</tr>
<tr>
<td>1 child</td>
<td>0.211</td>
<td>0.214</td>
</tr>
<tr>
<td>2 children</td>
<td>0.211</td>
<td>0.214</td>
</tr>
<tr>
<td>3 children</td>
<td>0.300</td>
<td>0.260</td>
</tr>
<tr>
<td>4 children</td>
<td>0.130</td>
<td>0.115</td>
</tr>
<tr>
<td>5 children</td>
<td>0.036</td>
<td>0.038</td>
</tr>
<tr>
<td>more than 6</td>
<td>0.013</td>
<td>0.017</td>
</tr>
</tbody>
</table>

Sources: IPUMS-CPS.

Notes: lfpr stands for labor force participation rates.

Table B.1. Full Set of Explanatory Variables and Measure of Closeness
APPENDIX C

FERTILITY RATE OF WOMEN AGED 25-45

Figure C.1. Pre-k Enrollment Rate in Oklahoma

Figure C.2. Synthetic Control Analysis on Oklahoma Pre-k Enrollment Rate
Figure C.3. Fertility Rate of Oklahoma Women Aged 25-45

Figure C.4. Synthetic Control Analysis on Fertility Rate of Women Aged 25-45
APPENDIX D
EFFECTS OF OKLAHOMA UNIVERSAL PRE-K ON OTHER SAMPLES

Figure D.1. Synthetic Control Analysis on Full-time Labor Force Participation Rate of Mothers with Children Younger than 5 Years Old

Figure D.2. Synthetic Control Analysis on Employment Rate of Mothers with Children Younger than 5 Years Old
Figure D.3. Synthetic Control Analysis on Employment Rate of Mothers with 4-year-olds Only

Figure D.4. Synthetic Control Analysis on Full-time Labor Force Participation Rate (Georgia)

Figure D.5. Synthetic Control Analysis on Employment Rate (Georgia)
APPENDIX E

SUBSAMPLE RESULTS OF THE EFFECT OF OKLAHOMA UNIVERSAL PRE-K

Figure E.1. Synthetic Control Analysis on Labor Force Participation Rate (Below Poverty Line)

Figure E.2. Synthetic Control Analysis on Weekly Working Hours (Below Poverty Line)
Figure E.3. Synthetic Control Analysis on Labor Force Participation Rate (Above Poverty Line)

Figure E.4. Synthetic Control Analysis on Weekly Working Hours (Above Poverty Line)

Figure E.5. Synthetic Control Analysis on Labor Force Participation Rate (Married)
Figure E.6. Synthetic Control Analysis on Weekly Working Hours (Married)

Figure E.7. Synthetic Control Analysis on Labor Force Participation Rate (Unmarried)

Figure E.8. Synthetic Control Analysis on Weekly Working Hours (Unmarried)
Figure E.9. Synthetic Control Analysis on Labor Force Participation Rate (Differential Education Level—Lower than Highschool, High School, College and Above College by Order)

Figure E.10. Synthetic Control Analysis on Weekly Working Hours (Differential Education Level—Lower than Highschool, High School, College and Above College by Order)
Figure E.11. Synthetic Control Analysis on Labor Force Participation Rate (Fewer than 2 Children)

Figure E.12. Synthetic Control Analysis on Weekly Working Hours (Fewer than 2 Children)

Figure E.13. Synthetic Control Analysis on Labor Force Participation Rate (More than 2 Children)
Figure E.14. Synthetic Control Analysis on Weekly Working Hours (More than 2 Children)

Figure E.15. Synthetic control analysis on women (25-45) with less than high school education level

Figure E.16. Synthetic control analysis on women (25-45) with high school education level
Figure E.17. Synthetic control analysis on women (25-45) with college and above education level

Figure E.18. Synthetic control analysis on women (25-45) with family income above poverty line

Figure E.19. Synthetic control analysis on women (25-45) with family income below poverty line
Figure E.20. Synthetic control analysis on married mothers

Figure E.21. Synthetic control analysis on unmarried mothers
Literature Review

We identify three general approaches to handling measurement error in the study of mobility in the existing literature: (i) ignore it, (ii) *ad hoc* data approaches, and (iii) *structural* approaches. The first, and most common, approach is to note the problem and then ignore it. Pavlopoulos et al. (2012, p. 750) state that “despite the enormous bias that measurement error can cause in the estimation of wage dynamics, most relevant studies ignore this phenomenon.” Lee et al. (2017, p. 37) write that “most studies of income and poverty dynamics have ignored potential measurement error biases in the transition matrices, although the presence of measurement error in both income and expenditure survey data has been widely acknowledged.”

The second strategy we refer to as *ad hoc* data approaches. Trimming is one example and refers to the practice of deleting a fraction (say, 1%) of the poorest and richest observations in the sample. Jäntti and Jenkins (2015, p. 862) note that trimming has been “applied in virtually every study cited.” The motivation for trimming is the removal of outliers that may represent measurement error (e.g., Maasoumi and Trede 2001). The drawbacks to trimming include the fact that outliers may arise for reasons other than measurement error and that it does not address measurement error outside the tails of the distribution.

A second example of an *ad hoc* approach is to average income data over several years. Thus, when computing mobility between two points in time, income in the initial (terminal) period is taken as, say, the three-year average around the true initial (terminal) period. Such a strategy was popularized in Solon (1992); see Bhattacharya and Mazumder (2011) and Bradbury (2016) for more recent examples. The motivation for averaging income over several periods is to smooth away measurement error. However, there are several drawbacks to this procedure. First, averaging smooths away all time-varying idiosyncratic sources of income variation, regardless of whether the variation arises from measurement error or
legitimate shocks to income. Second, averaging will not remove measurement error that is persistent over time. Finally, averaging requires data from more than two time periods, a requirement that may be prohibitive.

A third example of an ad hoc approach is the pseudo-panel estimator in Antman and McKenzie (2007), although the approach can also be applied with genuine panel data. Here, rather than averaging income over several periods for each observation, income is averaged over individuals assigned to the same cohort within each time period. Measures of mobility are then computed using panel data at the cohort level. As in the preceding case, the motivation for averaging income within cohorts and time periods is to smooth away measurement error. Again, though, there are several drawbacks. First, averaging smooths away all time-varying idiosyncratic sources of income variation. Second, cohorts must remain stable over time, which is not assured when using pseudo-panel data, and cohorts must be large. Finally, the definition of cohorts is arbitrary and shrinks the effective sample size.

The final general strategy used in the extant literature we refer to as structural approaches. Approaches falling under this category represent the forefront of the literature and can be sub-divided into two groups. The first group seeks to estimate a scalar measure of mobility: either the correlation coefficient between (true) log incomes in the initial and terminal periods, denoted by $\rho$, or the elasticity of (true) terminal period income with respect to (true) initial period income, denoted by $\beta$ in the following linear regression model

$$\ln(y_{0i}^*) = \alpha + \beta \ln(y_{0i}^*) + \varepsilon_i,$$  \hspace{1cm} (F.1)

where $y_{0i}^*$ ($y_{1i}^*$) is income in the true initial (terminal) period for observation $i$. Glewwe (2012) notes that if we define $\beta_R$ as the coefficient in the reverse regression, given by

$$\ln(y_{0i}^*) = \alpha_R + \beta_R \ln(y_{1i}^*) + \eta_i,$$  \hspace{1cm} (F.2)

then

$$\text{plim} \sqrt{\hat{\beta} \hat{\beta}_R} = \rho,$$  \hspace{1cm} (F.3)
where $\beta_\beta$ and $\beta_R$ denote the ordinary least squares (OLS) estimate of the corresponding population parameters.

With measurement error, the researcher observes $y_{0i}$ and $y_{1i}$. As such, $\beta_\beta$ and $\beta_R$ are inconsistent and the square root of their product provides a consistent estimate of the correlation between the logs of *observed* income, not *true* income. The solution proffered in the literature is to recover consistent estimates of $\beta$, $\beta_R$, and $\rho$ via instrumental variables (IV). There are two drawbacks to this approach. First, obtaining credible instruments is extremely difficult (if not impossible). Antman and McKenzie (2007) and Glewwe (2012) offer detailed examinations of this issue. Second, the scalar nature of $\rho$ and $\beta$ precludes examination of mobility at different parts of the distribution.

The second group of structural approaches are discussed in the main text.

