The Pennsylvania State University

The Graduate School

College of the Liberal Arts

Essays in Firm and Labor Market Dynamics

A Dissertation in
Economics
by
Jiaming Mao

© 2013 Jiaming Mao

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2013
The dissertation of Jiaming Mao was reviewed and approved* by the following:

James R. Tybout
Professor of Economics
Dissertation Adviser
Chair of Committee

Edward Green
Professor of Economics and Mathematics

Theodore Papageorgiou
Assistant Professor of Economics

Paul Grieco
Assistant Professor of Economics

Eric Plutzer
Professor of Political Science and Sociology

Robert C. Marshall
Head, Department of Economics & Professor of Economics

*Signatures are on file in the Graduate School.
Abstract

This thesis consists of three self-contained chapters on firm and labor market dynamics.

In Chapter 1, I analyze the evolution of the U.S. labor market by developing and estimating a dynamic general equilibrium model that simultaneously accounts for the evolution of wages, employment, and educational attainment in the U.S. from 1975 to 2010. The model features overlapping generations of individuals who self-select into education, employment, and occupations in an environment with competitive markets and exogenous demand and supply shifts. It departs from existing general equilibrium studies by adopting a rich characterization of heterogeneity in the form of continuous ability distributions.

I estimate the model on the Current Population Survey (CPS) using simulated method of moments. The estimated model successfully matches trends in the U.S. labor market and reveals changes in the underlying ability composition of the workforce. In particular, my results contradict earlier findings that increase in college attainment has led to a large decline in the average ability of college graduates and suggest that female selection into the workforce has become more positive over the last four decades.

Using the estimated model as a basis for counterfactual experiments, I first examine the impact of exogenous demand and supply forces by allowing them to happen one at a time. I then assess composition effects and highlight the sensitivity of labor market outcomes to underlying ability heterogeneity by simulating economies with different ability distributions. The results show that composition effects have played a positive but minor role in the increase of the
college premium and the narrowing of the gender wage gap, and suggest that the observed labor market changes are the results of a confluence of forces that have acted in concert or in balance with each other over the last four decades. Skill-biased technological change has increased the college premium and raised college attainment, but has also contributed to closing the gender wage and employment gaps. Gender-biased demand shifts and changing home sector values have increased female relative employment and wages, while depressing the growth in male college attainment and the college premium. I quantify the respective contributions of these exogenous forces and highlight their general equilibrium interactions.

In Chapter 2, I study the graduate school enrollment decision of U.S. college graduates. The key question is that although college enrollment has increased significantly in the U.S. over the last four decades, there has been limited growth in the enrollment rate of most graduate programs, despite the fact that the returns to graduate education have increased more than the returns to college education. To understand the graduate school enrollment behavior as well as to investigate possible barriers of entry, I develop and estimate a dynamic discrete choice model of post-college work and educational investment decisions. The model consists of two stages. In the first stage, individuals enter college and choose a major field of study based on expectations of grades, future earnings, and the probabilities of enrolling in different graduate programs in the future. In the second stage, individuals are graduated from college and in period choose whether to enroll in a graduate program, work, or engage in home production. In each case, the individual’s choice is rationalized as the optimal solution to a dynamic discrete choice problem that maximizes her expected lifetime utility.

One of the key features of the study is that it jointly models college major and graduate school choices. This includes the recognition that the choice of college major is both an outcome of a student’s talent and prior preparation and a determinant of her later options. The choice of a dynamic structural model captures the dynamics of post-college educational investment. While most people attend school continuously from grade school to college, years
of employment or home production can separate graduate school enrollment from college graduation. In addition, post-college education is characterized by frequent entries and exits as people move between work and school. Such dynamics can only be captured by a dynamic model.

I estimate the model using data from the 1992-1993 Baccalaureate and Beyond Longitudinal Study (B&B:93/03), which follows a representative sample of U.S. college graduates who earned their bachelor’s degrees in the 1992-93 academic year. My results find strong sorting of preferences toward undergraduate majors and graduate programs by mathematical ability. Individuals with high math ability favor non-education majors in college, are more likely to work rather than engage in home production, and have stronger preference toward enrolling in business, professional, and doctoral programs (as opposed to non-MBA master’s programs) after college. I also find substantial transition costs associated with going back to school after college, which significantly deters individuals from pursuing graduate education in response to their increasing returns.

In Chapter 3, I explore the use of Bayesian methods in estimating dynamic games of imperfect competition. Structural estimations of dynamic models have been known for their computational demands. One of the major sources of computational burden lies in the necessity of solving the dynamic programming (DP) problems of forward-looking agents. In the context of maximum likelihood estimation, the likelihood functions are based on the explicit solution of these DP problems. Estimations therefore have to proceed by iterating on a set of Bellman equations until convergence for all values of model parameters considered by the estimator. Compared to single agent problems, dynamic games of imperfect competition are more complex, as each agent’s choice of action depends on her belief about other agents’ private information and choice probabilities. Solution of such models require both value functions and choice probability functions to converge in equilibrium.

To alleviate the computational burden of estimating dynamic games of imperfect competition, I develop a Bayesian Markov chain Monte Carlo (MCMC) procedure based on an algorithm developed by Imai, Jain, and Ching (2009) for
inference in single-agent problems. The key is to employ local nonparametric regression techniques to approximate value functions and choice probability functions and to combine the parameter search step with the model solution step. I propose three algorithms that extend the Imai-Jain-Ching procedure to the estimation of multi-agent games. The performance of each algorithm is tested using an entry and exit model, from which I determine the optimal procedure.
Contents

List of Tables x

List of Figures xii

1 Heterogeneity, Technical Change, and the Evolution of the
U.S. Labor Market 1

1.1 Introduction 2

1.2 The Model 9

  1.2.1 The Environment 9

  1.2.2 The Individual Problem 13

    1.2.2.1 Education Decision 13

    1.2.2.2 Labor Supply and Occupation Choice 15

  1.2.3 Exogenous Processes 18

  1.2.4 Equilibrium 20

1.3 Estimation 22

  1.3.1 Distributional Assumptions 22

  1.3.2 Estimation Method 23

  1.3.3 Identification 25

1.4 Results 29

  1.4.1 Parameter Estimates 29

  1.4.2 Model-Data Comparison 31

  1.4.3 The Changing Composition of College Graduates and
  the U.S. Workforce 34

  1.4.4 Accounting for Labor Market Changes 37

vii
List of Tables

1.1 Employment Rates by Age (%) ........................................ 50
1.2 Employment Rates by Year (%) .................................... 50
1.3 Transition Rates by Age (%) ....................................... 51
1.4 Transition Rates by Year (%) ...................................... 52
1.5 Ability Endowment: Parameter Estimates ...................... 53
1.6 Ability Endowment: Summary Statistics ....................... 53
1.7 Production Function: Parameter Estimates ..................... 54
1.8 Skill Production: Parameter Estimates .......................... 54
1.9 Utility of Work: Parameter Estimates ............................ 54
1.10 Value of Residual Sector: Parameter Estimates .............. 55
1.11 Cost of Mobility: Parameter Estimates ....................... 55
1.12 Cost of Schooling: Parameter Estimates ...................... 55
1.13 Idiosyncratic Shocks: Parameter Estimates ................... 56

2.1 Wage Regressions: Bachelor’s and Master’s Degree ........ 92
2.2 Mean Wage: MBA, Professional, and Doctoral Degree ...... 93
2.3 GPA Regressions ..................................................... 94
2.4 Estimated Degree Attainment Probability Function .......... 95
2.5 Estimated transition Costs ......................................... 96
2.6 Estimated Type-specific Preferences ............................ 97
2.7 Estimated Type Proportions ....................................... 98
2.8 Goodness of Fit (a): Major Choice .............................. 99
2.9 Goodness of Fit (b): Activity .................................... 100
2.10 Goodness of Fit (c): Degree Attainment ...................... 100
2.11 Conditional Type Probability Parameters for the Extended Model 107
2.12 Aggregate Moments ............................................. 108
2.13 Major Distribution by SAT scores ............................... 109
2.14 Highest Graduate Enrollment 2003 by Highest Degree Attainment
1997 ................................................................. 110

3.1 Convergence Time* .................................................. 129
3.2 BDP estimation of dynamic entry and exit model* .............. 157
3.3 BDP estimation of dynamic entry and exit model with random grid
generation* ....................................................... 161
List of Figures

1.1 Skill-Experience Profile ................................................. 56
1.2 College Enrollment Probability ........................................ 57
1.3 College enrollment decision under homogeneous attendance cost . 57
1.4 GDP Components and the imputed \{\gamma_t\}, series .................. 58
1.5 Labor Supply and Occupation Choice Decision Tree .................. 59
1.6 Convergence in Gender Wage Disparities by Occupation ............ 59
1.7 College Enrollment Probability (a) given relative payoff; (b) given attendance cost. ......................................................... 60
1.8 Complementary Uniform and Beta(5,5) CDF ............................ 60
1.9 Ability Distribution, Idiosyncratic Shock Variance, and Elasticity of Supply. The horizontal axis is the value associated with a choice and the vertical axis is the percentage of individuals making the choice in the population. ................................................................. 61
1.10 Estimated Ability Distribution in Men. Cooler colors represent lower densities and warmer colors represent higher densities. Histograms of each ability dimension are plotted next to the vertical and horizontal axes. 61
1.11 Model-Data Comparison - Relative Wages. (a) log college/non-college wage ratio. (b) female/male wage ratio. .............................. 62
1.12 Model-Data Comparison - Education and Labor Supply. (a) college attainment rate by gender among individuals 25 - 29 years old. (b) rate of employment by gender ................................. 62
1.13 Model-Data Comparison - Wage Variance. (a) relative changes in the overall variance of log wages. (b) relative changes in the variance of log wages among college- and non-college-educated individuals. Relative change = \( \frac{V_t - V_{1975}}{V_{1975}} \). The initial values \( V_{1975} \) are reported within the legends.

1.14 Evolution of the Mean Cognitive Ability of College Graduates in Entering Cohorts...

1.15 Ability Distribution by Education among Females in Selected Entering Cohorts. (a) - (d) shows respectively the 1975, 1985, 1995 and 2010 cohorts. In each plot, the red dots represent individuals who have obtained college education and the green dots represent those who haven’t. Nonparametric density plot of the cognitive ability of college graduates is given next to the vertical axis. Nonparametric density plot of the physical labor endowment of college graduates is given below the horizontal axis.

1.16 Ability Distribution by Education among Males in Selected Entering Cohorts. (a) - (b) shows respectively the 1975 and 2010 cohorts. In each plot, the red dots represent individuals who have obtained college education and the green dots represent those who haven’t. Nonparametric density plot of the cognitive ability of college graduates is given next to the vertical axis. Nonparametric density plot of the physical labor endowment of college graduates is given below the horizontal axis.

1.17 Evolution of Mean Cognitive Ability and Physical Labor in the College-educated Population...

1.18 Evolution of Population Log College/Non-College Cognitive Ability Ratio...

1.19 Evolution of Mean Cognitive Ability and Physical Labor in Employment...

1.20 Decomposing Labor Market Changes - Relative Wages. (a) log college/non-college wage ratio. (b) female/male wage ratio. The labels in the legend refer to the specific component turned on in the experiment. “RES” denotes changes in residual sector value, “HN” Hicks-neutral technological growth, “SB” skill-biased technological change, and “GB” gender-biased demand shift.

1.22 Ability Distribution and Labor Market Outcomes - Relative Wages. (a) log college/non-college wage ratio. (b) female/male wage ratio. The labels in the legend refer to the specific ability distributions assumed in the experiment. “No CA” denotes no gender comparative advantage in ability endowments, “Uniform” uniform ability distributions, and “Degenerate” degenerate ability distributions.

1.23 Ability Distribution and Labor Market Outcomes - Education and Labor Supply. (a) college attainment rate by gender among individuals 25 - 29 years old. (b) female/male employment rate. The labels in the legend refer to the specific ability distributions assumed in the experiment. “No CA” denotes no gender comparative advantage in ability endowments, “Uniform” uniform ability distributions, and “Degenerate” degenerate ability distributions.
Chapter 1

Heterogeneity, Technical Change, and the Evolution of the U.S. Labor Market
Abstract

This paper analyzes the evolution of the U.S. labor market by estimating a dynamic general equilibrium model that simultaneously accounts for changes in wages, employment, and educational attainment in the U.S. from 1975 to 2010. The model features overlapping generations of individuals who self-select into education, employment, and occupations in an environment with competitive markets and exogenous demand and supply shifts. It departs from existing general equilibrium studies by adopting a rich characterization of heterogeneity in the form of continuous two-dimensional ability distributions. The model is fit to the Current Population Survey and successfully matches trends in the U.S. labor market. The estimated model reveals changes in the ability composition of the workforce. Using the model as a basis for experiments, I first quantify the impact of exogenous demand and supply forces and highlight their general equilibrium interactions. I then assess composition effects and highlight the importance of ability heterogeneity in shaping labor market outcomes. The paper contributes to a better understanding of U.S. labor market dynamics over the last four decades.

1.1 Introduction

The U.S. labor market has undergone a dramatic transformation in recent decades. The wage structure has become more unequal and skill differentials have risen, while gender disparities in education, labor supply and wages have shrunk.

While a large literature has examined these trends, most studies have treated them as separate phenomena. The growth of the college premium has been attributed to skill-biased technological change\(^1\). The narrowing of the gender wage gap has been attributed to increase in female human cap-

ital and decline in discrimination\textsuperscript{2}. Increased female labor supply has been attributed to a combination of rising wages, increase in home productivity and decline in marriage and fertility rates\textsuperscript{3}. The partial equilibrium nature of these studies, however, prevent them from studying the general equilibrium interactions of these factors. What is the impact of skill-biased technological change on relative female labor supply and the gender wage gap? How does the decline in discrimination affect male education attainment and labor supply? Few studies have answered these questions. Attempts to quantify the influence of each channel in a dynamic equilibrium framework are rarer still.

Given the substantial increase in educational attainment and female labor supply, the composition of the workforce, in terms of its underlying ability, has also likely changed. Thus a related set of questions concerns the direction and magnitude of composition effects. Has the increase in college attainment led to a decrease in the average quality of college graduates? Has the ability of women who enter the labor force decreased or increased over time? To what extent, if any, have the resulting composition effects contributed to the observed rising college premium and narrowing gender wage gap? The empirical literature on these questions is dominated by reduced form studies that have offered contradictory answers\textsuperscript{4}. The conclusion by some authors that composition effects can be substantial and are changing over time suggests that it is important to control for changes in ability composition, and that the neglect to do so may well distort our assessment of the drivers of labor market changes.