**Misclassification Probabilities**

We can write the elements of $P_{0,1}^*$ as

$$p_{kl}^* = \frac{\Pr(y^*_{0i} \in k, y^*_{1i} \in l)}{\Pr(y^*_{0i} \in k)}$$

$$= \frac{\Pr(y_0 \in k; y_1 \in l) + \sum_{k', l' = 1,2,...,K \neq k, k'} \theta^{(k' - k, l' - l)}_{(k,l)} - \sum_{k, l = 1,2,...,K} \theta^{(k - k, l - l)}_{(k,l)}}{\Pr(y_0 \in k) + \sum_{k', l', l'' = 1,2,...,K \neq k} \theta^{(k' - k, l' - l)}_{(k,l)} - \sum_{k, l = 1,2,...,K} \theta^{(k - k, l - l)}_{(k,l)}}$$

$$= \frac{r_{kl} + \sum_{k', l', l'' = 1,2,...,K \neq k, k'} \theta^{(k' - k, l' - l)}_{(k,l)} - \sum_{k, l = 1,2,...,K} \theta^{(k - k, l - l)}_{(k,l)}}{p_k + \sum_{k', l', l'' = 1,2,...,K \neq k} \theta^{(k' - k, l' - l)}_{(k,l)} - \sum_{k, l = 1,2,...,K} \theta^{(k - k, l - l)}_{(k,l)}}$$

$$= K \left[ \Pr(y_0 \in k; y_1 \in l) + \sum_{k', l', l'' = 1,2,...,K \neq k, k'} \theta^{(k' - k, l' - l)}_{(k,l)} - \sum_{k, l = 1,2,...,K} \theta^{(k - k, l - l)}_{(k,l)} \right]$$

117
where the final line holds only in the case of equal-sized partitions. The transition probabilities are not identified from the data alone. The data identify \( \Pr(y_0 \in k, y_1 \in l) \) and \( \Pr(y_0 \in k) \), but not the misclassification parameters, \( \theta \). In principal, one can compute sharp bounds by searching across the unknown misclassification parameters. There are \( K^2(K^2 - 1) \) misclassification parameters in \( P^*_0 \). However, the following constraints must hold.

\[
\begin{align*}
(i) \quad 0 & \leq \sum_{k,k',l,l'=1,2,...,K \atop k \neq k', l \neq l'} \theta_{(k,k,l-l')}^{(k-k,k',l-l')} \leq \Pr(y_0 \in k, y_1 \in l) \equiv r_{kl}, \quad k, l = 1, \ldots, K \\
(ii) \quad 0 & \leq \sum_{k,k',l,l'=1,2,...,K \atop k \neq k'} \theta_{(k,k,l-l')}^{(k-k,k',l-l')} \leq \Pr(y_0 \in k) \equiv p_k, \quad k = 1, \ldots, K \\
(iii) \quad 0 & \leq \sum_{k,k',l,l'=1,2,...,K \atop l \neq l'} \theta_{(k,k,l-l')}^{(k-k,k',l-l')} \leq \Pr(y_0 \in l) \equiv p_l, \quad l = 1, \ldots, K
\end{align*}
\]

The \( K^2 \) inequality constraints in (i) must hold since the fraction of observations incorrectly classified as belonging to partition \( (k, l) \) cannot exceed the fraction of observations classified as belonging to this partition. The \( K \) inequality constraints in (ii) and (iii) must hold since the fraction of observations incorrectly classified as belonging to partition \( k \) in period 0 or partition \( l \) in period 1 cannot exceed the fraction of observations classified as belonging to these partitions.

In addition, the following constraints must hold in the case of equal-sized partitions:

\[
\begin{align*}
(iv.a) \quad \sum_{k',l',l=1,2,...,K \atop k' \neq k} \theta_{(k',l',l)}^{(k'-k,k,l-l')} - \sum_{k,k',l',l=1,2,...,K \atop k \neq k, k' \neq k} \theta_{(k,k,l-l')}^{(k-k,k',l-l')} & = 0, \quad k = 1, \ldots, K \\
(v.a) \quad \sum_{k',l',l=1,2,...,K \atop l \neq l'} \theta_{(k,k,l-l')}^{(k-k,k',l-l')} - \sum_{k,k',l,l'=1,2,...,K \atop l \neq l, k \neq k'} \theta_{(k,k,l-l')}^{(k-k,k',l-l')} & = 0, \quad l = 1, \ldots, K \\
(vi.a) \quad -r_{kl} & \leq \sum_{k',l',l=1,2,...,K \atop (k'-k,l'-l) \neq (0,0)} \theta_{(k,k,l-l')}^{(k'-k,l'-l)} - \sum_{k,k',l,l'=1,2,...,K \atop (k,k,l-l') \neq (k,l)} \theta_{(k,k,l-l')}^{(k-k,k',l-l')} \leq \frac{1}{K} - r_{kl}, \quad k, l = 1, \ldots, K
\end{align*}
\]

The constraints in (iv.a) and (v.a) follow from the fact that \( \Pr(y_0 \in k) = \Pr(y_1 \in l) = 1/K \).

The constraints in (vi.a) follow from the fact that \( r_{kl}^* = \Pr(y_0^* \in k, y_1^* \in l) \in [0, 1/K] \).

If the partitions are of unequal size, then the following constraints must hold:
(iv.b) \[-p_k \leq \sum_{k',l',\tilde{u}=1,2,...,K} \theta^{(k'-k,l'-\tilde{u})}_{(k,\tilde{u})} - \sum_{k,k',l',\tilde{u}=1,2,...,K} \theta^{(k-k\tilde{u},l'-\tilde{u})}_{(k\tilde{u},\tilde{u})} \leq 1 - p_k, \quad k = 1, ..., K\]

(v.b) \[-p_l \leq \sum_{k',l',\tilde{u}=1,2,...,K} \theta^{(k'-k\tilde{u},l'-l)}_{(k\tilde{u},l)} - \sum_{k',l',\tilde{u}=1,2,...,K} \theta^{(k'-k\tilde{u},l-l)}_{(k\tilde{u},l)} \leq 1 - p_l, \quad l = 1, ..., K\]

(vi.b) \[-r_{kl} \leq \sum_{k',l'=1,2,...,K} \theta^{(k'-k,l'-l)}_{(k,l)} - \sum_{k,k',l',\tilde{u}=1,2,...,K} \theta^{(k-k\tilde{u},l'-l)}_{(k\tilde{u},l)} \leq 1 - r_{kl}, \quad k, l = 1, ..., K\]

The constraints in (iv.b) and (v.b) follow from the fact that \(p_k^* \equiv \Pr(y_0^* \in k) \in [0,1]\) and \(p_l^* \equiv \Pr(y_1^* \in l) \in [0,1]\). Finally, the constraints in (vi.b) follow from the fact that \(r_{kl}^* \equiv \Pr(y_0^* \in k, y_1^* \in l) \in [0,1]\).
Proofs for Propositions

Proof for Proposition 1

Assumption 1 implies \( Q_{1,kl} = Q_{2,kl} = Q_{3,k} = Q_{4,k} = 0, \forall k,l \). Therefore,

\[
p_{kl}^* = \frac{\Pr(y_0 \in k, y_1 \in l) + Q_{1,kl} - Q_{2,kl}}{\Pr(y_0 \in k) + Q_{3,k} - Q_{4,k}} = \frac{\Pr(y_0 \in k, y_1 \in l)}{\Pr(y_0 \in k)} = E \left[ \frac{I(y_0 \in k, y_1 \in l)}{I(y_0 \in k)} \right] = KE[I(y_0 \in k, y_1 \in l)]
\]

where the last line holds in the case of equal-sized partitions. The sample analog of \( p_{kl}^* \), denoted by \( \hat{p}_{kl} \), is thus given by

\[
\hat{p}_{kl} = \frac{\sum_i I(y_{0i} \in k, y_{1i} \in l)}{\sum_i I(y_{0i} \in k)} = \frac{\sum_i I(y_{0i} \in k, y_{1i} \in l)}{N/K} = \frac{K}{N} \sum_i I(y_{0i} \in k, y_{1i} \in l)
\]

where the last two line holds in the case of equal-sized partitions.