In this paper, I address these issues by estimating a dynamic general equilibrium model that simultaneously accounts for the evolution of wages, em-

\textsuperscript{3}See, e.g., Costa (2000), Greenwood et al. (2005), Goldin (2006), Attanasio et al. (2008).
\textsuperscript{4}Carneiro and Lee (2011) find that increases in college attainment led to a large decline in the average quality of male college graduates between 1960 and 2000. Juhn et al. (2005) find only weak adverse compositional shifts between 1940 and 1990. Taubman and Wales (1972) find that the average aptitude of those attended college actually rose relative to those who did not between 1920 and 1960. In terms of labor force composition, Mulligan and Rubinstein (2008) argue that the narrowing of the gender wage gap is almost entirely attributable to composition effects. Blau and Kahn (2007) find that female selection into the workforce was positive during the 1980s but negative during the 1990s.
ployment and educational attainment in the U.S. from 1975 to 2010. My framework is closely related to Lee and Wolpin (2006, 2010). The economy is populated by overlapping generations of individuals who self-select into education, employment, and occupations. A final good is produced by combining intermediate output from skill-intensive and labor-intensive occupations. Production within each occupation is governed by CES technologies with constant returns to scale in two factors, skill and physical labor. In addition, there is a home sector to allow for labor supply decisions. Individuals are \textit{ex ante} heterogeneous in their abilities, which comprise two dimensions, cognitive ability and physical labor. During the life cycle, cognitive ability governs the rate of skill formation. Labor markets are competitive and workers receive the marginal product of the skill and physical labor they supply. The evolution of the labor market is driven by a combination of exogenous demand and supply forces, including neutral and skill-biased technological change, gender-biased demand shifts, changes in home sector value, and variations in cohort sizes.

One of the key features of the model lies in its treatment of heterogeneity. In reality, all men (and women) are not created equal. Those endowed with great intelligence and communication ability benefit more from receiving formal education and are potentially more productive in skill-intensive jobs, while those endowed with great strength and coordination are potentially more productive in labor-intensive jobs. One of the most important functions of the labor market is to allocate the heterogeneous pool of talent to their most productive activities. I capture these features of the labor market by assuming individuals differ in their term of both their cognitive and their physical abilities, and by assuming both dimensions of heterogeneity are distributed continuously and time-invariantly in the population. In addition, to capture gender difference, men and women are allowed to have different physical labor distributions\textsuperscript{5,6}. I estimate the ability distributions and show that they

\textsuperscript{5}The distribution of cognitive ability is assumed to be the same among men and women. \textsuperscript{6}The interpretation of gender ability difference need not be a biological one. Persistent difference in the market valuation of men and women’s physical ability is likewise reflected in this paper as difference in \textit{ex ante} endowment. Thus, more precisely, the paper models difference in men and women’s \textit{marketable} physical labor, which serves as a source of gender
not only affect the observed wages through composition effects, but also affect labor force participation and college attainment rates through their effects on aggregate elasticities.

Importantly, the difference in ability distributions gives rise to gender comparative advantage in education and different occupations, and can make skill-biased technological change an inherently gender-biased process. The mechanism is similar to Welch (2000): if women are relatively intensive in their possession of cognitive ability, then an increase in the price of skill relative to physical labor benefits women more than men. This mechanism could contribute to the observed simultaneous increase in within-gender inequality and decrease in between-gender inequality. Indeed, my results suggest that this is the case.

I estimate the model using simulated method of moments. The procedure involves repeatedly simulating the economy, solving for its equilibrium, and matching the simulated moments on wages, employment, and educational attainment to data moments from the March supplements of the Current Population Surveys (CPSs) from 1975 to 2010. The estimated model successfully matches trends in the U.S. labor market. The model’s estimates reveal significant female-biased demand shifts that raise the relative market valuation of female inputs and skill-biased technological change in the form of both shifts toward the skill-intensive occupation in final good production and shifts toward skill inputs within skill-intensive and labor-intensive occupations. I find women on average are endowed with less physical labor, leading to female comparative advantage in education and skill-intensive jobs.

By explicitly modeling individual ability and accounting for selection, the comparative advantage.

\footnote{In addition to Welch (2000), several other authors have distinguished the roles of “brains” and “brawn” in arguing that skill-biased technological change has an inherent female bias. Weinberg (2000) finds that increase in computer use in the workplace is associated with increase in the demand for female workers. Rendall (2010) conducts a principal component analysis of occupations in the U.S. Dictionary of Occupational Titles (DOT) and projects the skill requirements of an occupation into three components, “brain,” “brawn,” and “motor coordination.” He concludes that there has been a “brain-biased technological change” in the U.S. since the 1950s that has been the main driver of gender wage convergence.}
model allows me to directly examine changes in ability composition. In particular, I show that as more people attend college and enter the workforce, the average cognitive ability of college graduates and that of the workforce do not necessarily go down, but can first increase then decrease due to changing selectivity into college and employment. Furthermore, since the population is composed of cohorts that make their education decisions at different times, the population trend in the average cognitive ability of college graduates displays hysteresis and can differ markedly from the trend observed in entering cohorts. Specifically, I find that (1) the average cognitive ability of 25–29 year old college graduates increased from 1975 to the late 1980s and then decreased; (2) the average cognitive ability of the college-educated population increased from 1975 to the early 2000s and then decreased; (3) the average cognitive ability of the workforce increased throughout the period of 1975–2010, while its average physical labor decreased throughout the period. Most of these changes are driven by female entry into college and the workforce.

Using the estimated model as a basis for counterfactual experiments, I am able to trace labor market changes to their underlying causes. Two sets of experiments are conducted. First, I assess the impact of exogenous demand and supply forces by allowing them to happen one at a time. The results suggest that (1) in addition to female-biased demand shifts, neutral and skill-biased technological change both contribute to increasing female relative labor supply and decreasing the gender wage gap; (2) female-biased demand shifts discourage male college attainment and labor supply, and would by themselves lead to a lower college premium; (3) the value of the home sector has been increasing for men and decreasing for women; this process contributes to increasing

As will be discussed later, this is possible because the model allows variation in both individual ability and preference/cost, which breaks the monotone relationship between the size and average ability of a self-selected group. Intuitively, in the case of college attainment, when the value of college is low, individuals are more influenced by their idiosyncratic costs in making enrollment decisions and self-selection into college is relatively random. As the value of college increases, it will first draw high ability individuals into college who previously are not going to enroll, then lower ability ones as more attend college. The degree of selectivity into college can thus first increase and then decrease with enrollment size. Juhn et al. (2005) illustrates the same point using a model of college enrollment with variations in ability and discount rates.
female relative labor supply and college attainment and helps to narrow the gender wage gap.

Second, I assess composition effects and the importance of female comparative advantage by simulating respectively an economy with degenerate ability distributions and with no gender difference in the distribution of physical labor. The results suggest that (1) composition effects have contributed to increasing the college premium and narrowing the gender wage gap, but the effects are quite small; (2) female comparative advantage is an important source in the convergence of gender wage and employment gaps and serves to slow down the growth in the college premium. The results also highlight the sensitivity of labor market outcomes to underlying ability distributions.

The paper builds on a tradition of prior work. The model embeds the Roy model of self-selection and earnings in a dynamic general equilibrium setting and extends the Ben-Porath human capital model by adopting the two factor human capital framework of Guvenen and Kuruscu (2010). It extends Heckman, Lochner and Taber (1998)'s study of male schooling and skill premia to account for female education, wages, and labor supply.

The paper is most closely related to Lee and Wolpin (2006, 2010), who develop a two-sector competitive labor market model with exogenous demand and supply shifts to study the rise of the service sector and changes in the U.S. wage and employment structure. There are several main differences between their papers and mine. First, their model contains only a limited number of \textit{ex ante} individual types. In this paper, I adopt a richer characterization of heterogeneity in the form of two-dimensional, continuous ability distributions. This allows me to better capture changes in the underlying ability composition of the workforce, and examine the direction and magnitude of composition effects. Second, in their model, gender difference in job preferences gives rise to female comparative advantage in service sector or white/pink-collar jobs.

\footnote{More precisely, an economy is simulated in which cognitive ability and physical labor distributions are degenerate within gender. Gender difference in mean physical labor endowment persists.}

\footnote{Lee and Wolpin (2006) and Lee and Wolpin (2010) assume, respectively, four and five individual types}

\footnote{The two papers give different estimates. In Lee and Wolpin (2006), women are esti-}
In my model, gender comparative advantage derives mainly from difference in ability endowments, which not only influences selection, but interacts with skill-biased technological change to raise female relative wages. Third, in their model, increase in female relative wages is due to selection and skill-upgrading. Within sector-occupation cells, no demand side forces exist to close the gender gap. In my model, the convergence of within-occupation gender gaps is driven on the demand side by both female-biased demand shifts and within-occupation skill-biased technological change. Finally, there are a number of features in their model that are missing in mine, including the rise of the service sector, capital-skill complementarity, and the effect of children on female labor supply.

Johnson and Keane (2013) conduct a similar analysis of the forces driving U.S. labor market changes by building on the model of Lee and Wolpin (2006, 2010). Their main innovations are a finer differentiation of labor by occupations and allowing more types of labor to be imperfect substitutes in production. The increase in the number of occupations and the complexity of substitution patterns is achieved through a reduction in ex ante individual types: the authors assume individuals are ex ante homogeneous. Our analyses therefore should be seen as complementary.

Finally, a number of authors have studied labor market dynamics in a trade context. Artuç, Chaudhuri, and McLaren (2010) and Artuç and McLaren estimated to strongly prefer service sector jobs, but within the service sector, they are estimated to have the strongest preference for blue-collar jobs. In Lee and Wolpin (2010), women do not have stronger preference for service sector jobs in comparison with men. Instead, men are estimated to strongly prefer blue-collar jobs in both sectors, giving women a comparative advantage in white- and pink-collar jobs.

More precisely, skill-biased technological change takes two forms in this paper. The first involves shifts toward skill-intensive jobs in the production of the final good. This change is also present in Lee and Wolpin (2006, 2010) and likewise results in an increase in female relative wages due to women’s comparative advantage generated by their job preferences. The second involves within-occupation shifts toward skill relative to physical labor. This change is not present in Lee and Wolpin (2006, 2010). In my model, this within-occupation skill-biased technological change interacts with gender difference in abilities to raise the within-occupation female relative wages.

Both my paper and Johnson and Keane (2013) can be viewed as extensions of Lee and Wolpin (2006, 2010). I adopt a richer characterization of ex ante heterogeneity, while Johnson and Keane adopt a richer specification of the production function.
(2013) develop a parsimonious structural dynamic model of the U.S. labor market to study the effects of trade liberalization on labor adjustment and wage inequality. Dix-Carneiro (2010) estimate a model similar to Lee and Wolpin (2006, 2010) with emphasis on costs of mobility and sector-specific human capital to study trade-induced transitional dynamics and assess the welfare implications of different labor market policies.

The rest of the paper is organized as follows. Section 1.2 describes the model. Section 1.3 describes the estimation procedure and discusses the strategy for identification. Section 1.4 presents the estimation results and conducts counterfactual experiments based on the estimated model. Finally, Section 1.5 concludes the paper.

1.2 The Model

1.2.1 The Environment

Time is discrete. In any given period, the economy is populated by agents of age \( a \in \{1, \ldots, A\} \). At the beginning of each period, a new cohort of age 1 agents enter the economy, while those reaching age \( A + 1 \) exit. Cohort sizes are exogenous and therefore the age distribution of the model population is nonstationary.

The economy consists of a formal sector and a non-productive residual sector. Workers in the formal sector work in two types of occupations: skill-intensive occupation and labor-intensive occupation, both utilizing a combination of skill and labor as factors of production.

Schooling and work take place sequentially. There are two education levels: college education and no college education. Individuals choose their optimal education prior to entering the economy. After entering the economy, individuals make labor supply decisions and occupational choices in each period, subject to idiosyncratic shocks and mobility costs.

The labor market is closed and perfectly competitive. Workers receive the marginal product of the labor and skill they bring to their jobs, which,
unobservable to the econometrician, are assumed perfectly revealed in the
market place. Preferences are defined by a momentary utility function linear in
consumption. Each agent maximizes her lifetime expected present discounted
value of flow utilities at a common discount rate $\beta \in (0, 1)$.

**Ability Endowments** Individuals are *ex ante* heterogeneous along two di-
mensions: physical labor ($\ell$) and cognitive ability ($\kappa$)\(^{14}\). The distribution of
these two attributes among entering cohorts is assumed to be time invariant
and gender-specific. Specifically, each man draws independently upon birth
an endowment of $(\ell_i, \kappa_i)$ according to

$$
\begin{align*}
\ell_i & \sim f_\ell (\ell) \equiv Beta (\ell; b_\ell^1, b_\ell^2, \ell, \ell) \\
\kappa_i & \sim f_\kappa (\kappa) \equiv Beta (\kappa; b_\kappa^1, b_\kappa^2, \kappa, \kappa)
\end{align*}
$$

(1.1)

, where

$$
Beta (x; b_x^1, b_x^2, x, \pi) = \frac{1}{B (b_x^1, b_x^2)} \frac{(x - \pi)^{b_x^1-1} (\pi - x)^{b_x^2-1}}{(\pi - x)^{b_x^1+b_x^2-1}}, \quad x < \pi < x, \ b_x^1, b_x^2 > 0
$$

, and $B(.,.)$ is the beta function.

Women are assumed to have the same cognitive ability distribution $f_\kappa (\kappa)$,
but draw their labor endowment instead from the distribution $\tilde{f}_\ell (\ell) \equiv f_\ell (\ell / \chi_\ell) / \chi_\ell$
($\chi_\ell > 0$), so that on average, female labor endowment is $\chi_\ell$ times that of men.

The difference in ability endowments between the sexes gives rise to the
possibility of gender comparative advantage. In particular, if $\chi_\ell < 1$, then
the female labor distribution is first-order stochastically dominated by that of
men, and women can be said to be relatively intensive in “brains”, while men
have comparative advantage in “brawn.” The family of distributions defined
in (1.1) are four-parameter beta distributions whose shapes are regulated by
$(b_\ell^1, b_\ell^2)$ and $(b_\kappa^1, b_\kappa^2)$, and whose support sit on $[\ell, \ell]$ and $[\kappa, \kappa]^{15}$. The choice of

\(^{14}\)In the remaining article, the terms “physical labor” and “labor” are used interchange-
ably.

\(^{15}\)The four-parameter beta distribution includes many commonly used distributions as
special cases, such as the t, F, and binomial distributions. See McDonald and Xu (1995)
for more discussions on the distribution family.
the beta distribution family reduces the dimensionality of the problem while maintaining flexibility and allowing a variety of ways in which the two factors are distributed in the population.

**Skill Production** Over the life cycle, an individual’s physical labor endowment stays constant, while she acquires skill through education and experience. Each newborn person is endowed with 1 unit of skill and the rate of her skill acquisition depends on cognitive ability. Specifically, the skill production function is given by

\[
\log s^j_{it} = \kappa_i \left( e_i + \omega_1 a^j_{it} + \omega_2 \left( a^j_{it} \right)^2 \right) \quad j \in \{H, L\} \tag{1.2}
\]

, where \( e_i \in \{0, 1\} \) denotes \( i \)'s education level and \( a_{it} \), \( i \)'s model age in time \( t \), serves as a proxy for her labor market experience\(^\text{16}\).

Figure 1.1 illustrates the typical skill-experience profile implied by this functional form\(^\text{17}\). Differences in cognitive ability translate into greater differences in skill as experiences accrue. Education amplifies such differences as individuals with higher cognitive abilities benefit more from acquiring college education.

**Aggregate Production** The economy produces a final good using two intermediate goods produced respectively by workers in skill-intensive and labor-intensive occupations. Aggregate production is governed by the following Cobb-Douglas technology

\[
Y_t = A_t \left( Y_t^H \right)^{\gamma_t} \left( Y_t^L \right)^{1-\gamma_t} \tag{1.3}
\]

, where \( A_t \) represents aggregate productivity and \( (Y_t^H, Y_t^L) \) denote respectively the output of skill-intensive and labor-intensive occupations. Output

\(^{16}\)The choice of modeling skill production as a function of model age rather than actual labor market experience is made out of convenience, since accounting explicitly for labor market experience would require the addition of two state variables (one for each occupation), significantly increasing the computational burden of the problem.

\(^{17}\)The figure assumes \( \omega_1 > 0 \) and \( \omega_2 < 0 \), which is consistent with estimation results.
share $\gamma_t$ is allowed to vary exogenously over time, reflecting changing occupational compositions of the economy.

Production in each occupation involves the use of labor and skill in different proportions. Labor and skill supplied by each gender are perfect substitutes, but weighted differently in production\textsuperscript{18}. Specifically, let $L^{j,g}$ and $S^{j,g}$ denote respectively the aggregate amount of labor and skill of supplied by workers of gender $g \in \{m, f\}$ employed in occupation $j \in \{H, L\}$. Intermediate goods production is governed by the following constant elasticity of substitution (CES) technology

$$Y^j_t = \left[ \alpha^j_t \left( S^j_t \right)^{\rho^j} + (1 - \alpha^j_t) \left( L^j_t \right)^{\rho^j} \right]^{1/\rho^j}, \quad j \in \{H, L\} \quad (1.4)$$

, where

$$S^j_t = \lambda_t S^{j,f}_t + (1 - \lambda_t) S^{j,m}_t$$

$$L^j_t = \lambda_t L^{j,f}_t + (1 - \lambda_t) L^{j,m}_t$$

In this specification, the elasticity of substitution between labor and skill in occupation $j$ is given by $\sigma^j = 1/(1 - \rho^j)$. The factor shares $\alpha^j_t$ govern the intensity in which skill is used in occupation $j$, and $\lambda_t$ reflects differences in the demand for male and female inputs. The existence of gender-specific demand could be attributable to a number of sources, including labor market discrimination (Darity and Mason 1998), occupational segregation and vertical gender gap (Blau and Kahn 2000), as well as social and cultural norms (Haveman and...\textsuperscript{12}

\textsuperscript{18} The assumption of male and female inputs being perfect substitutes in production follows Heathcote \textit{et al.} (2010). Estimates of the elasticity of substitution between male and female individuals of similar qualifications are high. For example, Johnson and Keane (2013) estimate an elasticity above five for men and women in the same education/occupation/age group.
Labor Markets  Labor markets are competitive. Wages are determined in spot equilibrium and workers receive the marginal product of the labor and skill they supply. Let \( r_{t}^{L,j,g} = \frac{\partial Y_t}{\partial L_{t}^{j,g}} \) and \( r_{t}^{S,j,g} = \frac{\partial Y_t}{\partial S_{t}^{j,g}} \) denote respectively the equilibrium price of a unit of labor and skill of gender \( g \) in occupation \( j \). Then a worker receives a wage in time \( t \) from occupation \( j \) equal to

\[
w_{it}^{j} = r_{t}^{L,j,g} \ell_{i} + r_{t}^{S,j,g} s_{it}, \quad j \in \{H, L\}
\]

where \( g_{i}, \ell_{i}, \) and \( s_{it} \) are respectively the worker’s gender, labor and skill endowment in time \( t \).

1.2.2 The Individual Problem

1.2.2.1 Education Decision

I now describe the individual’s decision problems. Before entering the economy, individuals choose their optimal education level, given their differences in sex, cognitive ability, and physical labor endowment. The cost of attending college, \( c^{e} \), is idiosyncratic and drawn from a time-invariant gender-specific distribution \( F_{g}^{c} \). This distribution captures, in reduced form, cross-sectional

\[19\]

Haveman and Beresford (2012) argue that widely held cultural schemas about what is appropriate for men and women to do (gender norms) and what it is that men and women do well (gender roles) may be the root cause of differences between men’s and women’s educational attainment, job preferences, and work experience. The authors identify several beliefs that can lead to difference in gender-specific demand, such as “men are better than women at math and science,” and “men are more natural managers and leaders than women.”

\[20\]

Other factors include gender comparative advantage in dimensions other than physical labor and skill. Borghans et al. (2006), for example, argue that technological and organizational changes have increased the importance of social skill, or “people skill”, in the workplace, which females have comparative advantage in.

\[21\]

In addition, since model age is used in this paper as a proxy for actual labor market experience (equation (1.2)), \( \lambda_{t} \) also reflects the gender gap in accrued labor market experiences. Intuitively, the term \( (1 - \lambda_{t})/\lambda_{t} \) creates a time-varying wedge between the wages of men and women who are otherwise identical in the context of the model. Therefore, \( \lambda_{t} \) can be understood as capturing the residual gender wage gap unexplained by gender differences in ability endowments, education achievement, occupational choice, and labor supply behavior that are modeled in this paper.
variation in the psychological and pecuniary factors that make acquiring a college degree costly, such as variation in tuition fees, parental resources, access to credit, and government aid programs.

When individuals consider whether to acquire college education, they weigh their draw of the college attendance cost against the expected benefit of entering the economy with a college degree. Let $\mathbb{W}_t^0(\ell, \kappa, g)$ and $\mathbb{W}_t^1(\ell, \kappa, g)$ be respectively the expected value of entering the economy with and without a college degree in time $t$ for an individual with labor and cognitive ability endowment $(\ell, \kappa)$ and of gender $g$. The optimal education decision is characterized by

$$e_i = \begin{cases} 
1 & \text{if } \mathbb{W}_t^1(\ell_i, \kappa_i, g_i) - c_i^e \geq \mathbb{W}_t^0(\ell_i, \kappa_i, g_i) \\
0 & \text{otherwise} 
\end{cases}$$

(1.6)

Figure 1.2 provides an illustration of $\Pr(e_i = 1|\ell_i, \kappa_i)$ for a given $(g_i, c_i^e)$. The probability of college enrollment is increasing in $\kappa$ and decreasing in $\ell$. This is a result of the complementarity between cognitive ability and skill acquisition (equation (1.2)) and of the increasing value of the outside option in physical labor endowment: The higher her cognitive ability, the more skill an individual acquires through education; and the higher her physical labor endowment, the more she earns without a college degree. In the extreme case when there is no variation in attendance cost within the population (i.e. $c^e$ is constant), the policy function (1.6) produces a cut-off line in the $(\ell, \kappa)$ plane above which all individuals in a cohort acquire college education. Figure 1.3 illustrates this scenario. As skill prices increase vis-à-vis labor prices, the cut-off line moves so that more individuals with lower cognitive ability and/or higher labor endowment will enroll in college. The magnitude of the resulting increase in college population depends on the shape of the $(\ell, \kappa)$ distribution.

---

22Education is assumed to be instantaneous in this model: individuals pay $c^e$ and become a college graduate. Hence the terms “enrolling in college”, “acquiring college education”, and “obtaining a college degree” are used interchangeably in this article. In practice, however, education takes time and $c^e$ also reflects the opportunity cost of college attendance in the form of lost wages.
1.2.2.2 Labor Supply and Occupation Choice

**State Space and Timing** I now characterize the individual’s labor supply and occupation choice decision. After entering the economy, individuals have to choose in each period whether to work in the formal sector, and if so, in which occupation. The choice set facing an individual in each period is $C \equiv \{0: \text{Residual Sector}; 1: \text{Skill-intensive Occupation}; 2: \text{Labor-intensive Occupation}\}$.

The sequence of events in a given period unfolds as follows: at the beginning of the period, aggregate technology shocks are realized. Let $z_t \equiv \{A_t, \gamma_t, \alpha_t^H, \alpha_t^L, \lambda_t\}$ denote the state of production technology in period $t$. Then each individual draws an idiosyncratic utility shock $\epsilon_{it}$ associated with each choice $d \in C$, which she will receive if $d$ is chosen during that period. The vector of idiosyncratic shocks $\epsilon_{it} \equiv (\epsilon_{it}^0, \epsilon_{it}^1, \epsilon_{it}^2)$ is independently distributed in the population according to a gender-specific distribution $F^g(\cdot)$. After observing the state of aggregate technology and learning her own realized utility shocks, each individual proceeds to choose an alternative $d_{it} \in C$ for period $t$. Her state at the time of choice can be represented by $(x_{it}, \epsilon_{it})$, where $x_{it} \equiv \{\ell_i, \kappa_i, g_i, a_{it}, e_i, d_{i,t-1}\}$.

The collection of individual choices determine the labor force composition in a given period. After these choices are made, production takes place and workers receive their compensation. Equilibrium labor and skill prices depend on the set of all individual states and choices as well as on the state of aggregate technology $z_t$. Let $x_t \equiv (x_{1t}, \ldots, x_{N(t),t})$ and $d_t \equiv (d_{1t}, \ldots, d_{N(t),t})$ denote respectively the collection of individual states and choices in period $t$, then equilibrium wage $w_{jt} = w^j(x_t, d_t, z_t)$, $j \in \{H, L\}$.

---

23In the remaining article, the skill- and labor-intensive occupations are labeled either by $\{H, L\}$ or $\{1, 2\}$.

24Respectively, physical labor, cognitive ability, gender, age, education, and last period’s choice.

25$N(t)$ denotes the period $t$ population size.

26See equation (1.5). The idiosyncratic shocks $\epsilon_{it}$, $i = 1, \ldots, N(t)$, affect choice but do not enter the wage function.
**Flow Utilities** Individuals make labor supply decisions and choose their occupation by maximizing lifetime expected present discounted value of flow utilities. There is no saving or borrowing. Wage received in any given period is used to finance final goods consumption in that period only. Given the assumption, the individual solves the following problem

\[
\max_{\{d_{i\tau}\}} E_t \left[ \sum_{\tau=t}^{1+A-a_{i\tau}} \beta^{\tau-t} \pi_{\tau} (x_{\tau}, \epsilon_{i\tau}, d_{\tau}, z_{\tau}) \right]
\]

(1.7)

, where \( \pi_t \) is the per-period payoff function defined by

\[
\pi_t (x_t, \epsilon_{it}, d_t, z_t) = \left( u^{d_{it}} + w^{d_{it}} (x_t, d_t, z_t) \right) I_{d_{it} \in \{1,2\}} + u^0_t (x_{it}) I_{d_{it} = 0} + c^M (x_{it}, d_{it})
\]

, where \( u^d \) is the (dis)utility of working in occupation \( d \in \{1, 2\} \), \( u^0_t (x) \) is the flow utility of choosing the residual sector\(^{27}\), and \( c^M (x, d) \) is the mobility cost one incurs by choosing \( d \)\(^{28}\).

**Value of the Residual Sector** The utility flow generated by the residual sector is a sum of leisure consumption value, non-market home production, and other benefits and costs associated with not being employed in the formal sector. The value is assumed to be age and gender dependent and given by

\[
u^0_t (x_{it}) = \exp \left\{ \phi^0_{0it} + \phi_1 a_{it} + \phi_2 a^2_{it} \right\}
\]

(1.8)

, where \( \phi^0_{0it} \) is allowed to change over time.

**Mobility Cost** Individuals incur mobility costs when they enter the formal sector from the residual sector or when they switch occupations. Specifically,

\[
c^M (x_{it}, d_{it}) = \begin{cases} 
\phi^0_{01} & \text{if } d_{i,t-1}, d_{it} \in \{1, 2\} \text{ and } d_{i,t-1} \neq d_{it} \\
\phi^0_{02} & \text{if } d_{i,t-1} = 0 \text{ and } d_{it} \in \{1, 2\} \\
0 & \text{otherwise}
\end{cases}
\]

(1.9)

\(^{27}\)Note that \( u^0_t (.) \) depends on time. Detailed specification is given below.

\(^{28}\)Note that \( x \) contains information about last period’s choice. Detailed specification of the mobility cost function is given below.
The existence of mobility costs explains the inability of workers to arbitrage away wage differentials and captures, in reduced form, several sources of barriers to labor adjustment including occupation-specific human capital (Kambourov and Manovskii 2009), and psychological costs associated with switching.

**Expectations** To reduce the dimensionality of the problem in (1.7), I follow Lee and Wolpin (2006, 2010) and adopt an approximate rational expectations technology. Specifically, let \( r_t \equiv \left\{ r_t^L, j, g \right\}_{j=\{H,L\}, g=\{m,f\}} \) denote the vector of equilibrium labor and skill prices. I assume that the process governing \( r_t \) can be approximated by

\[
\log r_t = \eta_0 + \eta_1 t + \varepsilon_t
\]

(1.10)

, where

\[
\varepsilon_t = \eta_2 \varepsilon_{t-1} + \varsigma_t, \quad \varsigma_t \sim i.i.d. \mathcal{N}(0, \Omega_r)
\]

Let \( F_t^r (r_{t+1} | r_t; \eta) \) denote the process specified in (1.10). Individuals use \( F_t^r (r_{t+1} | r_t; \eta) \) to form expectations about future prices. In bounded rational expectations equilibrium, the parameter \( \eta \equiv (\eta_0, \eta_1, \eta_2) \) provides the best fit to the equilibrium path of \( r_t \) generated by the model.

**Bellman Equation** Given \( r_t \), the per-period payoff defined in (1.7) is a function of individual state and choice only and can be written as \( \pi_t (x_{it}, \epsilon_{it}, d_{it}, r_t) \). Given expectation technology \( F_r \), the Bellman equations for problem (1.7) are given by

\[
V_t (x_{it}, \epsilon_{it}, r_t) = \max_{d \in \mathcal{C}} V^d_t (x_{it}, \epsilon_{it}, r_t)
\]

(1.11)

, where

\[
V^d_t (x_{it}, \epsilon_{it}, r_t) = \pi_t (x_{it}, \epsilon_{it}, d, r_t) + \mathbb{I}_{\{a_{it} < A\}} \cdot \beta \mathbb{E}_t \left[ V_{t+1} (x_{i,t+1}, \epsilon_{i,t+1}, r_{t+1}) | x_{it}, r_t, d_{it} = d \right]
\]

This is because equilibrium wages are completely determined by \( x_{it} \) and \( r_t \) and \( (z_{it}, x_{-i,t}, d_{-i,t}) \) enter \( \pi_t \) only through the wage function.
\[ E_t [V_{t+1} (x_{i,t+1}, \epsilon_{i,t+1}, r_{t+1}) | x_{it}, r_t, d_{it}] = \int_{r_{t+1}} \int_{x_{i,t+1}} \sum_{x_{i,t+1}} V_{t+1} (x_{i,t+1}, \epsilon_{i,t+1}, r_{t+1}) \cdot f_x (x_{i,t+1} | x_{it}, d_{it}) \cdot dF^g_t (\epsilon_{i,t+1}) \cdot dF^r_t (r_{t+1} | r_t) \]

, where \( f_x (x_{i,t+1} | x_{it}, d_{it}) \) is the discrete state transition function.

Note that since the environment is non-stationary\(^{30}\), there is a different value function associated with each time period. To complete the description of individual’s problems, I link the value function, \( V_t \), with the expected value of entering the economy, \( W^e_t \):\(^{31}\)

\[ W^e_t (\ell_i, \kappa_i, g_i) = \int V_t (x_{it}, \epsilon_{it}, r_t) \cdot dF^g_t (\epsilon_{it}), \quad e = 0, 1 \quad (1.12) \]

, where \( x_{it} = (\ell_i, \kappa_i, g_i, a_{it} = 1, e_i = e, d_{i,t-1} = 0) \)\(^{32}\).