Proof for Proposition 3

Let it be known that \( Q \leq 1 \), then

\[
p_{kl}^* = [0, 1] \cap \left[ \frac{r_{kl} - Q_{2,kl}}{p_k + Q_{3,k}}, \frac{r_{kl} + Q_{1,kl}}{p_k - Q_{4,k}} \right]
\]

The component before intersection follows the definition of probability, such that \( p_{kl}^* \in [0, 1] \). The lower bound of in the second component in the right hand side is obtained by setting \( Q_{1,kl} = Q_{4,k} = 0 \), and the upper bound is obtained by setting \( Q_{2,kl} = Q_{3,k} = 0 \).
The bound can be written as

\[ \max \left\{ \frac{r_{kl} - Q_{2,kl}}{p_k + Q_{3,k}}, 0 \right\} \leq \hat{p}_{kl} \leq \min \left\{ \frac{r_{kl} + Q_{1,kl}}{p_k - Q_{4,k}}, 1 \right\} \]

When the lower bounds are obtained, \( Q_{1,kl} = Q_{4,k} = 0 \). This implies

\[ 0 \leq r_{kl} - Q_{2,kl} \leq 1, \text{ and } 0 \leq p_k + Q_{3,k} \leq 1 \]

Equivalently,

\[ 0 \leq Q_{2,kl} \leq r_{kl}, \text{ and } 0 \leq Q_{3,k} \leq 1 - p_k \]

When the upper bounds are obtained, \( Q_{2,kl} = Q_{3,k} = 0 \). This implies

\[ 0 \leq r_{kl} + Q_{1,kl} \leq 1, \text{ and } 0 \leq p_k - Q_{4,k} \leq 1 \]

Equivalently,

\[ 0 \leq Q_{1,kl} \leq 1 - r_{kl}, \text{ and } 0 \leq Q_{3,k} \leq p_k \]

Since setting \( Q_{1,kl} = Q_{4,k} = 0 \) will not increase \( Q_{2,kl} \) and \( Q_{3,k} \); and setting \( Q_{2,kl} = Q_{3,k} = 0 \) will not increase \( Q_{1,kl} \) and \( Q_{4,k} \). Under Assumption 2(i), we have

\[ Q_{1,kl} + Q_{4,k} \leq Q \]
\[ Q_{2,kl} + Q_{3,k} \leq Q \]

Under Assumption 2(ii), we have

\[ Q_{1,kl} + Q_{4,k} \leq Q/K \]
\[ Q_{2,kl} + Q_{3,k} \leq Q/K \]
Thus we can bound the transition probability by assigning a value to $Q$. It follows that if $r_{kl} = Q_{2,kl}$ the lower bound on the true transition probability is zero. And if $\frac{r_{kl} + Q_{1,kl}}{p_{kl} - Q_{1,kl}} > 1$, the upper bound is one. The bound becomes wider if $Q$ increases.

**Proof for Proposition 4**

In the $2 \times 2$ transition matrix case, following equation (5) we have,

$$p_{11}^* = \frac{r_{11} + Q_{1,11} - Q_{2,11}}{p_{1} + Q_{3,1} - Q_{4,1}}$$

$$= \frac{r_{11} + (\theta_{11}^{10} + \theta_{11}^{01} + \theta_{11}^{11}) - (\theta_{12}^{01} + \theta_{21}^{10} + \theta_{22}^{10})}{p_{1} + (\theta_{11}^{10} + \theta_{11}^{11} + \theta_{12}^{10} + \theta_{12}^{11}) - (\theta_{21}^{10} + \theta_{21}^{11} + \theta_{22}^{10} + \theta_{22}^{11})}$$

$$p_{12}^* = \frac{r_{12} + Q_{1,12} - Q_{2,12}}{p_{1} + Q_{3,1} - Q_{4,1}}$$

$$= \frac{r_{12} + (\theta_{12}^{10} + \theta_{12}^{01} + \theta_{12}^{11}) - (\theta_{11}^{01} + \theta_{21}^{10} + \theta_{22}^{10})}{p_{1} + (\theta_{11}^{10} + \theta_{11}^{11} + \theta_{12}^{10} + \theta_{12}^{11}) - (\theta_{21}^{10} + \theta_{21}^{11} + \theta_{22}^{10} + \theta_{22}^{11})}$$

$$p_{21}^* = \frac{r_{21} + Q_{1,21} - Q_{2,21}}{p_{2} + Q_{3,2} - Q_{4,2}}$$

$$= \frac{r_{21} + (\theta_{21}^{10} + \theta_{21}^{01} + \theta_{21}^{11}) - (\theta_{11}^{01} + \theta_{12}^{10} + \theta_{12}^{11})}{p_{2} + (\theta_{11}^{10} + \theta_{11}^{11} + \theta_{12}^{10} + \theta_{12}^{11}) - (\theta_{21}^{10} + \theta_{21}^{11} + \theta_{22}^{10} + \theta_{22}^{11})}$$

$$p_{22}^* = \frac{r_{22} + Q_{1,22} - Q_{2,22}}{p_{2} + Q_{3,2} - Q_{4,2}}$$

$$= \frac{r_{22} + (\theta_{22}^{10} + \theta_{22}^{01} + \theta_{22}^{11}) - (\theta_{11}^{10} + \theta_{12}^{10} + \theta_{12}^{01})}{p_{2} + (\theta_{11}^{10} + \theta_{11}^{11} + \theta_{12}^{10} + \theta_{12}^{11}) - (\theta_{21}^{10} + \theta_{21}^{11} + \theta_{22}^{10} + \theta_{22}^{11})}$$

Since all the $\theta$s with negative superscripts are the probabilities of misreporting to be in a lower partition, specifically in the $2 \times 2$ case, superscript -1 means in the time period, the person reports herself to be in Partition 1 while she is truely in Parition 1, under Assumption 3, are all the $\theta$s with negative superscripts are zero. Therefore, we have

$$p_{11}^* = \frac{r_{22} + (\theta_{11}^{10} + \theta_{11}^{01} + \theta_{11}^{11})}{p_{2} + (\theta_{11}^{10} + \theta_{11}^{11} + \theta_{12}^{10})}$$

$$p_{12}^* = \frac{r_{12} + \theta_{12}^{10} - \theta_{12}^{01}}{p_{1} + (\theta_{11}^{10} + \theta_{11}^{11} + \theta_{12}^{10})}$$

$$p_{21}^* = \frac{r_{21} + \theta_{21}^{01} - \theta_{21}^{10}}{p_{2} + (\theta_{11}^{10} + \theta_{11}^{11} + \theta_{12}^{10})}$$

$$p_{22}^* = \frac{r_{22} - (\theta_{11}^{10} + \theta_{12}^{10} + \theta_{12}^{01})}{p_{2} - (\theta_{11}^{10} + \theta_{11}^{11} + \theta_{12}^{10})}$$
1) For $P_{11}^*$, the lower bound is obtained when $\theta_{10}^{10} = \theta_{11}^{01} = \theta_{11}^{11} = 0$ and $\theta_{12}^{10} = Q$, or equivalently, $Q_{1,11} = 0, Q_{3,1} = Q$. And since $p_i^* \in [0, 1]$, we have $0 \leq p_1 + Q \leq 1$. Therefore,

$$p_{11}^* \geq \frac{r_{11}}{\min\{p_1 + Q, 1\}}$$

The upper bound is obtained when $\theta_{10}^{10} = \theta_{10}^{12} = \theta_{11}^{11} = 0$ and $\theta_{11}^{01} = Q$, or equivalently, $Q_{1,11} = Q, Q_{3,1} = 0$. And since $p_{11}^* \in [0, 1]$, we have

$$p_{11}^* \leq \min\{r_{11} + Q, 1\}$$

Therefore, the bound for $P_{11}^*$ is given by

$$\frac{r_{11}}{\min\{p_1 + Q, 1\}} \leq p_{11}^* \leq \min\{\frac{r_{11} + Q}{p_1}, 1\}$$

2) For $P_{12}^*$, the lower bound is obtained when $\theta_{12}^{10} = \theta_{11}^{10} = \theta_{11}^{11} = 0$ and $\theta_{11}^{01} = Q$, or equivalently, $Q_{1,12} = Q_{3,1} = 0, Q_{2,12} = Q$. And since $p_{12}^* \in [0, 1]$, we have

$$p_{12}^* \geq \max\{\frac{r_{12} - Q}{p_1}, 0\}$$

The upper bound is obtained when $\theta_{11}^{10} = \theta_{11}^{01} = \theta_{11}^{11} = 0$ and $\theta_{12}^{10} = Q$, or equivalently, $Q_{1,12} = Q_{3,1} = 0, Q_{2,12} = 0$. And since $p_{12}^* \in [0, 1]$, we have

$$p_{12}^* \leq \min\{\frac{r_{12} + Q}{p_1 + Q}, 1\}$$

Therefore, the bound for $P_{12}^*$ is given by

$$\max\{\frac{r_{12} - Q}{p_1}, 0\} \leq p_{12}^* \leq \min\{\frac{r_{12} + Q}{p_1 + Q}, 1\}$$