### 1.2.3 Exogenous Processes

The economy is nonstationary, subject to both exogenous supply and demand shifts. On the supply side, cohort sizes vary and values attached to the residual sector shift over time. On the demand side, aggregate productivity \( A_t \), occupation output share \( \gamma_t \), within-occupation factor shares \( \{\alpha_t^H, \alpha_t^L\} \), and gender weight \( \lambda_t \) are all allowed to change over time.

Cohort sizes are taken from data. Given data on aggregate labor income, \( \{Y_t\} \), and the assumed expectation technology, the model can also be solved without the need to recover \( \{A_t\} \). I characterize the other exogenous processes here.

**Changes in residual sector value** Values associated with the residual sector can shift over time as a result of a number of forces, including technological

\(^{30}\)See Section 1.2.3.

\(^{31}\)Since education is instantaneous in the model, individuals entering the economy in period \( t \) are assumed to have observed \( r_t \) when making their education decisions. This may create a potential bias as individuals, in practice, typically form their college enrollment decisions several years before they can graduate and enter the labor market.

\(^{32}\)I set the last-period activity to be the residual sector \( (d_{i,t-1} = 0) \) for new entrants into the economy. In addition, it is assumed that these individuals of model age 1 do not incur mobility costs when they choose to work in the formal sector.
progress in household production (Greenwood and Ananth 2005), the introduction of infant formula and the reduction in childbearing cost (Albanesi and Olivetti 2009), the innovation in contraceptive practices and the decline in marriage and fertility rates (Greenwood and Guner 2010), as well as the considerable shift in social norms regarding gender roles in housework and child care (Goldin 2006). To reflect the impact of these changes (in reduced form), the constant term in the residual sector utility function (1.8), \( \phi_{0t}^g \), is assumed to follow a linear trend, given by

\[
\phi_{0t}^g = \tilde{\phi}_{01}^g + \frac{1}{35} \left( \tilde{\phi}_{02}^g - \tilde{\phi}_{01}^g \right) (t - 1975), \quad g \in \{m, f\}
\]  

(1.13)

where \( \left( \tilde{\phi}_{01}^g, \tilde{\phi}_{02}^g \right) \) are respectively the value of \( \phi_{0t}^g \) in 1975 and 2010.

**Changes in occupation output share**  
Given the Cobb-Douglas specification of the aggregate production function (1.3), \( \gamma_t \) can be directly calculated from the skill-intensive occupation’s share of labor income in each year:\textsuperscript{33}

\[
\gamma_t = \frac{\text{Total wage \\& salary compensations of workers in skill-intensive occupation}}{\text{Total wage \\& salary compensations of employed workers}}
\]  

(1.14)

Figure 1.4 plots the imputed \( \{\gamma_t\} \) series along with selected components of GDP. As the figure shows, \( \gamma_t \) has risen considerably between 1975 and 2010. The increase comes almost entirely from rising aggregate wage of skill-intensive occupations, as the aggregate wage of labor-intensive occupations has remained stagnant in real terms during this period.

**Within-occupation factor-biased technological change**  
In addition to changing occupation output shares, I allow technological change within each occupation in the form of changing \( \alpha_{jt}^j \), \( j \in \{H, L\} \). Since \( \alpha_{jt}^j \) regulates the intensity in which skill is used, an increase in \( \alpha_{jt}^j \) will represent skill-biased technological change within occupation \( j \), while a decrease in \( \alpha_{jt}^j \) represents

\textsuperscript{33}In this article, I use the terms “labor income” and “wage \\& salary” interchangeably. In practice, labor income also includes bonuses and benefits (if any) and is typically greater than wage \\& salary income.
labor-biased technological change. I parametrize the process governing changes in $\alpha_j^t$ as
\[
\log \frac{\alpha_j^t}{1 - \alpha_j^t} = \tilde{\alpha}_0^j + \tilde{\alpha}_1^j t, \quad j \in \{H, L\}
\]
(1.15)
, where the pace and direction of factor-biased technological change is allowed to differ by occupation.

**Gender-biased Demand Shift**  Following Heathcote *et al.* (2010), I label changes in $\lambda_t$ as “gender-biased demand shift” and parametrize the process as
\[
\log \frac{\lambda_t}{1 - \lambda_t} = \tilde{\lambda}_0 + \tilde{\lambda}_1 t
\]
(1.16)

As Section 1.2.1 indicates, this shift could be driven by a number of forces, including antidiscrimination legislation\(^{34}\), increasing importance of social skills in the labor market (Borghans *et al.* 2006), and changes in social norms that make employers more willing to hire women (Goldin 2006)\(^{35}\).

### 1.2.4 Equilibrium

Let $\mathcal{N}(t, a, g)$ denote the size of the gender-$g$ cohort that enters the economy in time $t - a + 1$. The model equilibrium is defined as follows.

**Definition** An equilibrium for given paths of aggregate productivity $\{A_t\}$, occupation output shares $\{\gamma_t\}$, and cohort sizes $\{\mathcal{N}(t, a, g)\}$, is a sequence of aggregate output $\{Y_t\}$, intermediate output $\{Y^H_t, Y^L_t\}$, aggregate factor supply $\{S^j,g_t, L^j,g_t\}_{j\in\{H,L\},g\in\{m,f\}}$, intermediate goods prices $\{p^H_t, p^L_t\}$, and factor prices $\{r_t\}$\(^{36}\), satisfying

---

\(^{34}\)Important pieces of antidiscrimination legislation include the Fair Labor Standards Act of 1938, the Equal Pay Act of 1963 and the Lilly Ledbetter Fair Pay Act of 2009. The Paycheck Fairness Act, the latest legislation that seeks to address the gender income disparity, has been twice introduced and rejected by the United States Congress as of this writing.

\(^{35}\)In addition, since model age is used in this paper as a proxy for actual labor market experience (equation (1.2)), changes in $\lambda_t$ also reflect the changing gender gap in accrued labor market experiences.

\(^{36}\)Recall that $r_t \equiv \{r^L_t, r^S_t\}_{j\in\{H,L\},g\in\{m,f\}}$.
1. Individual education, labor supply, and occupation choice decisions are generated by choice probability functions that are solutions to the individual’s problems described in Section 1.2.2:

\[
\Pr (e_i = 1 | \ell_i, \kappa_i, g_i) = \int \mathcal{I}{\sigma^e_i(\ell_i, \kappa_i, g_i, c^e_i) = 1} d\mathcal{F}^g_{c^e_i}(c^e_i)
\]

\[
\Pr (d_{it} = d | x_{it}, r_{it}) = \int \mathcal{I}{\sigma^d_t(x_{it}, \epsilon_{it}, r_{it}) = d} d\mathcal{F}^{\epsilon}_{g}(\epsilon_{it})
\]

where \(\sigma^e_t(\ell_i, \kappa_i, g_i, c^e_i)\) and \(\sigma^d_t(x_{it}, \epsilon_{it}, r_{it})\) are policy functions associated respectively with the problems defined in (1.6) and (1.11).

2. Aggregate supply of skill and labor inputs in each occupation are given by

\[
S^{j,g}_t = \sum_{a=1}^{A} \sum_{i=1}^{N(t,a,g)} s_{it} \mathcal{I}_{a_{it} = a, g_i = g, d_{it} = j}, \quad j \in \{H, L\}, \ g \in \{m, f\}
\]

\[
L^{j,g}_t = \sum_{a=1}^{A} \sum_{i=1}^{N(t,a,g)} \ell_{it} \mathcal{I}_{a_{it} = a, g_i = g, d_{it} = j}, \quad j \in \{H, L\}, \ g \in \{m, f\}
\]

3. Aggregate product \(Y_t\) and intermediate products \(\{Y^H_t, Y^L_t\}\) are given by (1.3) and (1.4).

4. Goods market clears in each period

\[
p^j_t = \gamma_t Y^j_t / Y^j_t, \quad j \in \{H, L\}
\]

5. Labor market clears in each period

\[
r^{L,g,j}_t = p^j_t \frac{\partial Y^j_t}{\partial L^{j,g}_t}, \quad j \in \{H, L\}, \ g \in \{m, f\}
\]

\[
r^{S,g,j}_t = p^j_t \frac{\partial Y^j_t}{\partial S^{j,g}_t}, \quad j \in \{H, L\}, \ g \in \{m, f\}
\]

6. The expectation technology \(F_t^r (r_{t+1} | r_t; \eta)\) is consistent with the equilibrium path of \(\{r_t\}\).


## 1.3 Estimation

### 1.3.1 Distributional Assumptions

To estimate individual choices, further distributional assumptions about schooling cost $c_i^e$ and utility shocks $\epsilon_{it}$ are needed. I assume $c_i^e = \bar{c}^e + \xi_{it}$, where $\xi_{it}$ is Gumbel-distributed with gender-specific parameters $\nu_{gi}^e$:

$$
\xi_{it} \sim F_{\xi}^g(\xi) \equiv \exp\left\{ -e^{-\xi/\nu_{gi}^e} \right\}
$$

(1.17)

The distributional assumption implies that the variance of schooling cost $\text{Var}(c_i^e) = \frac{1}{6} \pi^2 (\nu_{gi}^e)^2$ and that the educational choice probability has a closed-form solution given by

$$
\Pr(e_i = 1|\ell_i, \kappa_i, g_i) = \frac{\exp \left\{ \left(W_t(\ell_i, \kappa_i, g_i, e_i = 1) - \bar{c}^e \right)/\nu_{gi}^e \right\}}{\exp \left\{ \left(W_t(\ell_i, \kappa_i, g_i, e_i = 1) - \bar{c}^e \right)/\nu_{gi}^e \right\} + \exp \left\{ \left(W_t(\ell_i, \kappa_i, g_i, e_i = 0) \right)/\nu_{gi}^e \right\}}.
$$

(1.18)

To estimate labor supply and occupation choice decisions, I assume the idiosyncratic utility shocks, $\epsilon_{it} \equiv (\epsilon_{it}^0, \epsilon_{it}^1, \epsilon_{it}^2)$, are distributed according to a generalized extreme value (GEV) distribution

$$
\epsilon_{it} \sim F_{\epsilon}^g(\epsilon) = \exp\left\{ -\left( e^{-\epsilon^0/\nu_{gi}^0} + \left( \sum_{j=1}^{2} e^{-\epsilon^j/\nu_{gi}^j} \right)^{\nu_{gi}^1/\nu_{gi}^0} \right) \right\}
$$

(1.19)

This is equivalent to a nested logit model in which individuals first choose whether to work in the formal sector or stay in the residual sector and then, conditional on choosing the formal sector, choose in which occupation to work. Figure 1.5 illustrates the choice model. Let $\tilde{V}_t^d(x_{it}, r_{it})$, $d \in C$ denote the deterministic part of $\tilde{V}_t^d(x_{it}, \epsilon_{it}, r_{it})$\footnote{See equation (1.11)}, then the choice probabilities are given

$$
\tilde{V}_t^d(x_{it}, r_{it}) = \tilde{\pi}_t(x_{it}, d, r_t) + \mathbb{I}_{\{a_{it} < A\}} \cdot \beta \mathbb{E}_t [V_{t+1}(x_{i,t+1}, \epsilon_{i,t+1}, r_{t+1})|x_{it}, r_t, d_{it} = d],
$$

where $\pi_t(x_{it}, \epsilon_{it}, d, r_t) \equiv \pi_t(x_{it}, d, r_t) + \epsilon_{it}^d$. 


by

$$\Pr (d_{it} = d | x_{it}, r_{it}, d_{it} \in \{1, 2\}) = \frac{\exp \left\{ \tilde{V}^d_t (x_{it}, r_{it}) / \nu_o^{git} \right\}}{\exp \left\{ \tilde{V}^1_t (x_{it}, r_{it}) / \nu_o^{git} \right\} + \exp \left\{ \tilde{V}^2_t (x_{it}, r_{it}) / \nu_o^{git} \right\}}, \quad d \in \{1, 2\}$$ (1.20)

and

$$\Pr (d_{it} = 0 | x_{it}, r_{it}) = \frac{\exp \left\{ \tilde{V}^0_t (x_{it}, r_{it}) / \nu_o^{git} \right\}}{\exp \left\{ \tilde{V}^0_t (x_{it}, r_{it}) / \nu_o^{git} \right\} + \exp \left\{ \tilde{W}_t (x_{it}, r_{it}) / \nu_o^{git} \right\}}$$ (1.21)

where

$$\tilde{W}_t (x_{it}, r_{it}) = \nu_o^{git} \log \left\{ \sum_{d=1}^2 \exp \left\{ \tilde{V}^d_t (x_{it}, r_{it}) / \nu_o^{git} \right\} \right\}$$

is the value associated with working in the formal sector. As (1.20) and (1.21) make clear, given the distributional assumption (1.19), the labor supply and occupation choice probabilities are regulated respectively by the variance parameters ($\nu_o^{git}$, $\nu_o^{git}$).

### 1.3.2 Estimation Method

The parameters in this model can be divided into three sets. Let $\theta_1$ describe the demand-side parameters, describing production technology and exogenous technology processes governing the evolution of factor shares and gender-specific demand\(^{38}\). Let $\theta_2$ denote the supply side parameters, describing \textit{ex ante} ability distributions, skill production, cost of schooling, value of the residual sector, cost of mobility, and the distribution of idiosyncratic shocks\(^{39}\). The third set is contained in $\eta$, which fits the dynamic evolution of equilibrium skill and labor prices to the expectation technology $F^t_t (r' | r; \eta)$.

The estimation procedure is adapted from Lee and Wolpin (2006) and involves the iteration of a model solution step and a parameter search step. I describe each of them below.

\(^{38}\)Specifically, $\theta_1$ includes the parameters reported in Tables 1.7

\(^{39}\)Specifically, $\theta_2$ includes the parameters reported in Tables 1.5, 1.8–1.13.
Model Solution  For each $\theta \equiv (\theta_1, \theta_2)$, I can solve for $\eta$ and for model equilibrium using observed aggregate labor income $\tilde{Y} \equiv \{Y_t\}$, cohort sizes $\tilde{N} \equiv \{N(t,a,g)\}$, and imputed occupation output shares $\tilde{\gamma} \equiv \{\gamma_t\}$. The solution procedure can be described as follows:

1. Given $\theta_1$, $\tilde{Y}$, $\tilde{\gamma}$, and the path of equilibrium prices $\tilde{r} \equiv \{r_t\}$, the demand function can be completely specified by $G^D \left( \tilde{r}, \tilde{Y}, \tilde{\gamma}, \theta_1 \right)$.

2. Given $\theta_2$, $\tilde{N}$, $\tilde{r}$, and $\eta$, the supply function is completely specified by $G^S \left( \tilde{r}, \tilde{N}, \theta_2, \eta \right)$.

3. The equilibrium path of factor prices $\tilde{r}$ can be solved by equating $G^D \left( \tilde{r}, \tilde{Y}, \tilde{\gamma}, \theta_1 \right)$ and $G^S \left( \tilde{r}, \tilde{N}, \theta_2, \eta \right)$.

4. $\eta$ is updated by re-fitting $\mathcal{F}_t^r \left( r_{t+1} | r_t; \eta \right)$ to the equilibrium path of $\tilde{r}$. Steps 3–4 are then repeated until $\eta$ converges. Denote the converged value by $\eta (\theta)$.