3) For $P_{21}^*$, the lower bound is obtained when $\theta_{21}^{10} = \theta_{12}^{10} = \theta_{11}^{11} = 0$ and $\theta_{11}^{01} = Q$, or equivalently, $Q_{1,21} = Q_{4,2} = Q, Q_{2,21} = 0$. And since $r_{21}^*, p_2^* \in [0, 1]$, we have

$$p_{21}^* \geq \max\{\frac{r_{21} - Q}{p_2 - Q}, 0\}$$
The upper bound is obtained when \( \theta_{11}^{10} = 0, \theta_{21}^{01} = Q_1 \) and \( \theta_{12}^{10} = \theta_{11}^{11} = Q_4 \), or equivalently, \( Q_{1,21} = Q_1, \ Q_{4,2} = Q_4 \) and \( Q_{2,21} = 0 \). And since \( p_2^* \in [0, 1] \), we have \( 0 \leq p_2 - Q \leq 1 \). Therefore,

\[
p_2^* \leq \min \left\{ \frac{r_{21} + Q_1}{\max\{p_2 - Q_4, 0\}}, 1 \right\}
\]

where

\[
Q_1, Q_4 \geq 0
\]

\[
Q_1 + Q_4 \leq Q
\]

Therefore, the bound for \( P_{12}^* \) is given by

\[
\max\left\{ \frac{r_{21} - Q}{p_2 - Q}, 0 \right\} \leq p_{21}^* \leq \min \left\{ \frac{r_{21} + Q_1}{\max\{p_2 - Q_4, 0\}}, 1 \right\}
\]

In addition to

\[
Q_1, Q_4 \geq 0
\]

\[
Q_1 + Q_4 \leq Q
\]

4) For \( P_{22}^* \), the lower bound is obtained when \( \theta_{12}^{10} = \theta_{11}^{10} = \theta_{21}^{11} = 0 \) and \( \theta_{21}^{01} = Q_4 \), or equivalently, \( Q_{4,2} = 0, \ Q_{2,22} = Q \). And since \( p_{22}^* \in [0, 1] \), we have

\[
p_{22}^* \geq \max\left\{ \frac{r_{22} - Q}{p_2}, 0 \right\}
\]

The upper bound is obtained when \( \theta_{11}^{11} = \theta_{12}^{10} = \theta_{21}^{01} = 0 \) and \( \theta_{11}^{10} = Q, \) or equivalently, \( Q_{4,2} = Q, \ Q_{2,22} = 0 \). And since \( p_2^* \in [0, 1] \), we have \( 0 \leq p_2 - Q \leq 1 \). Therefore,

\[
p_{22}^* \leq \min \left\{ \frac{r_{22}}{\max\{p_2 - Q, 0\}}, 1 \right\}
\]
Therefore, the bound for $P_{12}^*$ is given by

$$\max\{r_{22} - Q, 0\} \leq P_{22}^* \leq \min\{\frac{r_{22}}{\max\{p_2 - Q, 0\}}, 1\}$$

Altogether, we have

$$\frac{r_{11}}{\min\{p_1 + Q, 1\}} \leq p_{11}^* \leq \min\{\frac{r_{11} + Q}{p_1}, 1\}$$

$$\max\{r_{12} - Q, 0\} \leq p_{12}^* \leq \min\{\frac{r_{12} + Q}{p_1 + Q}, 1\}$$

$$\max\{r_{21} - Q, 0\} \leq p_{21}^* \leq \min\{\frac{r_{21} + Q_1}{\max\{p_2 - Q, 0\}}, 1\}$$

$$\max\{r_{22} - Q, 0\} \leq p_{22}^* \leq \min\{\frac{r_{22}}{\max\{p_2 - Q, 0\}}, 1\}$$

In addition to

$$Q_1, Q_4 \geq 0$$

$$Q_1 + Q_4 \leq Q$$

**Proof for Proposition 5**

To make the notations simple, denote the bounds on $p_{kl}^*$ under some combination of Assumption 2-3 as

$$LB_{kl} \leq p_{kl}^* \leq UB_{kl} \tag{F.4}$$

1) First, if $k=l$ then $|k - l| = 0$, since $p_{kl}^*$ is weakly decreasing in $|k - l|$, $p_{kl}^*$ is maximized at $k=l$. Formally,

$$p_{kl}^* \geq LB_{kl'} \quad \forall l' = 1, ..., K \iff p_{kl}^* \geq \sup_{l' = 1, ..., K} LB_{kl'}$$
and

\[
p^*_kl \geq LB_{k'l} \quad \forall k' = 1, ..., K, \quad \iff \quad p^*_kl \geq \sup_{k'=1,\ldots,K} LB_{k'l}
\]

We also know the condition (C.1), therefore we obtain equation (21)

\[
\max \left\{ \sup_{l' \geq l} LB_{kl'} , \sup_{k' \leq k} LB_{k'l} \right\} \leq p^*_kl \leq UB_{kl} \quad \text{if } k = l
\]

2) Second, if \( k < l \), Suppose i) there exists an \( l' \geq l \), so that \( k < l' \leq l \), then \( |k - l| \leq |k - l'| \), since \( p^*_kl \) is weakly decreasing in \( |k - l| \), we have

\[
p^*_kl \geq LB_{k'l} \quad \forall l \geq l' \quad \iff \quad p^*_kl \geq \sup_{l' \geq l} LB_{k'l}
\]

Suppose ii) there exists an \( l' \leq l \), so that \( k \leq l' \leq l \), then \( |k - l'| \leq |k - l| \), since \( p^*_kl \) is weakly decreasing in \( |k - l| \), we have

\[
p^*_kl \leq UB_{k'l} \quad \forall k \leq l' \leq l \quad \iff \quad p^*_kl \leq \inf_{k \leq l' \leq l} UB_{k'l}
\]

Suppose iii) there exists a \( k' \geq k \), so that \( k \leq k' \leq l \), then \( |k' - l| \leq |k - l| \), since \( p^*_kl \) is weakly decreasing in \( |k - l| \), we have

\[
p^*_kl \leq UB_{k'l} \quad \forall k \leq k' \leq l \quad \iff \quad p^*_kl \leq \inf_{k \leq k' \leq l} UB_{k'l}
\]

Suppose v) there exists a \( k' \leq k \), so that \( k' \leq k < l \), then \( |k' - l| \geq |k - l| \), since \( p^*_kl \) is weakly decreasing in \( |k - l| \), we have

\[
p^*_kl \geq LB_{k'l} \quad \forall k' \leq k \quad \iff \quad p^*_kl \geq \sup_{k' \leq k} LB_{k'l}
\]

We also know the condition (C.1), therefore we obtain equation (22)

\[
\max \left\{ \sup_{l' \geq l} LB_{kl'}, \sup_{k' \leq k} LB_{k'l} \right\} \leq p^*_kl \leq \min \left\{ \inf_{k \leq l' \leq l} UB_{kl'}, \inf_{k \leq k' \leq l} UB_{k'l} \right\} \quad \text{if } k < l
\]
3) Third, if \( k > l \), Suppose i) there exists a \( k' \geq k \), so that \( l < k \leq k' \), then \( |k - l| \leq |k' - l| \), since \( p_{kl}^* \) is weakly decreasing in \( |k - l| \), we have

\[
p_{kl}^* \geq LB_{k'} \quad \forall k' \geq k \iff p_{kl}^* \geq \sup_{k' \geq k} LB_{k'}
\]

Suppose ii) there exists a \( k' \leq k \), so that \( l \leq k' \leq k \), then \( |k' - l| \leq |k - l| \), since \( p_{kl}^* \) is weakly decreasing in \( |k - l| \), we have

\[
p_{kl}^* \leq UB_{k'} \quad \forall k' \leq k \iff p_{kl}^* \leq \inf_{l \leq k' \leq k} UB_{k'}
\]

Suppose iii) there exists an \( l' \geq l \), so that \( l \leq l' \leq k \), then \( |k - l| \leq |k - l'| \), since \( p_{kl}^* \) is weakly decreasing in \( |k - l| \), we have

\[
p_{kl}^* \leq UB_{l'} \quad \forall l \leq l' \leq k \iff p_{kl}^* \leq \inf_{l \leq l' \leq k} UB_{l'}
\]

Suppose v) there exists a \( l' \leq l \), so that \( l' \leq l < k \), then \( |k - l'| \leq |k - l| \), since \( p_{kl}^* \) is weakly decreasing in \( |k - l| \), we have

\[
p_{kl}^* \geq LB_{l'} \quad \forall l' \leq l \iff p_{kl}^* \geq \sup_{l' \leq l} LB_{l'}
\]

We also know the condition (C.1), therefore we obtain equation (23)

\[
\max \left\{ \sup_{l' \leq l} LB_{k'}, \sup_{k' \geq k} LB_{k'} \right\} \leq p_{kl}^* \leq \min \left\{ \inf_{l \leq l' \leq k} UB_{k'}, \inf_{l \leq l' \leq k} UB_{k'} \right\} \text{ if } k > l
\]

Proposition 5 is proved.
Proof for Corollary 6.1

From Equation (24) and Assumption 6, we have

\[ p^*_{kl} = \frac{r_{kl}(x) + Q_{1,kl}(x) - Q_{2,kl}(x)}{p_k(x) + Q_{3,k}(x) - Q_{4,k}(x)} \]

\[ = \frac{r_{kl}(x) + Q_{1,kl} - Q_{2,kl}}{p_k(x) + Q_{3,k} - Q_{4,k}} \]

In the case of equal-sized partitions, we have \( p^*_k = p_k = \frac{1}{K} \), so that \( Q_{3,k} = Q_{4,k} \). Therefore, we have

\[ = \frac{r_{kl}(x) + Q_{1,kl} - Q_{2,kl}}{p_k(x)} \]

Under Assumption 2(i), which implies \( Q_{1,kl} + Q_{2,kl} + Q_{3,k} + Q_{4,k} \leq Q \), the bound is given by

\[ \max \left\{ \frac{1}{p_k(x)}(r_{kl}(x) - Q), 0 \right\} \leq p^*_{kl}(x) \leq \min \left\{ \frac{1}{p_k(x)}(r_{kl}(x) + Q), 1 \right\}. \quad (F.5) \]

where the lower bound is obtained by setting \( Q_{1,kl} = 0 \) and \( Q_{2,kl} = Q \), and the upper bound is given by setting \( Q_{1,kl} = Q \) and \( Q_{2,kl} = 0 \).

**NOTE:** Professor, please check the lower bounds, if the nonzero item of lower the LB is possible to be (always) positive.

Under Assumption 2(ii), which implies \( Q_{1,kl} + Q_{2,kl} + Q_{3,k} + Q_{4,k} \leq Q/K \), the bound is given by

\[ \max \left\{ \frac{1}{p_k(x)}(r_{kl}(x) - Q/K), 0 \right\} \leq p^*_{kl}(x) \leq \min \left\{ \frac{1}{p_k(x)}(r_{kl}(x) + Q/K), 1 \right\}. \quad (F.6) \]

where the lower bound is obtained by setting \( Q_{1,kl} = 0 \) and \( Q_{2,kl} = Q \), and the upper bound is given by setting \( Q_{1,kl} = Q \) and \( Q_{2,kl} = 0 \).

Now without Assumption 6, given Equation (24) and following the Proof for Proposition 3, the bound of \( \Pr_{kl}(x) \) is given by

\[ \max \left\{ \frac{r_{kl}(x) - Q_{2,kl}(x)}{p_k(x) + Q_{3,k}(x)}, 0 \right\} \leq p^*_{kl}(x) \leq \min \left\{ \frac{r_{kl}(x) + Q_{1,kl}(x)}{p_k(x) - Q_{4,k}(x)}, 1 \right\} \]

128
where the lower bounds are obtained by setting \( Q_{1,k}(x) = Q_{4,k}(x) = 0 \) and the upper bounds are obtained by setting \( Q_{2,k}(x) = Q_{3,k}(x) = 0 \).

We also have,

\[
0 \leq r_{kl}(x) - Q_{2,kl}(x) \leq 1 \\
0 \leq r_{kl}(x) + Q_{1,kl}(x) \leq 1 \\
0 \leq p_{k}(x) + Q_{3,k}(x) \leq 1 \\
0 \leq p_{k}(x) - Q_{4,k}(x) \leq 1
\]

Or,

\[
0 \leq Q_{2,kl}(x) \leq r_{kl}(x) \quad (F.7) \\
0 \leq Q_{1,kl}(x) \leq 1 - r_{kl}(x) \quad (F.8) \\
0 \leq Q_{3,k}(x) \leq 1 - p_{k}(x) \quad (F.9) \\
0 \leq Q_{4,k}(x) \leq p_{k}(x) \quad (F.10)
\]

The law of total probability implies

\[
Pr(y_0 \in k)^* = Pr(y^*_0 \in k|X = x)Pr(X = x) + Pr(y^*_0 \in k|X = \bar{x})Pr(X = \bar{x}) \\
= Pr(y^*_0 \in k|X = x)Pr(X = x) + Pr(y^*_0 \in k|X = \bar{x})[1 - Pr(X = x)]
\]

In the case of equal-sized partitions, we know \( Pr(y_0 \in k)^* = \frac{1}{K} \), therefore we have

\[
\frac{1}{K} = p^*_k(x)Pr(X = x) + p^*_k(\bar{x})[1 - Pr(X = x)] \\
\iff p^*_k(\bar{x}) = \frac{\frac{1}{K} - Pr(X = x)p^*_k(x)}{1 - Pr(X = x)}
\]

129
And since $p_k^*(\bar{x}) \in [0, 1]$, we have

\[
0 \leq \frac{\frac{1}{K} - Pr(X = x)p_k^*(x)}{1 - Pr(X = x)} \leq 1
\]

\[\iff \frac{\frac{1}{K} - [1 - Pr(X = x)]}{Pr(X = x)} - p_k(x) \leq Q_{3,k}(x) \leq \frac{1}{KPr(X = x)} - p_k(x)\] (F.11)

Now we derive the bounds using the information above.

1) When the lower bounds are obtained, $Q_{1,k_{1}}(x) = Q_{4,k}(x) = 0$, we have $p_k^*(x) = p_k(x) + Q_{3,k}(x)$, or $Q_{3,k}(x) = p_k^*(x) - p_k(x)$, applying (C.6), we have

\[\frac{\frac{1}{K} - [1 - Pr(X = x)]}{Pr(X = x)} - p_k(x) \leq Q_{3,k}(x) \leq \frac{1}{KPr(X = x)} - p_k(x)\] (F.12)

From (C.3) and (C.7), we obtain the following bound

\[
\max\left\{\frac{\frac{1}{K} - [1 - Pr(X = x)]}{Pr(X = x)} - p_k(x), 0\right\} \leq Q_{3,k}(x) \leq \min\left\{1 - p_k(x), \frac{1}{KPr(X = x)} - p_k(x)\right\}
\] (F.13)

2) When the upper bounds are obtained, $Q_{2,k_{1}}(x) = Q_{3,k}(x) = 0$, we have $p_k^*(x) = p_k(x) - Q_{4,k}(x)$, or $Q_{4,k}(x) = p_k(x) - p_k^*(x)$, applying (C.6), we have

\[p_k - \frac{1}{kPr(X = x)} \leq Q_{4,k}(x) \leq p_k(x) - \frac{\frac{1}{K} - [1 - Pr(X = x)]}{Pr(X = x)}\] (F.14)

From (C.4) and (C.9), we obtain the following bound

\[
\max\left\{p_k - \frac{1}{kPr(X = x)}, 0\right\} \leq Q_{4,k}(x) \leq \min\{p_k(x), p_k(x) - \frac{\frac{1}{K} - [1 - Pr(X = x)]}{Pr(X = x)}\}\] (F.15)
Also, Assumption 2(i) implies $Q_{1,kl} + Q_{2,kl} + Q_{3,k} + Q_{4,k} \leq Q$, and Assumption 2(ii) implies $Q_{1,kl} + Q_{2,kl} + Q_{3,k} + Q_{4,k} \leq Q/K$. We know that under Assumption 2(i),

$$Q_{1,kl} + Q_{2,kl} + Q_{3,k} + Q_{4,k} = Q_{1,kl}(x)Pr(X = x) + Q_{2,kl}(x)Pr(X = x) + Q_{3,k}(x)Pr(X = x) + Q_{4,k}(x)Pr(X = x) \leq Q$$

and under Assumption 2(ii),

$$Q_{1,kl} + Q_{2,kl} + Q_{3,k} + Q_{4,k} = Q_{1,kl}(x)Pr(X = x) + Q_{2,kl}(x)Pr(X = x) + Q_{3,k}(x)Pr(X = x) + Q_{4,k}(x)Pr(X = x) \leq Q/K$$

Therefore, we have

$$Q_{1,kl}(x) + Q_{4,k}(x) \leq \min\{Q/Pr(X = x), 1\}$$

$$Q_{2,kl}(x) + Q_{3,k}(x) \leq \min\{Q/Pr(X = x), 1\}$$

under Assumption 2(i) or

$$Q_{1,kl}(x) + Q_{4,k}(x) \leq \min\{Q/[K \cdot Pr(X = x)], 1\}$$

$$Q_{2,kl}(x) + Q_{3,k}(x) \leq \min\{Q/[K \cdot Pr(X = x)], 1\}$$

under Assumption 2(ii). Corollary 6.1 is proved.
Proof for Corollary 6.2

Given Equation (24) and following the Proof for Proposition 3, with unequal-sized partitions, the bound of $\Pr_{kl}^*(x)$ is given by

$$\max \{ \frac{r_{kl}(x) - Q_{2,kl}(x)}{p_k(x) + Q_{3,k}(x)}, 0 \} \leq p_{kl}^*(x) \leq \min \{ \frac{r_{kl}(x) + Q_{1,kl}(x)}{p_k(x) - Q_{4,k}(x)}, 1 \}$$

where the lower bounds are obtained by setting $Q_{1,kl}(x) = Q_{4,k}(x) = 0$ and the upper bounds are obtained by setting $Q_{2,kl}(x) = Q_{3,k}(x) = 0$.