5. Once the model is solved given $\theta$, the aggregate productivity series $\{A_t\}$ can be calculated from equation (1.3) using $\tilde{Y}$ and $\tilde{\gamma}$.\footnote{Note that there is no need to recover $\{A_t\}$ for the purpose of model solution and estimation. However, $\{A_t\}$ are needed to perform counterfactual experiments in Section 1.4.}

The solution procedure requires repeatedly solving the individual’s problems and simulating the economy from 1975 – 2010 for each value of $\eta$ until its convergence to $\eta (\theta)$. The convergence of $\eta$ makes sure that individual expectations are consistent with the realized path of equilibrium prices. To simulate the economy, however, knowledge of the initial distribution of individual states in 1975 is needed. Since individual ability is unobserved by the econometrician, such a distribution cannot be obtained directly from data. To circumvent the initial conditions problem, I adopt the method used by Lee and Wolpin (2006) and start the economy at an earlier date (1940). The initial joint distribution of age, sex, education, and last period activity is obtained from the 1940 U.S. Census and then random values of cognitive ability and physical labor, drawn
from the ability distribution specified by $\theta$, are assigned to each individual in the simulated 1940 population. The economy is then simulated from 1940 to 2010. For each cohort, 400 individuals are generated. A cross-sectional population therefore contains 12,400 observations and the entire simulated sample contains 880,400 observations, spanning cohorts born as early as 1915 and as late as 1985\footnote{Individuals enter the economy at age 25. Therefore, the entering cohort in 1940 was born in 1915 and the entering cohort in 2010 was born in 1985.}

**Parameter Search** Estimation of $\theta$ is by simulated method of moments (SMM), where $\theta$ is chosen to minimize the distance between observed and simulated outcomes. Specifically, for each $\theta \in \Theta$, the model is solved using the solution procedure outlined above and a panel data set containing the simulated economy from 1975–2010 is created\footnote{As mentioned above, the economy is actually simulated from 1940 to 2010, but moments are calculated only for the 1975–2010 period.}. From the simulated data set, a set of moments can be calculated and compared with the corresponding sample moments. Section 1.6.4 documents the set of moments I use in this estimation. Formally, let $m^D$ and $m^S(\theta)$ denote respectively the vector of sample and simulated moments, then $\theta$ is chosen to solve the optimization problem

$$\arg\min_{\theta \in \Theta} Q(\theta)$$

, where $Q(\theta)$ is the weighted distance function defined by

$$Q(\theta) = (m^D - m^S(\theta))^\prime \Omega (m^D - m^S(\theta))$$

, where $\Omega$ is a positive definite weighting matrix. Section 1.6.5 provides more details on the calculation of the weighting matrix and the standard errors.

**1.3.3 Identification**

It may be helpful to review the strategy for identifying the various parameters of interest in this model. First, consider the skill production function. Identification of the function parameters follows from standard selection correction
arguments. Since self-selection is fully taken into account, following Heckman (1979), the parameters are identified by an exclusion restriction, namely the existence of variables that affect choices but not wages. The exclusion restriction here is the sector and occupation the individual is in during the previous period, which enters the utility function but not the wage equation. Fixing the wage equation parameters, the utility parameters are identified from choice probabilities conditional on individual characteristics. In particular, employment rates (choice probabilities unconditioned on last period sector and occupation) help pin down the residual sector value parameters, while transition rates (choice probabilities conditioned on last period sector and occupation) help pin down the mobility cost parameters. Next consider the production function. The model assumes that the demand for skill and labor inputs are stable up to an additively separable trend, which is a standard assumption in the literature. Given this assumption about the demand curve, the production function parameters are identified because exogenous changes in cohort sizes shift the supply curve but do not affect demand and serve as valid instrument for input levels.

The time-varying change in the gender wage gap provides the source of identification for the gender-biased demand shift process. The parameter measuring the difference between male and female physical labor endowment, $\chi_f$, is identified by the differential change in the gender wage gap within the skill-intensive and the labor-intensive occupation. The intuition is that when $\chi_f < 1$, gender comparative advantage interacts with within-occupation shift in skill demand to increase the female/male wage ratio. It can be proved that a reduction in $(1 - \alpha^j), j \in \{H, L\}$, share of physical labor in production, increases the relative wage of female workers. Furthermore, the higher $(1 - \alpha^j)$ is, or the smaller $\chi_f$ is, the greater the increase will be. Since production in the labor-intensive occupation uses a larger share of labor inputs, a marginal reduction in $(1 - \alpha^L)$ benefits the female workers in the labor-intensive occu-

\[\text{From the model's perspective, the gender weight ratio } \lambda_t/(1 - \lambda_t) \text{ is equivalent to a parameter in the wage equation measuring differences in female and male productivity and therefore is identified using the same selection correction arguments for the identification of the skill production function.}\]
pation more than a marginal reduction in \((1 - \alpha^H)\) benefits the female workers in the skill-intensive occupation. The difference in the resulting gender wage convergence rates reflects the magnitude of \(\chi_f\). Figure 1.6 plots the log female/male wage ratio by occupation from 1975 to 2010 and shows that, as expected, the gender gap closes faster in the labor-intensive occupation\(^{44}\).

**Ability Distribution and Idiosyncratic Shocks** I now discuss how the ability distribution and the idiosyncratic shock parameters are identified. Essentially, observing variations in aggregate choice patterns across populations subject to the same idiosyncratic shock variances but with different ability compositions help pin down the parameters of the ability distribution, while observing populations with the same ability composition but subject to different idiosyncratic shock variances help pin down the variance parameters\(^{45}\).

To see this more clearly, first, it should be noted that both the distribution of *ex ante* ability and of idiosyncratic shocks play a role in regulating how aggregate choice patterns change over time. Consider the college enrollment decision in (1.20) and consider a simplified version of the model in which individuals differ only by cognitive ability \(\kappa\). Then the equation be written as

\[
\Pr (e_i = 1 | \kappa_i) = \frac{1}{1 + \exp \left(- \frac{\Delta W_t (\kappa_i) - \bar{c} \nu_e}{\nu_e}\right)}
\]

(1.23)

, where \(\Delta W_t (\kappa_i) \equiv W_t (\kappa_i, e_i = 1) - W_t (\kappa_i, e_i = 0)\) is the relative payoff of entering the labor market with a college degree. Aggregate enrollment rate,

\(^{44}\)Note that the figure plots the raw gender gap. Regressions exercises show that after controlling for age and education, the residual gender gap likewise closes faster in the labor-intensive occupation, but consistently remains larger than the gap in the skill-intensive occupation. This is what should be expected, since the labor-intensive occupation involves the use of more labor inputs in production, to the disadvantage of women. *Ceteris paribus*, the gender gap in the occupation should be larger, despite faster convergence.

\(^{45}\)For example, the male population in successive entering cohorts share the same ability distribution. Thus, the elasticity of male college enrollment in response to changes in the college premium identifies \(\nu^m_e\). The populations of college-educated and non-college-educated men are subject to the same idiosyncratic shock variance \(\nu^m_w\). Thus, differences in their labor supply elasticities as well as absolute choice percentages reflect, in addition to differences in payoffs, the difference in their ability composition and help to identify the parameters of the ability distribution.
\[ \Pr(e_i = 1), \text{obtains from integration over } i: \]

\[ \Pr(e_i = 1) = \mathbb{E}_{\kappa_i}[\Pr(e_i = 1|\kappa_i)] = \int \Pr(e_i = 1|\kappa_i) d\mathcal{F}_\kappa(\kappa_i) \quad (1.24) \]

It should be clear from observing (1.23) and (1.24) that both \( \nu_e \) and \( \mathcal{F}_\kappa(.) \) determine the responsiveness of the aggregate enrollment rate to changes in the relative payoff of college education. The larger \( \nu_e \) is, the less responsive each individual \( i \) is to changes in \( \Delta W \). This is easy to understand, as a rise in \( \nu^e \) flattens the tails of the education cost distribution, increasing the probability that a given individual would choose to enroll in college if \( \Delta W - \bar{c}^e \) is negative and the probability that she would choose not to do so if \( \Delta W - \bar{c}^e \) is positive (Figure 1.7a). Viewed differently, a high value of \( \nu^e \) implies that psychological and pecuniary factors that affect idiosyncratic college attendance costs are more dominant in enrollment decisions, so that individuals do not pay as much attention to the relative payoff of college degrees.

On the other hand, \( \mathcal{F}_\kappa(.) \) determines how changes in individual enrollment probabilities translate into changes in the aggregate enrollment rate. Greater dispersion in \( \mathcal{F}_\kappa(.) \), for example, will translate into an aggregate enrollment rate that is less responsive to changes in \( \Delta W \). This is easy to understand, as a more dispersed \( \mathcal{F}_\kappa(.) \) increases the probability that a given individual would have a \( \Delta W \) greater than a given \( c \) when \( c \) is large and smaller than a given \( c \) when \( c \) is small (Figure 1.7b). In the case where \( \nu^e = 0 \), the aggregate enrollment rate simply becomes \( \Pr(e_i = 1) = \Pr(\kappa_i > \Delta W^{-1}_i(\bar{c}^e)) = 1 - \mathcal{F}_\kappa(\Delta W^{-1}_i(\bar{c}^e)), \) which is equivalent to the complementary CDF function of \( \kappa \) when \( \Delta W^{-1}(\bar{c}^e) \) goes from \( \bar{\kappa} \) to \( \underline{\kappa} \). Figure 1.8 plots, for example, the complementary Beta(5,5) and uniform CDFs and one can see that the uniform distribution, being a mean-preserving spread of Beta(5,5), produces a flatter and more inelastic “enrollment rate curve.” It is also worth noting that when \( \nu^e = 0 \), i.e. when there is no uncertainty in individual choice, the ability distribution completely determines how the aggregate enrollment rate changes in response to shifts in \( \Delta W \).\(^{46}\)

\(^{46}\)See Figure 1.3 as well as more discussions in Section 1.2.2.
More generally, both the ability distribution $F(\ell, \kappa)$ and the idiosyncratic shock variances $\nu \equiv \{\nu_\ell^g, \nu_w^g, \nu_\phi^g\}$ regulate the elasticity of aggregate response to changing present value of choices. Higher values of $\nu$ and more dispersion in $F(\ell, \kappa)$ lead to lower elasticities of response and a flatter supply curve, while lower values of $\nu$ and a more concentrated $F(\ell, \kappa)$ lead to higher elasticities of response and a steeper supply curve. Figure 1.9 illustrates this intuition. Observing variations in aggregate choice patterns across time and across populations with different ability compositions or subject to different idiosyncratic shock variances together identify the parameters of $\nu$ and $F(\ell, \kappa)$.

1.4 Results

1.4.1 Parameter Estimates

Estimation results are presented in Tables 1.5–1.13. A number of normalizations are needed to estimate the model. The nonpecuniary benefit associated with employment in the skill-intensive occupation is normalized to zero (for both males and females). Thus, the nonpecuniary benefit of working in the labor-intensive occupation as well as the value of the residual sector are relative to this normalization. Since skill and physical labor are not directly observable, but must be inferred from wages, their levels cannot be disentangled from the level of equilibrium skill and labor prices. Therefore, the upper bounds of both ability distributions are set to $\ell = \pi = 1$.

Table 1.5 reports the estimated ability distribution parameters. The cognitive ability distribution, common to both sexes, is fairly spread out and slightly positively skewed. The physical labor distribution, on the other hand, is highly concentrated and negatively skewed. Figure 1.10 plots the male ability distribution. The female cognitive ability distribution is by assumption the same as that of men, while the female physical labor distribution is $\tilde{f}_\ell(\ell|\chi_\ell) = f_\ell(\ell|\chi_\ell)/\chi_\ell$, where $f_\ell(\cdot)$ is the distribution of physical labor in men. The parameter $\chi_\ell$ is estimated at 0.41, suggesting a significant gender
comparative advantage\footnote{The results partially agree with calibration results from Guvenen and Kuruscu (2010b) and Rendall (2010). Guvenen and Kuruscu (2010b) calibrates the distribution of “raw labor” and “ability to accumulate human capital” among men and find that the coefficient of variance in “ability to accumulate human capital” is almost five times that in “raw labor.” Rendall (2010) conducts a similar calibration exercise to determine the distributions of “brain” and “brawn” for both men and women, and find that men have about twice as much brawn than women. However, contrary to Guvenen and Kuruscu (2010b) and to the results of this paper, Rendall (2010) also finds that the variance in “brawn” is considerable higher than that in “brain”. In both Guvenen and Kuruscu (2010b) and Rendall (2010), the authors assume specific distributional shapes (uniform and normal) and calibrate the distribution parameters from wage data, which differs from both the assumption and the identification strategy of this paper.}

Table 1.6 displays the mean and variance of the ability distributions by gender. The result that population variation in cognitive ability is substantially larger than that in physical labor is consistent with microeconomic evidence from Huggett et al. (2006), who find that differences in “learning ability” are essential to produce the observed increase in earnings dispersion over the life cycle and account for the bulk of the variation in the present value of earnings across agents. Overall, heterogeneity in cognitive ability has a much more significant effect on labor market outcomes than heterogeneity in labor.

Table 1.7 reports the estimated production function parameters. In both occupations, skill and labor are gross complements. The elasticities of substitution are 0.64 in the skill-intensive occupation and 0.45 in the labor-intensive occupation\footnote{The elasticity of substitution in occupation $j, j \in \{L, H\}$ is $\sigma_j = 1/ (1 - \rho^j)$.}. There are skill-biased technological changes within both occupations, with the ratio of skill-to-labor share rising faster in the skill-intensive occupation than in the labor-intensive occupation. The estimates also reveal the presence of significant gender-biased demand shift over time. The relative weight of female inputs used in production increases from 0.74 in 1975 to 0.91 in 2010, suggesting a continuing convergence in the demand for gender-specific inputs.

The estimated residual sector parameter values are presented in Table 1.10.\footnote{Although, to my knowledge, no comparable estimates exist in the literature, Jeong et al. (2008) have estimated a model with an aggregate production function that features complementarity between stocks of labor and experience and find the two to be gross complements with the elasticity of substitution equal to 0.3.}
Female values fell during the period of 1975 – 2010, while male values rose. For both sexes, the values people attach to the residual sector increase with age. Mobility costs are substantial, but do not differ significantly by gender (Table 1.11). The costs of entering employment from the residual sector and the costs of switching occupations are about the same. The estimated cost of attending college, in 2005 dollars, is $125,000 for men and $120,858 for women during the 1975 – 2010 period (Table 1.12). The standard deviation of idiosyncratic shocks associated with education, labor supply and occupation decisions are given in Table 1.13. The standard deviation of shocks associated with the labor supply decision is greater than that associated with occupational choice, reflecting more frequent transitions in and out of the residual sector. Females have smaller standard deviation of shocks associated with all three choices, implying that, everything else being equal, women are more elastic in their college enrollment, labor supply, and occupation decisions. On the other hand, larger variances of shocks among men means that idiosyncratic and nonpecuniary factors are more important, so that male education, labor supply and occupation decisions are less responsive to changing labor market returns.

1.4.2 Model-Data Comparison

In this section, I consider the model’s ability to generate the salient trends observed in the U.S. labor market over the last thirty five years.