We also have,

$$0 \leq r_{kl}(x) - Q_{2,kl}(x) \leq 1$$
$$0 \leq r_{kl}(x) + Q_{1,kl}(x) \leq 1$$
$$0 \leq p_k(x) + Q_{3,k}(x) \leq 1$$
$$0 \leq p_k(x) - Q_{4,k}(x) \leq 1$$

Or,

$$0 \leq Q_{2,kl}(x) \leq r_{kl}(x)$$
$$0 \leq Q_{1,kl}(x) \leq 1 - r_{kl}(x)$$
$$0 \leq Q_{3,k}(x) \leq 1 - p_k(x)$$
$$0 \leq Q_{4,k}(x) \leq p_k(x)$$

Under Assumption 6, we have

$$Q_{1,kl}(x) = Q_{1,kl}, Q_{2,kl}(x) = Q_{2,kl}, Q_{3,k}(x) = Q_{3,k}, and Q_{4,k}(x) = Q_{4,k}$$

the bound in (C.13) is then given by

$$\max \{ \frac{r_{kl}(x) - Q_{2,kl}}{p_k(x) + Q_{3,k}}, 0 \} \leq p_{kl}^*(x) \leq \min \{ \frac{r_{kl}(x) + Q_{1,kl}}{p_k(x) - Q_{4,k}}, 1 \}$$
where the lower bounds are obtained by setting $Q_{1,kl} = Q_{4,k} = 0$ and the upper bounds are obtained by setting $Q_{2,kl} = Q_{3,k} = 0$.

Applying (C.18) to (C.14)-(C.17), we have

\[
0 \leq Q_{2,kl} \leq r_{kl}(x)
\]
\[
0 \leq Q_{1,kl} \leq 1 - r_{kl}(x)
\]
\[
0 \leq Q_{3,k} \leq 1 - p_k(x)
\]
\[
0 \leq Q_{4,k} \leq p_k(x)
\]

Following Proposition 3, we also have

\[
Q_{1,kl} + Q_{4,k} \leq Q
\]
\[
Q_{2,kl} + Q_{3,k} \leq Q
\]

under Assumption 2(i), and

\[
Q_{1,kl} + Q_{4,k} \leq Q/K
\]
\[
Q_{2,kl} + Q_{3,k} \leq Q/K
\]

under Assumption 2(ii).

Under Assumption 2(i) or 2(ii) without 6, (C.13)-(C.17) hold, following Corollary 6.1, with unequal-sized partition, we have

\[
Q_{1,kl}(x) + Q_{4,k}(x) \leq \min\{Q/\Pr(X = x), 1\}
\]
\[
Q_{2,kl}(x) + Q_{3,k}(x) \leq \min\{Q/\Pr(X = x), 1\}
\]

under Assumption 2(i) or
\[ Q_{1,k}(x) + Q_{4,k}(x) \leq \min\{Q/[K \cdot \Pr(X = x)], 1\} \]
\[ Q_{2,k}(x) + Q_{3,k}(x) \leq \min\{Q/[K \cdot \Pr(X = x)], 1\} \]

under Assumption 2(ii).

Corollary 6.2 is proved.

**Proof for Corollary 6.3**

Following the Proof of Proposition 4, Equation (24) and Assumption 3, in the 2 x 2 transition matrix case, we have

\[
\begin{align*}
p_{11}^* &= \frac{r_{22}(x) + (\theta_{11}^{10}(x) + \theta_{11}^{01}(x) + \theta_{11}^{11}(x))}{p_2(x) + (\theta_{11}^{10}(x) + \theta_{11}^{01}(x) + \theta_{11}^{12}(x))} \\
p_{12}^* &= \frac{r_{12}(x) + \theta_{12}^{10}(x) - \theta_{11}^{01}(x)}{p_1(x) + (\theta_{11}^{10}(x) + \theta_{11}^{01}(x) + \theta_{11}^{12}(x))} \\
p_{21}^* &= \frac{r_{21}(x) + \theta_{21}^{01}(x) - \theta_{11}^{01}(x)}{p_2(x) - (\theta_{11}^{10}(x) + \theta_{11}^{11}(x) + \theta_{11}^{12}(x))} \\
p_{22}^* &= \frac{r_{22}(x) - (\theta_{11}^{11}(x) + \theta_{11}^{01}(x) + \theta_{11}^{12}(x))}{p_2(x) - (\theta_{11}^{10}(x) + \theta_{11}^{11}(x) + \theta_{11}^{12}(x))}
\end{align*}
\]

Under Assumption 6, \( \theta_{k,l}^{k',l'}(x) = \theta \quad \forall k, l, k', l' \), then we have

\[
\begin{align*}
p_{11}^* &= \frac{r_{22}(x) + (\theta_{11}^{10} + \theta_{11}^{01} + \theta_{11}^{11})}{p_2(x) + (\theta_{11}^{10} + \theta_{11}^{01} + \theta_{11}^{12})} \\
p_{12}^* &= \frac{r_{12}(x) + \theta_{12}^{10} - \theta_{11}^{01}}{p_1(x) + (\theta_{11}^{10} + \theta_{11}^{01} + \theta_{11}^{12})} \\
p_{21}^* &= \frac{r_{21}(x) + \theta_{21}^{01} - \theta_{11}^{01}}{p_2(x) - (\theta_{11}^{10} + \theta_{11}^{11} + \theta_{11}^{12})} \\
p_{22}^* &= \frac{r_{22}(x) - (\theta_{11}^{11} + \theta_{11}^{01} + \theta_{11}^{12})}{p_2(x) - (\theta_{11}^{10} + \theta_{11}^{11} + \theta_{11}^{12})}
\end{align*}
\]
Following the same argument in the Proof for Proposition 4, we have

\[
\frac{r_{11}(x)}{\min\{p_1(x) + Q, 1\}} \leq p_{11}^*(x) \leq \min\left\{\frac{r_{11}(x) + Q}{p_1(x)}, 1\right\}
\]

\[
\max\left\{\frac{r_{12}(x) - Q}{p_1(x)}, 0\right\} \leq p_{12}^*(x) \leq \min\left\{\frac{r_{12}(x) + Q}{p_1(x) + Q}, 1\right\}
\]

\[
\max\left\{\frac{r_{21}(x) - Q}{p_2(x) - Q}, 0\right\} \leq p_{21}^*(x) \leq \min\left\{\frac{r_{21}(x) + Q}{\max\{p_2(x) - Q, 0\}}, 1\right\}
\]

\[
\max\left\{\frac{r_{22}(x) - Q}{p_2(x)}, 0\right\} \leq p_{22}^*(x) \leq \min\left\{\frac{r_{22}(x)}{\max\{p_2(x) - Q, 0\}}, 1\right\}
\]

where the following constraints must hold

\[
\tilde{Q}_1, \tilde{Q}_4 \geq 0
\]
\[
\tilde{Q}_1 + \tilde{Q}_4 \leq Q
\]

under Assumption 2(i) or

\[
\tilde{Q}_1, \tilde{Q}_4 \geq 0
\]
\[
\tilde{Q}_1 + \tilde{Q}_4 \leq Q/2
\]

Now suppose without Assumption 6, For \(P_{11}^*(x)\), the lower bound is obtained when \(\theta_{11}^{10}(x) = \theta_{11}^{11}(x) = 0\) and \(\theta_{12}^{10}(x) = Q = \frac{Q}{\Pr(X = x)}\), or equivalently, \(Q_{1,11}(x) = 0\), \(Q_{3,1}(x) = Q(x) = \frac{Q}{\Pr(X = x)}\). And since \(p_1^* \in [0, 1]\), we have \(0 \leq p_1(x) + Q(x) \leq \frac{1}{\Pr(X = x)}\). Therefore,

\[
p_{11}^*(x) \geq \frac{r_{11}(x)}{\min\{p_1(x) + Q/\Pr(X = x), 1\}}
\]