Relative Wages Figure 1.11 shows the model’s prediction on relative wages. Figure 1.11a plots the evolution of the actual and predicted college premium,

---

50 The estimates are generally higher than the mobility cost estimates in Lee and Wolpin (2006, 2010). One reason is that Lee and Wolpin account for sector-occupation-specific experience, while this paper does not.

51 Lee (2005) estimates the cost of attending college to be $7908 per year in 1983 dollars, which is about $15,000 in 2005 dollars. Lee and Wolpin (2006) estimate the cost to be $18,664 per year in 1983 dollars, which is about $36,000 in 2005 dollars. In Lee and Wolpin (2010), the cost is estimated to be $14,494 per year, in 1983 dollars, for the first two years of college, and $26,688 per year for the next two years, which amounts respectively to about $28,000 and $52,344 in 2005 dollars.
defined as the log of college/non-college wage ratio. As the figure shows, the actual college premium declined between 1975 and 1980, and then increased significantly between 1980 and 2006. The recent decline in the college premium since 2007 reflects the large impact of the economic recession. The model is able to capture both the level of and the growth in the college premium during the majority of this period, although it does not generate the initial decline in the 1970s or the recent drop in the college premium as a result of the recession.

Figure 1.11b plots the evolution of the actual and predicted gender gap, measured by the female/male wage ratio. As the figure shows, the actual female/male wage ratio has been increasing since the late 1970s. The model does well in capturing this significant increase, although, incongruent to data, it also predicts a steep rise in the late 1970s, and understates the growth in the wage ratio toward the end of the period. By 2010, the gender gap in the model is about 5 percent larger than that observed in the data.

Education and Labor Supply  Figure 1.12 shows the model’s prediction on education attainment and labor supply behavior. Figure 1.12a compares the observed changes in the college attainment rate among 25 – 29 years olds with the prediction by the model. There has been a rapid increase in the actual college attainment rate among females, from less than 20 percent in 1975 to more than 35 percent in 2010. The male college attainment rate, on the other hand, declined in the late 1970s, remained stagnant throughout the 1980s and early 1990s, and then increased about 5 percentage points between 1990 and 2000.
1994 and 2000, and had been stagnant during the 2000s. Thus, the short period of rise in the late 1990s essentially captures the entire gain in the male college attainment rate among entering cohorts between 1975 and 2010. The pattern generated by the model is broadly consistent. In particular, the model generates increases in male and female college attainment rates that match the magnitude of the observed increases. Trend deviations, however, are matched less well. The model does not capture the decline in the male college attainment rate in the late 1970s or its stagnation during much of the sample period.

Figure 1.12b plots the actual and predicted employment rates by gender. The male employment rate stays relatively constant during the majority of the period, fluctuating around 80 percent in the model, as in the data. The actual male employment rate fell precipitously after 2008 as a result of the economic recession, which the model does not capture. Overall, the male employment rate in the model experiences a slight decline of about 3 percentage points over the entire period. The actual female employment rate rose substantially, from 35 percent in 1975 to almost 60 percent in 2000. The rate became stagnant during the 2000s and, just like male employment, suffered a drop after 2008 as a consequence of the recession. The model prediction closely matches the data between 1975 and 2000, but does not quite capture the slow-down in the growth of female employment rate during the 2000s or its drop during the recession.

Wage Inequality  Figure 1.13 examines the model’s implications on overall and within-group wage inequality. Figure 1.13a and 1.13b compare, respectively, the actual and predicted growth in overall and within-group log wage variances since 1975. Both the model and data series are expressed in terms of percentage changes relative to the 1975 value. In terms of trends, the model predicts continuously rising overall variance, as well as rising within-group variance among both college- and non-college-educated individuals, which is qualitatively consistent with the empirical pattern. In particular, in both model and data, the percentage increase in log wage variance among college-educated
individuals is greater than that among non-college-educated individuals.

Although the model fits the data qualitatively, quantitatively it does not match well the absolute level of or the degree of growth in the observed overall and within-group log wage variances. Between 1975 and 2010, overall variance increased by 42 percent in the data, but only by 20 percent in the model. The model also underestimates the degree of growth in the variance among college-educated individuals over much of the sample period, while slightly over-predicts that among non-college-educated individuals. In terms of absolute values, overall log wage variance is 0.056 in the model in 1975, compared to 0.275 in the data. The variance among college-educated individuals is 0.04 in the model in 1975 compared to 0.255 in the data, and that among non-college-educated individuals is 0.034 in the model in 1975 compared to 0.248 in the data. Thus, although the model is able to generate trends in overall and within-group inequality that are qualitatively consistent, this exercise shows that more is at play in determining the changes in the variance of the U.S. wage structure.

1.4.3 The Changing Composition of College Graduates and the U.S. Workforce

By explicitly modeling individual ability and accounting for selection, the model allows me to “see” how the ability composition of college graduates and of the workforce has changed over time. Figure 1.14 plots the evolution the mean cognitive ability of college graduates in entering cohorts. As the figure shows, the cognitive ability of female college graduates in entering cohorts increased between 1975 and 1986, and has been decreasing since. The result implies that the degree of women’s selectivity into college first increased then decreased with increasing female enrollment. To see this more clearly, Figure 1.15 plots the female ability distribution in selected entering cohorts. The first observation is that in any cohort, individuals with higher cognitive ability and lower physical labor are more likely to self-select into college. In 1975, the cognitive abilities of female college graduates, however, are relatively dispersed.
As discussed, this is because when the relative payoff of college is low, individuals are more influenced by their idiosyncratic costs when making enrollment decisions. By 1985, entry into college has become more sorted along cognitive ability, indicating that the majority of new entrants between 1975 and 1985 comes from high cognitive ability individuals. As female enrollment continues to increase in each cohort, the marginal entrant then becomes more and more likely to come from low cognitive ability ones. Figure 1.15c–1.15d illustrate this process. By 2010, 35 percent of the females in the entering cohort have a college degree and their abilities are widely distributed.

Figure 1.14 also shows that the cognitive ability of male college graduates stays relatively constant, with perhaps a slight decline between 1975 and 2005. The overall cognitive ability of college graduates follows a trend similar to that of female’s, increasing from 1975 to around 1986 and decreasing afterward. Figure 1.16 plots the male ability distribution in the entering cohorts of 1975 and 2010 and it can be seen that although there is a slight increase in male enrollment during this time, the distribution of abilities among male college graduates in entering cohorts has not changed much over this period of time.

Since the population is composed of cohorts that make their education decisions at different times, the population trend in the average ability of college graduates can differ markedly from the trend observed in entering cohorts, as the education decisions of older cohorts reflect the economic and social realities of earlier times. This gives rise to hysteresis in population dynamics. Figure 1.17 plots the population average of cognitive ability and physical labor among college graduates from 1975 to 2010. The figure shows that the cognitive ability of female college graduates in the population has been increasing until around 2000, after which it has slightly declined. The cognitive ability of male college graduates in the population has been increasing until around 1995 and has been declining since then. This implies that for females, from 1975 to 2000, the average cognitive ability of college graduates in entering cohorts is likely higher than that in retiring cohorts, while the reverse is true.

\footnote{In 2010, for example, the oldest cohort in the model population was born in 1950, whose college enrollment decisions were made before they entered the model in 1975.}
after 2000. The same can be said about male college graduates in the periods of 1975 – 1995 and 1995 – 2010. These findings are consistent with the study by Taubman and Wales (1972), who, by examining test scores, report that the average aptitude of those who attended college rose during the first half of the 20th century relative to those who did not attend college. On the other hand, the results do not support Carneiro and Lee (2011)’s finding that increases in college attainment has lead to a significant decline in the average quality of male college graduates between 1960 and 2000. Figure 1.17 also shows that the population average of physical labor among college graduates has been decreasing throughout the period. Since the dispersion in individual labor endowment is small, the decrease is almost entirely attributable to increased female share of the college-educated population. Finally, Figure 1.18 shows that the cognitive ability of college graduates relative to non-college-educated individuals in the population has been increasing throughout the period.

To examine the changes in the ability composition of the workforce, Figure 1.19 plots the mean cognitive ability and physical labor from 1975 to 2010. It shows that both male and female cognitive ability has been increasing in the workforce throughout the period, with female cognitive ability increasing faster than male’s. The result suggests that both female and male employment growth during this period have been positively selected along the dimension of cognitive ability. This is consistent with the findings of Juhn and Murphy (1997), who document that employment gains for married women over the 1970s and 1980s were largest for wives of middle- and high-wage men who themselves tended to be more skilled. Like Juhn and Murphy (1997), Blau and Kahn (2007) find that female labor force growth during the 1980s was positively selected, but conclude that the nature of female labor force growth.

Note that while the average cognitive ability of college graduates may be increasing or decreasing with enrollment sizes, the average cognitive ability of the non-college-educated group always declines as a result of increasing college enrollment.

Individuals of higher cognitive ability have been entering the workforce, while those of lower cognitive abilities have been leaving. Since female employment has grown tremendously during the period, the result suggests that the grown mainly involves women of higher cognitive ability entering the workforce. Since male employment has stayed relatively constant, it suggests that men of lower cognitive abilities have been leaving.
changed during the 1990s to that of negative selection. I do not find the same change here. The result also lends partial support to Mulligan and Rubinstein (2008), who argue that selection into labor force participation explains a major part of the narrowing of the gender gap, due to women with less skill dropping out of the workforce, and those with more entering. The results in this paper indicate that while the cognitive ability of the female workforce has indeed been increasing, it is only one of a number of contributing mechanisms toward the closing of the gender wage gap. Finally, Figure 1.19b shows that the physical labor endowment of the workforce has been decreasing throughout the period, due almost entirely to the increasing female share of the workforce.

1.4.4 Accounting for Labor Market Changes

The dynamics predicted by the model and documented in Section 1.4.2 are primarily driven by the following exogenous processes: (i) Hicks-neutral technological growth; (ii) Skill-biased technological change; (iii) Gender-biased demand shift; (iv) Changes in residual sector value.

In this section, I assess, through the lens of the estimated model, the relative importance of each of these four channels in shaping the observed changes in the U.S. labor market over the last 35 years. To do so, I fix Hicks-neutral technology, $A_t$, the occupational share of labor income, $\gamma_t$, the factor shares, $\{\alpha_H^t, \alpha_L^t\}$, the gender weights, $\lambda_t$, and the value of the residual sector, $u_0^t$, at their 1975 values. Then I allow each of the four exogenous processes one at a time. Figures 1.20–1.21 show the results of these counterfactual experiments for equilibrium outcomes over the 1975–2010 period. To focus on changes over time, all results are presented as changes from their 1975 values.

---

59 Another factor that contributes to the non-stationary environment is changing cohort sizes.

60 I define skill-biased technological change as a combination of increasing output share of the skill-intensive occupation and within-occupation shifts toward skill inputs.

61 Throughout the discussions in this section and the next, it is assumed in every counterfactual experiment that gender-specific demand remains exogenous. As discussed in Section 1.2.1, however, the shift in gender-specific demand reflects in part changes in gender-specific labor market experiences, which are likely to be different in different experiments. This is
College Attainment and the College Premium  Figure 1.21a shows the impact of exogenous processes on education attainment. All four processes contribute to the rise in female college attainment, while countervailing forces govern the evolution of male college attainment. Skill-biased technological change, by raising the demand for skill inputs and for skill-intensive jobs, raises the return to skill and attracts both men and women into college. Hicks-neutral technological growth raises the return to college education relative to the college attendance cost, assumed to be constant in real terms, and likewise induces both men and women into college. On the other hand, gender-biased demand shift increases the relative weight of female inputs in production and therefore raises the female return to college relative to men’s, which encourages more women to obtain college education and depresses the male incentive to do so. Similarly, since the estimates show that residual sector value is increasing for men and decreasing for women, it has the effect of encouraging more women to go to college, while keeping more men at home, diminishing their incentives to pursue college education. Dynamics also play an important role, as the large entry of women into college depresses the growth in male college premium and adds to the unfavorable demand and supply shifts to further limit the growth in male college attainment.

Figure 1.20a considers the determinants of changes in the college wage premium. The experiments reveal several off-setting forces. Both skill-biased and Hicks-neutral technological changes drive up the increase in the college premium. In particular, skill-biased technological change by itself would have led to a 128% increase in the college premium over the time period, compared to the 81% increase predicted by the model. Hicks-neutral technological growth has a large positive impact on the college premium early on, but its effect becomes muted especially after the mid-1980s. On the other hand, gender-biased demand shift and changing residual sector values both lower the college premium over time. The intuition for these results stems from the direct impact these processes have on college and non-college wages and the indirect effects

\[62\] Clearly, these effects are not additive.
they have through their differential impact on college attainment. Skill-biased technological change, for example, directly raises the college premium by raising the relative price of skill and by raising the demand for skill-intensive jobs which college graduates have a comparative advantage in. Hicks-neutral technological growth, by raising the return to work and increasing the wage differentials between occupations, increases the sorting of college- and non-college-educated individuals by employment and occupational skill-intensity, thereby raising the college premium. Gender-biased demand shift, on the other hand, by increasing female wages and depressing male wages, lowers the college premium, as there are more college-educated men than women in the population. Meanwhile, as Figure 1.21a shows, all four processes lead to an increase in college attainment and thus have an indirect effect in depressing the wage premium. As the decomposition result shows, the direct effect of skill-biased technological change in increasing the college premium dominates the indirect effect of its induced supply response, while the direct effect of Hicks-neutral technological growth dominates early on but is neutralized by its indirect effect later. Both the direct effect and the indirect effect of gender-biased demand shift work in the same direction and therefore have the largest impact on decreasing the college premium.

**Labor Supply and the Gender Wage Gap** Figure 1.21b examines the relative importance of each exogenous process in contributing to the increase in female/male employment ratio over the last 35 years. As the figure shows, all four processes contribute in various degrees to the observed increase, of which gender-biased demand shift and changing residual sector value have the largest impact. Since gender-biased demand shift directly increases female wages at the expense of male wages, the process encourages women to enter the labor market from the residual sector and discourages men to do so. The fall in women’s residual sector value that is estimated to have occurred likewise drives women to leave home and become a full-time employee in the workforce. The

---

This is, in a nutshell, the thesis of the Tinbergen race literature that the growth in the demand for skill has outpaced the growth in the supply of skill to continually increase the college premium and wage inequality over the last century (Goldin and Katz 2009).
rise in men’s residual sector value, to the contrary, encourages more men to stay home and forgo labor market earnings. Skill-biased technological change also contributes. By increasing the share of skill-intensive jobs in production and raising the relative price of skill, the process encourages men and women with high skill-to-labor ratios to enter the labor market and those with high labor-to-skill ratios to exit. As discussed, given the presence of gender comparative advantage, the process itself has a gender-biased nature: more women would join the workforce due to their relative intensity in cognitive ability and more men would exit as a result of their relative abundance in physical. Finally, Hicks-neutral technological growth, by increasing the returns to labor market participation relative to staying home, induces both men and women to join the workforce. The result of the experiment shows that its effect on female labor supply dominates its effect on male labor supply so that it contributes on net to increasing the female/male employment ratio.

Figure 1.20b considers the determinants of changes in the gender wage gap. As the experiments reveal, all four processes also help to close the gap between male and female wages. Gender-biased demand shift has the largest impact and by itself would have led to a 22% growth in female/male wage ratio between 1975 and 2010, compared to the 32% growth generated by the model. This is not surprising as gender-biased demand shift direly closes the gender gap in production. Moreover, as the results in Figure 1.21a show, it also helps to raise female college attainment, while discouraging male college attainment, and thus has an indirect effect on closing the gender gap through the induced education response. Like gender-biased demand shift, skill-biased technological change works in combination with gender comparative advantage to directly close the gender gap. It also induces an education response: as Figure 1.21a shows, the process increases both male and female college attainment rates, but has a larger effect on women than men, and thus contribute on net to decreasing gender wage disparities. The same indirect effect through induced education response is present in both Hicks-neutral technolog-

---

64 The result should not be surprising given the higher elasticity of female labor supply (lower variance in \( \nu_w \)) estimated by the model.
logical growth and changes in residual sector values, the latter having a larger impact as it increases female college attainment, while decreasing the supply of male college graduates. Note that as Figure 1.21b shows, all four processes also lead to a labor supply response that results in an increase in the female/male employment ratio. However, the results here indicate that they do not negate the positive effects these processes generate on the relative wage of females. In particular, the indirect effects are only possible in a dynamic model where women increase their education investment in anticipation of higher earnings and increased labor market participation, so that an increase in the supply of women to the workforce can raise their relative wages.

1.4.5 Ability Distribution and Labor Market Outcomes

In this section, I assess the importance of ability distributions and gender comparative advantage in affecting the equilibrium outcomes of the economy. To do this, I perform counterfactual experiments assuming: (i) both physical labor and cognitive ability are uniformly distributed; (ii) all individuals of the same gender are endowed with the same physical labor and cognitive ability; (3) males and females have the same physical labor distribution. Figures 1.22–1.23 show the results of these experiments for equilibrium outcomes over the 1975–2010 period. I compare them with the results obtained from the estimated model, which serves as the benchmark in the following discussions.

College Attainment and the College Premium

Figure 1.23a shows the evolution of college attainment rates among 25–29 years olds under the three

65Specifically, I assume that, for men, \((\ell_i, \kappa_i)\) are uniformly distributed on the intervals of \([\tilde{\ell}, 1]\) and \([\tilde{\kappa}, 1]\), while for women, they are uniformly distributed on the intervals of \([\chi \ell, \chi]\) and \([\chi, 1]\).

66Specifically, I assume that \((\ell_i, \kappa_i) = (1, 1)\) for all men and \((\chi_i, 1)\) for all women.

67I.e., \(\chi = 1\).

68Since the economy is simulated from 1940 to 2010, taking the individual states in 1940 as given, the experiments amount to assuming a sudden change in the distribution of ex ante abilities in 1940. This is a limitation of the exercise that cannot be circumvent.
counterfactual scenarios. When abilities are uniformly distributed, more men and women between the ages of 25 and 29 obtain college education compared to benchmark, although the rate of growth in college attainment is lower for both sexes. The opposite is true when the ability distributions are degenerate: fewer men and women obtain college education, but the attainment rates rise faster with time. The intuition for these results stems from the effect of ability distributions on aggregate choice elasticities. As discussed in Section 1.3.3, more dispersion in abilities lead to lower elasticities of aggregate response to changing values and a flatter supply curve. Thus, under uniform ability distributions, more men and women obtain college education early on, while aggregate attainment rates rise more slowly with increasing returns to college. The opposite is true when there is no dispersion in abilities.

When there is no gender comparative advantage in ability endowments, Figure 1.23a shows that their college attainment rate is lower and grows more slowly relative to benchmark. Skill-biased technological change now has less effect on the relative returns to college for women and more women would choose to work in the labor-intensive occupation, diminishing their incentive to acquire college education. With less female competition in higher education and in skill-intensive jobs and more female competition in labor-intensive jobs, men, instead, have more incentives to go to college. Thus, as the result shows, their college attainment rate is higher and grows faster relative to benchmark.

Figure 1.22a considers the implications of different ability distributions on the evolution of the college premium. When abilities are uniformly distributed, the premium is higher in comparison with benchmark. When they are homogeneous, the premium is lower and grows relatively slower. To understand the results, note that under homogeneous abilities, the average cognitive ability of college graduates is the same as that of non-graduates. The premium is therefore lower due to the absence of ability differentials. Since the differential in cognitive ability between college and non-college graduates is growing over

---

69 This is because there are more high cognitive ability individuals when the ability is uniformly distributed.
70 among 25 – 29 years old.
71 among 25 – 29 years old.
time in the benchmark model (Figure 1.18), the difference between the college premium under homogenous abilities and benchmark becomes larger over time. On the other hand, the level of college attainment in the population is lower under homogeneous abilities and increasingly so over the sample period relative to benchmark\(^{72}\), which should lead to faster growth in the premium. Similarly, under uniform ability distributions, the cognitive ability differential between college and non-college graduates is larger than in the benchmark model\(^{73}\), while the level of college attainment in the population is higher\(^{74}\). The predicted college premium behavior results from these countervailing forces.

When there is no gender comparative advantage in ability endowments, the college premium is both higher and rises faster compared to benchmark. One reason, as Figure 1.23a suggests, is that the female college enrollment response is significantly weaker in this case, which helps to drive up the premium. On the other hand, increased supply of physical labor and more entry of women into the labor-intensive occupation drive down labor price and raise the relative price of skill, thereby also contributing to a higher premium.

**Labor Supply and the Gender Wage Gap** Figure 1.23 looks at the female/male employment ratio under these counterfactual scenarios. When abilities are uniformly distributed, the employment ratio is higher in 1975, but grows more slowly and is at about the same level in 2010 as the ratio predicted by the benchmark model. When abilities are homogeneous, the employment

\(^{72}\)See Figure 1.23a, which plots the college attainment rates among 25 – 29 year old individuals. Under homogeneous abilities, although these entry cohort rates rise faster than they do in the benchmark model, the fact that they are consistently lower in level means that the population college attainment rate is increasingly lower relative to benchmark. Their increase in difference will last until college attainment in entry cohorts under homogeneous abilities catch up with that in the benchmark model. Similarly, under uniform ability distributions, although attainment rates in entry cohorts rise more slowly, their higher levels translate into increasingly higher population rates relative to benchmark. See also the discussion in Section 1.4.3 on the difference between entry cohort and population dynamics.

\(^{73}\)This is because there are more high cognitive ability college graduates and more low cognitive ability non-college graduates when the ability is uniformly distributed. Selection effects are more pronounced when there is more dispersion in abilities.

\(^{74}\)See Figure 1.23a. Higher levels of college attainment in entering cohorts translate into increasingly higher population rates relative to benchmark, which depresses the level and growth of the college premium.
ratio is initially lower but grows significantly faster and, by 2010, is about 4 points higher than benchmark. The explanation for these results comes from the direct influence ability distributions have on male and female employment and the indirect effects they have through their differential impact on college attainment. When abilities are uniformly distributed, more women and fewer men would enter the workforce early on and their respective employment rates change more slowly. In addition, both men and women have higher levels of population college attainment rates relative to benchmark, leading to higher labor market participation for both sexes. Conversely, when abilities are homogeneous, fewer women and more men would enter the workforce early on and their respective employment rates change faster. Meanwhile, lower college attainment rates in the population depress the level and growth of employment for both sexes. These effects combine to generate the predicted behavior of the employment ratio series.

When there is no gender comparative advantage in ability endowments, Figure 1.23 shows that the employment ratio is significantly higher than benchmark in 1975 as well as growing more slowly. Over the sample period, the employment ratio grows around 40%, as opposed to 73% in the benchmark model. This is because when women have on average the same amount of physical labor as men, more women would choose to enter the workforce and to work in the labor-intensive occupation when the relative price of labor is high, leading to higher female employment rates and depressing male employment rates early on. Over time, with skill-biased technological change no longer having a gender-biased nature and with more subdued female education response, the female employment rate necessarily rises more slowly, while the male employment rate, benefiting from less female competition and higher levels of male college attainment, actually increases over the period, leading to a more limited growth in the female/male employment ratio.

---

75 This is because, under uniform ability distributions, more high ability individuals would enter the workforce when the employment rate is low and more low ability individuals would drop out of the workforce when the employment rate is high.
76 See Figure 1.23a.
77 See Figure 1.23a.
78 See Figure 1.23a.
Figure 1.22b plots the evolution of the gender wage gap under the different ability distribution assumptions. When abilities are uniformly distributed, the female/male wage ratio is higher and grows faster than benchmark. When abilities are homogeneous, the wage ratio is lower and grows more slowly with respect to benchmark. Several channels can be at work here. First, in the benchmark model, the average cognitive ability of employed female workers is higher than that of employed male workers, which is in turn higher than that of unemployed workers of both sexes. The differentials among groups grow over time, contributing to closing the gender gap (Figure 1.19a). When abilities are homogeneous, the gender gap is larger due to the absence of such ability selection effects and when abilities are uniformly distributed, the gap is smaller due to increased selection effects associated with more dispersed abilities. Second, since female employment is more elastic under homogeneous abilities, the relative wage of female workers rises faster when abilities are uniformly distributed than when they are homogeneous. Finally, more men and women are college-educated under uniformly distributed abilities than under homogeneous abilities.\textsuperscript{79} This can lead to higher relative wages for women both because the college attainment rate is initially lower among women and because relatively more college-educated women work in the skill-intensive occupation.

When there is no gender comparative advantage in ability endowments, the female/male wage ratio is initially higher, but grows significantly slower. The reason is that by virtue of both sexes having the same average amount of physical labor, the female/male wage ratio is going to be smaller, \textit{ceteris paribus}. Over time, however, slower college attainment increase among women, faster college attainment increase among men, higher share of females in the labor-intensive occupation and lower share of females in the skill-intensive occupation retards the growth in the female/male wage ratio.

\textsuperscript{79}See Figure 1.23a.
1.5 Conclusion

This paper analyzes changes in the U.S. labor market by estimating a dynamic general equilibrium model that simultaneously accounts for the evolution of wages, employment, and educational attainment in the U.S. from 1975 to 2010. The estimated model successfully matches trends in the U.S. labor market and reveals changes in the underlying ability composition of the workforce. I find that (1) the average cognitive ability of 25 – 29 year old college graduates increased from 1975 to the late 1980s and then decreased; (2) the average cognitive ability of the college-educated population increased from 1975 to the early 2000s and then decreased; (3) the average cognitive ability of the workforce increased throughout the period of 1975 – 2010, while its average physical labor decreased throughout the period. The results contradict earlier findings that increase in college attainment has led to a large decline in the average ability of college graduates and suggest that female selection into the workforce has become more positive over the last four decades.

I use the estimated model as a basis for counterfactual experiments to understand the sources of labor market changes. The main findings are: (1) in addition to female-biased technological change, neutral and skill-biased technological change both contribute to increasing female relative labor supply and decreasing the gender wage gap; (2) female-biased demand shifts discourage male college attainment and labor supply, and would by themselves lead to a lower college premium; (3) the value of the home sector has been increasing for men and decreasing for women; this process contributes to increasing female relative labor supply and college attainment and helps to narrow the gender wage gap; (4) female comparative advantage in education and skill-intensive job is an important source in the convergence of gender wage and employment gaps and serves to slow down the growth in the college premium; (5) composition effects have contributed to increasing the college premium and narrowing the gender wage gap, but the effects are quite small. The experiments also highlight the sensitivity of labor market outcomes to underlying ability distributions.
It is useful to contrast the results with those of Lee and Wolpin (2010). Although my model and theirs do not nest each other, there are common exogenous forces that are present in both models. This includes neutral and skill-biased technological change and changes in home sector value. Since the modeling of these forces are also different, the numbers are not directly comparable, but several comparisons can be made. First, similar to this paper, Lee and Wolpin find that skill-biased technological change has contributed to increasing college attainment and the college premium, and has worked to increase the female-to-male employment and wage ratios. There is nevertheless a difference in mechanisms. In Lee and Wolpin (2010), skill-technological change can help close the gender wage and employment gaps because of stronger female preference for service sector or white/pink-collar jobs, while in this paper, the source of gender comparative advantage lies in the difference in ex ante ability endowments. It is worth noting that my model also allows gender difference in preference toward white- and blue-collar jobs, but once difference in ability endowments is accounted for, estimation results reveal little gender difference in occupational preference. Second, Lee and Wolpin find that Hicks-neutral technological change has similar effects, both in direction and magnitude, as skill-biased technological change, in helping to increase educational attainment and closing the gender wage and employment gaps. The two processes differ significantly only in their contributions to the increase in the college premium, where the effects of skill-biased technological change dominate. According to the results in this paper, Hicks-neutral technological change plays a significantly smaller role in closing the gender wage and employment gaps. The difference in results may be attributable to the fact that what Lee and Wolpin call Hicks-neutral technological change is actually a combination of neutral technological change in the goods and service sectors and changes in the two sectors’ respective product prices. Hicks-neutral tech-

Due to differences in modeling, the exact meaning of these processes are different in the two papers. In Lee and Wolpin (2010), skill-biased technological change refers to demand shifts toward more skill-intensive occupations within the goods and service sectors. Hicks-neutral technological change includes both neutral technological change in the goods and service sectors and changes in their respective product prices, which are indistinguishable in the model.
nological change in their model therefore captures demand shifts toward the service sector. The results in this paper should serve as a clearer assessment of the effects of neutral technological change. Third, the effects of changes in home sector value are found to be similar in both papers. Finally, a number of exogenous factors are present in their model but not mine and vice versa. In this paper, I find that there are gender-biased demand shifts that close the gender wage and employment gaps. These shifts also have a moderating effect on the rise of the college premium. The results are consistent with findings by Heathcote et al. (2010) and Johnson and Keane (2013), who both include gender-biased demand shifts in their models. On the other hand, Lee and Wolpin include capital-skill complementarity and exogenously rising wage variances in their model. They find that the former has a negligible effect on the U.S. employment and wage structure, while the latter plays a significant role in increasing college attainment as well as closing the gender wage and employment gaps.

One of the key features of this model lies in its treatment of heterogeneity, as I allow individuals to be endowed with two-dimensional, continuously distributed abilities. In this paper, I show that if the population is assumed to be ex ante homogeneous, the economy would be very different. The college premium would be lower. The gender gap would be larger, while college attainment and female labor supply would increase faster. This implies that if we had estimated the model assuming that the population is more homogeneous than it actually is, then we are likely to obtain upward biased estimates of the rate of skill- and gender-biased demand shifts. On the other hand, although the analysis highlights the sensitivity of labor market outcomes to assumptions regarding the underlying ability distributions, the sensitivity of model estimates to such assumptions is unknown. It would therefore be a helpful exercise to re-estimate this model assuming different structures of ex ante heterogeneity and compare the results. I am current conducting such an exercise.