The upper bound is obtained when \(\theta_{11}^{10}(x) = \theta_{12}^{10}(x) = \theta_{11}^{11}(x) = 0\) and \(\theta_{11}^{11}(x) = Q(x) = \frac{Q}{\Pr(X = x)}\), or equivalently, \(Q_{1,11}(x) = Q = \frac{Q}{\Pr(X = x)}\), \(Q_{3,1}(x) = 0\). And since \(p_{11}^* \in [0, 1]\), we have

\[
p_{11}^*(x) \leq \min\left\{\frac{r_{11}(x) + Q/\Pr(X = x)}{p_1(x)}, 1\right\}
\]
Therefore, the bound for $P_{11}^*(x)$ is given by

$$\frac{r_{11}(x)}{\min\{p_1(x) + Q/Pr\{X = x\}, 1\}} \leq \min\left\{\frac{r_{11}(x) + Q/Pr\{X = x\}}{p_1(x)}, 1\right\}$$

Similarly, for $P_{12}^*(x)$, $P_{21}^*(x)$ and $P_{22}^*(x)$, we have

$$\max\left\{\frac{r_{12}(x) - Q/Pr\{X = x\}}{p_1(x)}, 0\right\} \leq p_{12}^*(x) \leq \min\left\{\frac{r_{12}(x) + Q/Pr\{X = x\}}{p_1(x) + Q/Pr\{X = x\}}, 1\right\}$$

$$\max\left\{\frac{r_{21}(x) - Q/Pr\{X = x\}}{p_2(x) - Q/Pr\{X = x\}}, 0\right\} \leq p_{21}^*(x) \leq \min\left\{\frac{r_{21}(x) + Q_1(x)}{\max\{p_2(x) - Q_4(x), 0\}}, 1\right\}$$

$$\max\left\{\frac{r_{22}(x) - Q/Pr\{X = x\}}{p_2(x)}, 0\right\} \leq p_{22}^*(x) \leq \min\left\{\frac{r_{22}(x)}{\max\{p_2(x) - Q/Pr\{X = x\}, 0\}}, 1\right\}$$

where the following constraints must hold

$$\tilde{Q}_1(x), \tilde{Q}_4(x) \geq 0$$

$$\tilde{Q}_1(x) + \tilde{Q}_4(x) \leq Q/Pr\{X = x\}$$

under Assumption 2(i) or

$$\tilde{Q}_1(x), \tilde{Q}_4(x) \geq 0$$

$$\tilde{Q}_1(x) + \tilde{Q}_4(x) \leq Q/2Pr\{X = x\}$$

under Assumption 2(ii).

Altogether, we have

$$\frac{r_{11}(x)}{\min\{p_1(x) + Q/Pr\{X = x\}, 1\}} \leq \min\left\{\frac{r_{11}(x) + Q/Pr\{X = x\}}{p_1(x)}, 1\right\}$$

$$\max\left\{\frac{r_{12}(x) - Q/Pr\{X = x\}}{p_1(x)}, 0\right\} \leq p_{12}^*(x) \leq \min\left\{\frac{r_{12}(x) + Q/Pr\{X = x\}}{p_1(x) + Q/Pr\{X = x\}}, 1\right\}$$

$$\max\left\{\frac{r_{21}(x) - Q/Pr\{X = x\}}{p_2(x) - Q/Pr\{X = x\}}, 0\right\} \leq p_{21}^*(x) \leq \min\left\{\frac{r_{21}(x) + Q_1(x)}{\max\{p_2(x) - Q_4(x), 0\}}, 1\right\}$$

$$\max\left\{\frac{r_{22}(x) - Q/Pr\{X = x\}}{p_2(x)}, 0\right\} \leq p_{22}^*(x) \leq \min\left\{\frac{r_{22}(x)}{\max\{p_2(x) - Q/Pr\{X = x\}, 0\}}, 1\right\}$$

136
where the following constraints must hold

\[ \tilde{Q}_1(x), \tilde{Q}_4(x) \geq 0 \]
\[ \tilde{Q}_1(x) + \tilde{Q}_4(x) \leq \max\{Q/Pr\{X = x\}, 1\} \]

under Assumption 2(i) or

\[ \tilde{Q}_1(x), \tilde{Q}_4(x) \geq 0 \]
\[ \tilde{Q}_1(x) + \tilde{Q}_4(x) \leq \max\{Q/2Pr\{X = x\}, 1\} \]

under Assumption 2(ii).

Corollary 6.3 is proved.
Simulated Misclassification Rates

The total misclassification rate in income data, \( Q \), is unknown, but figures prominently in the width of the bounds. While one of the advantages of our approach is that the transition probabilities may be bounded under various assumptions concerning \( Q \), having some prior knowledge concerning plausible values of \( Q \) is important. Here, we discuss one possible way to obtain reasonable choices of \( Q \).

Prior to discussing our approach, it is worth noting that some previous studies have attempted to quantify the measurement error rate in income data by matching self-reported income to administrative data. For example, Pedace and Bates (2000) compare self-reported earnings for over 50,000 respondents in the 1992 SIPP longitudinal file to respondents’ earnings documented in the Social Security Administration’s Summary Earnings Record (SER). The authors find that 3.6% (6.4%) of the final sample report no (positive) earnings in the SIPP despite having positive (no) earnings in the SER. A similar exercise in Kapteyn and Ypma (2007) using Swedish data on older individuals finds that 18.2% (4.6%) report no (positive) earnings in the survey data despite having positive (no) earnings in the administrative records. While interesting, such results are of limited usefulness in the current context of estimating transition probabilities since rates of measurement error in self-reported earnings do not directly translate into misclassification rates when examining transition matrices. Moreover, the amount of misclassification generated by a given amount of measurement error will depend on the dimensionality of the transition matrix.

In light of this, we employ a simulation-based approach to quantify \( Q \) when the data is discretized into a different number of equal-sized partitions. To proceed, we utilize a structural model for income dynamics and measurement error based on an \( AR(1) \) model as in Lee et al. (2017). Note, while Lee et al. (2017) rely on the model to assess mobility, we only rely on the model to suggest plausible values of the misclassification rate, \( Q \).

The model for income dynamics is given by

\[
\ln (y_{it}) = \gamma \ln(y_{i,t-1}) + \alpha_i + \varepsilon_{it}, \quad t = 1, ..., T \tag{F.22}
\]
where $y_{it}^{*}$ is the true income of household $i$ at time $t$, $y_{i,t-1}^{*}$ is the true income at time $t-1$, $\alpha_i$ is a time invariant household-specific intercept, and $\varepsilon_{it}$ is an idiosyncratic error term.

The initial condition is drawn from the stationary distribution of $y$, given by

$$\ln(y_{i0}^{*}) \sim \text{iid } N\left(\frac{\alpha_i}{1-\gamma}, \frac{\sigma_{\alpha}^2}{(1-\gamma)^2} + \frac{\sigma_{\varepsilon}^2}{1-\gamma^2}\right), \quad (F.23)$$

where $\sigma_{\alpha}^2$ and $\sigma_{\varepsilon}^2$ are the variances of $\alpha$ and $\varepsilon$, respectively. The fixed effect, $\alpha$, is drawn from a uniform distribution, given by

$$\alpha_i \sim \text{iid } U(\alpha_l, \alpha_u). \quad (F.24)$$

The mean of $\alpha$, $\bar{\alpha} = 0.5(\alpha_l + \alpha_u)$, is chosen to pin down the expected value of $\ln(y^{*})$ for a given choice of $\gamma$ as

$$E[\ln(y^{*})] = \frac{\bar{\alpha}}{1-\gamma}$$

assuming $E[\varepsilon_{it}] = 0$. The bounds on $\alpha$ are given by

$$\alpha_l = \bar{\alpha} - \kappa$$
$$\alpha_u = \bar{\alpha} + \kappa.$$ 

implying that $\sigma_{\alpha}^2 = \kappa^2/3$. The idiosyncratic error, $\varepsilon$, is drawn from a normal distribution, given by

$$\varepsilon_{it} \sim \text{iid } N\left(0, \sigma_{\varepsilon}^2\right).$$

The variance of $\varepsilon$ is chosen to pin down the variance of $\ln(y^{*})$ for a given choice of $\gamma$ as given in (F.23).

We do not observe true income, $y_{it}^{*}$, but rather mismeasured income, $y_{it}$. Given that there is substantial evidence that measurement error is mean-reverting and serially correlated (Bound and Krueger 1991; Bound et al. 1994; Bollinger and Chandra 2005; Kim and Solon 2005), we express observed income as
\[ \ln(y_{it}) = \rho \ln(y^*_{it}) + \eta_{it}, \quad (F.25) \]

where \( \rho \) denotes the mean reversion parameter, \( \rho \in (0, 1) \). The error, \( \eta_{it} \), is assumed to be independent of \( y^* \) and is decomposed into a time invariant and time-varying component as follows

\[ \eta_{it} = e_i + v_{it}, \quad (F.26) \]

where \( \text{Cov}(e_i, v_{it}) = 0 \) and \( e \) and \( v \) are each normally distributed.

Define the realized measurement error as

\[ \mu_{it} = \ln(y_{it}) - \ln(y^*_{it}) \]
\[ = (\rho - 1) \ln(y^*_{it}) + \eta_{it}, \quad (F.27) \]

where

\[ \text{E}[\mu_{it}] = (\rho - 1) \text{E}[\ln(y^*)] + \text{E}[\eta_{it}] \]
\[ \text{Var}(\mu_{it}) = (\rho - 1)^2 \text{Var}[\ln(y^*)] + \text{Var}[\eta_{it}] \]

The mean and variance of \( \eta \) are chosen to pin down the mean and the variance of \( \mu \) for a given choice of \( \rho \). Finally, we set

\[ \text{E}[e_i] = 0 \quad (F.28) \]
\[ \text{E}[v_{it}] = \text{E}[\mu_{it}] - (\rho - 1) \text{E}[\ln(y^*)] \quad (F.29) \]
\[ \sigma_e^2 = \zeta \left( \text{Var}(\mu_{it}) - (\rho - 1)^2 \text{Var}[\ln(y^*)] \right) \quad (F.30) \]
\[ \sigma_v^2 = (1 - \zeta) \left( \text{Var}(\mu_{it}) - (\rho - 1)^2 \text{Var}[\ln(y^*)] \right) \quad (F.31) \]

140
where $\sigma^2_e$ and $\sigma^2_v$ are the variances of $e$ and $v$, respectively.

In sum, we need to choose values for $\gamma$ (auto-regressive parameter in \eqref{eq:22}), $\rho$ (mean-reverting parameter in \eqref{eq:25}), $\kappa$ (variance of fixed effect in \eqref{eq:24}), $\zeta$ (share of measurement error variance due to time invariant component in \eqref{eq:30}), and $E[\mu_{it}]$ and $\text{Var}(\mu_{it})$ (mean and variance of measurement error in \eqref{eq:27}). However, instead of directly choosing values for $\text{Var}(\mu_{it})$, we instead choose values for the reliability statistic (Gottschalk and Huynh 2010; Abowd and Stinson 2013), given as

$$R = \frac{\text{Var}[\ln(y^*)]}{\text{Var}[\ln(y^*)] + \text{Var}(\mu_{it})}. \quad (\text{F.32})$$

The remaining parameters are chosen to pin down $E[\ln(y^*)]$ and $\text{Var}[\ln(y^*)]$ (mean and variance of true income in \eqref{eq:23}).

We consider the following cases

$$
\begin{align*}
E[\ln(y^*)] & = 10 \\
\text{Var}[\ln(y^*)] & = 2 \\
E[\mu_{it}] & = -0.15 \\
(\gamma, \kappa) & = \{(0.9, 0.2), (0.8, 0.45), (0.7, 0.7), (0.6, 0.95)\} \\
(R, \rho) & = \{(0.95, 0.8), (0.9, 0.7), (0.85, 0.6), (0.8, 0.55)\} \\
\zeta & = \{0.25, 0.50, 0.75\}
\end{align*}
$$

Note, $\gamma$ and $\kappa$ are chosen together to ensure that $\sigma^2_e > 0$ and $R$ and $\rho$ are chosen together to ensure that $\sigma^2_e, \sigma^2_v > 0$.

The mean and the variance of $\ln(y^*)$ are set at values a bit above those observed in our self-reported income data. The mean measurement error, $E[\mu_{it}] = -0.15$, is from Gottschalk and Huynh (2010) and is based on annual earnings of males in the 1996 SIPP panel matched to tax records and assumes the tax records are correct. In contrast, Gibson and Kim (2010) use the Panel Study of Income Dynamics Validation Study (PSIDVS) based on 1981–1986
data and find the mean measurement error to range from -0.22 to 0.22 across the different years. However, in results not shown, \( Q \) is not particularly sensitive to this value.

For the reliability statistic, \( R \), Gottschalk and Huynh (2010) and Abowd and Stinson (2013) both report values around 0.7 when assuming the administrative data are correct. However, when Abowd and Stinson (2013) give equal probability to the SIPP and administrative data being correct, the reliability statistic of the SIPP data rises to above 0.9. Hyslop and Townsend (2016) perform a similar analysis using survey and administrative data from New Zealand and report reliability statistics between 0.83 and 0.85. Thus, we consider values of \( R \) from 0.8 to 0.95. These values for \( R \) then allow us to consider values of \( \rho \) from 0.55 to 0.8. This is a reasonable range based on the literature. Specifically, Bound and and Krueger (1991) and Bollinger (1998) examine the structure of response error when income is the natural log of annual labor market earnings using the Social Security income data matched to the 1977 and 1978 Current Population Survey. The authors obtain estimates of \( \rho \) roughly equal to 0.90. However, using the PSIDVS, Gibson and Kim (2010) find much lower values, with \( \rho \) ranging from 0.3 to 0.7. Kim and Solon (2005) discuss a number of estimates from prior studies, citing values between roughly 0.6 and 0.8.

For the autoregressive parameter, \( \gamma \), we consider a range from 0.6 to 0.9. Gustavsson and Österholm (2014) use Swedish longitudinal data on males observed from 1968-2005 to estimate individual-level time series models of income dynamics. The authors obtain median values of \( \gamma \) ranging from 0.6 to 0.8 across samples and specifications. Finally, without an \textit{a priori} information, we consider a range of values of \( \zeta \) from 0.25-0.75.

With four values of \((\gamma, \kappa)\), four values of \((R, \rho)\), and three values of \(\zeta\), we end up with 48 total cases. For each case, we simulate 1000 data sets with \( N = 10,000 \) and \( T = 100 \). We then retain periods \( t = 97, 100 \) in order compute transition matrices over four periods. The choice of \( N \) is comparable to the data in our application. The choice of \( T \) ensures that initial conditions play a limited role. The choice of a four-year transition matrix follows from our application. Finally, we compute the misclassification rate in each of the 1000 data sets for different dimensions of the transition matrix, \( K = 2, \ldots, 5 \). Denote this misclassification rate as \( Q(K) \).
Table D1 reports the mean misclassification rates. The following findings stand out. First, for any value of the parameters, $Q(K)$ unsurprisingly increases significantly with $K$. Second, for any given combination of $\gamma$, $\kappa$, and $\zeta$, $Q(K)$ increases significantly as $R$ and $\rho$ decrease. This is not surprising since $R$ directly controls the salience of the measurement error. However, the misclassification rate does not increase linearly across the columns (for a given $\zeta$). In particular, the misclassification rate increases by roughly 50% as $R$ falls from 0.95 to 0.9, but then only rises marginally as $R$ falls from 0.9 to 0.85. The misclassification rate then increases by roughly 70% as $R$ falls from 0.85 to 0.80. Third, for any given $K$, $Q(K)$ is marginally decreasing with $\zeta$. This holds for all values of $\gamma$ and $\rho$. This implies that misclassification is less likely to occur when the time invariant component of the measurement error, $\eta$, is more salient. Finally, for any given combination of $R$, $\rho$, and $\zeta$, $Q(K)$ decreases as one moves from Panel A to D. As one moves from Panel A to D, $\gamma$ declines and $\kappa$ increases. As $\gamma$ declines, $\text{Var}[\ln(y^*)]$ declines as shown in (F.23). Because we wish to hold $\text{Var}[\ln(y^*)]$ constant (and equal to two), $\sigma^2_\alpha$ and/or $\sigma^2_\varepsilon$ must increase. The corresponding rise in $\kappa$ ensures that some of the increase is due to the fixed effect, $\alpha$. The fall in the misclassification rate across the panels of the table implies that misclassification is less likely to occur when income shocks are less permanent.

As it relates to our application, the assumption of $Q = 20\%$ seems reasonable in the case of tercile transition matrices. The corresponds to roughly the maximum misclassification rate when $K = 3$ and the reliability statistic is at least 0.85. In the case of poverty transition matrices, we use $Q = 10\%$. This corresponds to roughly the maximum misclassification rate when $K = 2$ and the reliability statistic is at least 0.85.
BIBLIOGRAPHY