81 One exercise is to re-estimate the model assuming four ex ante individual types as in Lee and Wolpin (2010).
Finally, several important limitations of my analysis are worth emphasizing. One is that the distribution of schooling costs is assumed to be time-invariant. This is likely not true. Another limitation is that I do not model on-the-job skill acquisition and instead use age as a proxy for experience. Gender-biased demand shifts in my model thus captures increased female work experiences and should be interpreted as such. The choice of the model, of course, is based on computational concerns and future work should allow these features to be embedded to achieve a more complete modeling of the U.S. labor market.
# 1.6 Appendix

## 1.6.1 Tables

Table 1.1: Employment Rates by Age (%)

<table>
<thead>
<tr>
<th></th>
<th>Skill-intensive</th>
<th>Labor-intensive</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 - 36</td>
<td>38.45</td>
<td>41.39</td>
<td>20.15</td>
</tr>
<tr>
<td>37 - 48</td>
<td>41.53</td>
<td>40.13</td>
<td>18.34</td>
</tr>
<tr>
<td>49 - 60</td>
<td>37.75</td>
<td>36.29</td>
<td>25.95</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 - 36</td>
<td>40.45</td>
<td>11.38</td>
<td>48.17</td>
</tr>
<tr>
<td>37 - 48</td>
<td>41.56</td>
<td>12.91</td>
<td>45.53</td>
</tr>
<tr>
<td>49 - 60</td>
<td>36.02</td>
<td>11.98</td>
<td>52.00</td>
</tr>
</tbody>
</table>

Table 1.2: Employment Rates by Year (%)

<table>
<thead>
<tr>
<th></th>
<th>Skill-intensive</th>
<th>Labor-intensive</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975 - 1986</td>
<td>38.01</td>
<td>42.26</td>
<td>19.72</td>
</tr>
<tr>
<td>1987 - 1998</td>
<td>39.38</td>
<td>39.93</td>
<td>20.68</td>
</tr>
<tr>
<td>1999 - 2010</td>
<td>40.18</td>
<td>37.74</td>
<td>22.08</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975 - 1986</td>
<td>30.75</td>
<td>11.88</td>
<td>57.36</td>
</tr>
<tr>
<td>1987 - 1998</td>
<td>40.99</td>
<td>12.33</td>
<td>46.68</td>
</tr>
<tr>
<td>1999 - 2010</td>
<td>44.90</td>
<td>11.97</td>
<td>43.13</td>
</tr>
</tbody>
</table>
Table 1.3: Transition Rates by Age (%)

<table>
<thead>
<tr>
<th></th>
<th>Skill-intensive</th>
<th>Labor-intensive</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>25 - 36</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill-intensive</td>
<td>92.03</td>
<td>1.61</td>
<td>6.35</td>
</tr>
<tr>
<td>Labor-intensive</td>
<td>1.63</td>
<td>86.66</td>
<td>11.71</td>
</tr>
<tr>
<td>Residual</td>
<td>10.33</td>
<td>17.97</td>
<td>71.71</td>
</tr>
<tr>
<td><strong>37 - 48</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill-intensive</td>
<td>94.08</td>
<td>0.78</td>
<td>5.14</td>
</tr>
<tr>
<td>Labor-intensive</td>
<td>0.88</td>
<td>89.22</td>
<td>9.90</td>
</tr>
<tr>
<td>Residual</td>
<td>6.40</td>
<td>14.19</td>
<td>79.41</td>
</tr>
<tr>
<td><strong>49 - 60</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill-intensive</td>
<td>92.21</td>
<td>0.60</td>
<td>7.19</td>
</tr>
<tr>
<td>Labor-intensive</td>
<td>0.59</td>
<td>87.96</td>
<td>11.45</td>
</tr>
<tr>
<td>Residual</td>
<td>4.48</td>
<td>8.02</td>
<td>87.50</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>25 - 36</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill-intensive</td>
<td>87.59</td>
<td>0.66</td>
<td>11.75</td>
</tr>
<tr>
<td>Labor-intensive</td>
<td>2.89</td>
<td>78.43</td>
<td>18.68</td>
</tr>
<tr>
<td>Residual</td>
<td>8.14</td>
<td>4.42</td>
<td>87.43</td>
</tr>
<tr>
<td><strong>37 - 48</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill-intensive</td>
<td>91.20</td>
<td>0.42</td>
<td>8.37</td>
</tr>
<tr>
<td>Labor-intensive</td>
<td>1.39</td>
<td>84.48</td>
<td>14.13</td>
</tr>
<tr>
<td>Residual</td>
<td>6.92</td>
<td>3.69</td>
<td>89.39</td>
</tr>
<tr>
<td><strong>49 - 60</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill-intensive</td>
<td>91.21</td>
<td>0.28</td>
<td>9.50</td>
</tr>
<tr>
<td>Labor-intensive</td>
<td>0.86</td>
<td>85.08</td>
<td>14.06</td>
</tr>
<tr>
<td>Residual</td>
<td>4.23</td>
<td>2.51</td>
<td>93.27</td>
</tr>
</tbody>
</table>
Table 1.4: Transition Rates by Year (%)

<table>
<thead>
<tr>
<th></th>
<th>Skill-intensive</th>
<th>Labor-intensive</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975 - 1986</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill-intensive</td>
<td>92.61</td>
<td>1.11</td>
<td>6.28</td>
</tr>
<tr>
<td>Labor-intensive</td>
<td>1.10</td>
<td>86.76</td>
<td>12.14</td>
</tr>
<tr>
<td>Residual</td>
<td>8.27</td>
<td>17.54</td>
<td>74.19</td>
</tr>
<tr>
<td>1987 - 1998</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill-intensive</td>
<td>92.73</td>
<td>1.21</td>
<td>6.05</td>
</tr>
<tr>
<td>Labor-intensive</td>
<td>1.28</td>
<td>88.03</td>
<td>10.69</td>
</tr>
<tr>
<td>Residual</td>
<td>7.40</td>
<td>14.51</td>
<td>78.09</td>
</tr>
<tr>
<td>1999 - 2010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill-intensive</td>
<td>93.02</td>
<td>0.91</td>
<td>6.07</td>
</tr>
<tr>
<td>Labor-intensive</td>
<td>1.02</td>
<td>88.52</td>
<td>10.46</td>
</tr>
<tr>
<td>Residual</td>
<td>6.95</td>
<td>11.25</td>
<td>81.80</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975 - 1986</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill-intensive</td>
<td>87.29</td>
<td>0.52</td>
<td>12.19</td>
</tr>
<tr>
<td>Labor-intensive</td>
<td>1.61</td>
<td>90.05</td>
<td>18.34</td>
</tr>
<tr>
<td>Residual</td>
<td>5.96</td>
<td>3.68</td>
<td>90.36</td>
</tr>
<tr>
<td>1987 - 1998</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill-intensive</td>
<td>89.85</td>
<td>0.52</td>
<td>9.63</td>
</tr>
<tr>
<td>Labor-intensive</td>
<td>1.87</td>
<td>82.85</td>
<td>15.28</td>
</tr>
<tr>
<td>Residual</td>
<td>6.96</td>
<td>3.85</td>
<td>89.19</td>
</tr>
<tr>
<td>1999 - 2010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill-intensive</td>
<td>90.33</td>
<td>0.44</td>
<td>9.23</td>
</tr>
<tr>
<td>Labor-intensive</td>
<td>1.87</td>
<td>83.71</td>
<td>14.42</td>
</tr>
<tr>
<td>Residual</td>
<td>7.02</td>
<td>3.41</td>
<td>89.58</td>
</tr>
</tbody>
</table>
Table 1.5: Ability Endowment: Parameter Estimates

<table>
<thead>
<tr>
<th>Physical Labor</th>
<th>Cognitive Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_1$</td>
<td>43.70 (0.3747)</td>
</tr>
<tr>
<td>$b_2$</td>
<td>2.02 (0.0182)</td>
</tr>
<tr>
<td>$(\ell, \kappa)$</td>
<td>0.52 (0.0035)</td>
</tr>
</tbody>
</table>

Gender Comparative Advantage

| $\chi_f$ | 0.41 (0.0027)

Standard errors in parentheses.

Table 1.6: Ability Endowment: Summary Statistics

<table>
<thead>
<tr>
<th>Physical Labor</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>$E(\ell_i)$</td>
<td>0.97</td>
</tr>
<tr>
<td>$Var(\ell_i)$</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cognitive Ability</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(\kappa_i)$</td>
<td>0.47</td>
</tr>
<tr>
<td>$Var(\kappa_i)$</td>
<td>0.047</td>
</tr>
</tbody>
</table>
Table 1.7: Production Function: Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th>Skill-intensive</th>
<th>Labor-intensive</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{\alpha}_0$</td>
<td>1.57 (0.0120)</td>
<td>0.79 (0.0039)</td>
</tr>
<tr>
<td>$\tilde{\alpha}_1$</td>
<td>0.0183 (.0006)</td>
<td>0.0091 (0.0003)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.56 (0.0007)</td>
<td>-1.20 (0.0140)</td>
</tr>
<tr>
<td>$\tilde{\lambda}_0$</td>
<td>0.74 (0.0043)</td>
<td></td>
</tr>
<tr>
<td>$\tilde{\lambda}_1$</td>
<td>0.0058 (0.0002)</td>
<td></td>
</tr>
</tbody>
</table>

Standard errors in parentheses.

Table 1.8: Skill Production: Parameter Estimates

<table>
<thead>
<tr>
<th>Age Variable</th>
<th>Skill-intensive</th>
<th>Labor-intensive</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Age - 25)</td>
<td>0.054 (0.0004)</td>
<td>0.068 (0.0005)</td>
</tr>
<tr>
<td>(Age - 25)$^2$</td>
<td>-0.0010 (0.0000)</td>
<td>-0.0015 (0.0000)</td>
</tr>
</tbody>
</table>

Standard errors in parentheses.

Table 1.9: Utility of Work: Parameter Estimates

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor-intensive</td>
<td>1.56 (0.0083)</td>
<td>1.46 (0.0567)</td>
</tr>
</tbody>
</table>

### Table 1.10: Value of Residual Sector: Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>$\tilde{\phi}_{00}$</td>
<td>3.85 (0.0292)</td>
</tr>
<tr>
<td></td>
<td>$\tilde{\phi}_{01}$</td>
<td>2.49 (0.0192)</td>
</tr>
<tr>
<td></td>
<td>$\tilde{\phi}_{02}$</td>
<td>2.83 (0.0169)</td>
</tr>
<tr>
<td>(Age - 25)</td>
<td></td>
<td>0.0113 (0.0001)</td>
</tr>
<tr>
<td>(Age - 25)$^2$</td>
<td></td>
<td>0.0001 (0.0000)</td>
</tr>
</tbody>
</table>

Standard errors in parentheses.

### Table 1.11: Cost of Mobility: Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occupation Switch from Residual to Formal</td>
<td>22.00 (0.1637)</td>
<td>26.25 (0.1958)</td>
</tr>
<tr>
<td></td>
<td>24.17 (0.1656)</td>
<td>22.91 (0.1536)</td>
</tr>
</tbody>
</table>


### Table 1.12: Cost of Schooling: Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_1$</td>
<td>165.86 (1.4911)</td>
<td>150.78 (1.2548)</td>
</tr>
<tr>
<td>$c_2$</td>
<td>125.00 (0.7634)</td>
<td>120.86 (0.8786)</td>
</tr>
</tbody>
</table>

Table 1.13: Idiosyncratic Shocks: Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu_w$</td>
<td>14.41 (0.0837)</td>
<td>10.77 (0.0583)</td>
</tr>
<tr>
<td>$\nu_o$</td>
<td>7.34 (0.0650)</td>
<td>4.00 (0.0348)</td>
</tr>
<tr>
<td>$\nu_e$</td>
<td>75.58 (0.4055)</td>
<td>30.21 (0.2196)</td>
</tr>
</tbody>
</table>


1.6.2 Figures

![Figure 1.1: Skill-Experience Profile](image)

Figure 1.1: Skill-Experience Profile
Figure 1.2: College Enrollment Probability

Figure 1.3: College enrollment decision under homogeneous attendance cost
Figure 1.4: GDP Components and the imputed \( \{ \gamma_t \} \) series
Figure 1.5: Labor Supply and Occupation Choice Decision Tree

Figure 1.6: Convergence in Gender Wage Disparities by Occupation
Figure 1.7: College Enrollment Probability (a) given relative payoff; (b) given attendance cost.

Figure 1.8: Complementary Uniform and Beta(5,5) CDF
Figure 1.9: Ability Distribution, Idiosyncratic Shock Variance, and Elasticity of Supply. The horizontal axis is the value associated with a choice and the vertical axis is the percentage of individuals making the choice in the population.

Figure 1.10: Estimated Ability Distribution in Men. Cooler colors represent lower densities and warmer colors represent higher densities. Histograms of each ability dimension are plotted next to the vertical and horizontal axes.
Figure 1.11: Model-Data Comparison - Relative Wages. (a) log college/non-college wage ratio. (b) female/male wage ratio.

Figure 1.12: Model-Data Comparison - Education and Labor Supply. (a) college attainment rate by gender among individuals 25 - 29 years old. (b) rate of employment by gender
Figure 1.13: Model-Data Comparison - Wage Variance. (a) relative changes in the overall variance of log wages. (b) relative changes in the variance of log wages among college- and non-college-educated individuals. Relative change = \((V_t - V_{1975})/V_{1975}\). The initial values \(V_{1975}\) are reported within the legends.

Figure 1.14: Evolution of the Mean Cognitive Ability of College Graduates in Entering Cohorts
Figure 1.15: Ability Distribution by Education among Females in Selected Entering Cohorts. (a) - (d) shows respectively the 1975, 1985, 1995 and 2010 cohorts. In each plot, the red dots represent individuals who have obtained college education and the green dots represent those who haven’t. Nonparametric density plot of the cognitive ability of college graduates is given next to the vertical axis. Nonparametric density plot of the physical labor endowment of college graduates is given below the horizontal axis.
Figure 1.16: Ability Distribution by Education among Males in Selected Entering Cohorts. (a) - (b) shows respectively the 1975 and 2010 cohorts. In each plot, the red dots represent individuals who have obtained college education and the green dots represent those who haven’t. Nonparametric density plot of the cognitive ability of college graduates is given next to the vertical axis. Nonparametric density plot of the physical labor endowment of college graduates is given below the horizontal axis.

Figure 1.17: Evolution of Mean Cognitive Ability and Physical Labor in the College-educated Population
Figure 1.18: Evolution of Population Log College/Non-College Cognitive Ability Ratio

Figure 1.19: Evolution of Mean Cognitive Ability and Physical Labor in Employment
Figure 1.20: Decomposing Labor Market Changes - Relative Wages. (a) log college/non-college wage ratio. (b) female/male wage ratio. The labels in the legend refer to the specific component turned on in the experiment. “RES” denotes changes in residual sector value, “HN” Hicks-neutral technological growth, “SB” skill-biased technological change, and “GB” gender-biased demand shift.

Figure 1.21: Decomposing Labor Market Changes - Education and Labor Supply. (a) college attainment rate by gender among individuals 25 - 29 years old. (b) female/male employment rate. The labels in the legend refer to the specific component turned on in the experiment. “RES” denotes changes in residual sector value, “HN” Hicks-neutral technological growth, “SB” skill-biased technological change, and “GB” gender-biased demand shift.
Figure 1.22: Ability Distribution and Labor Market Outcomes - Relative Wages. (a) log college/non-college wage ratio. (b) female/male wage ratio. The labels in the legend refer to the specific ability distributions assumed in the experiment. “No CA” denotes no gender comparative advantage in ability endowments, “Uniform” uniform ability distributions, and “Degenerate” degenerate ability distributions.

Figure 1.23: Ability Distribution and Labor Market Outcomes - Education and Labor Supply. (a) college attainment rate by gender among individuals 25 - 29 years old. (b) female/male employment rate. The labels in the legend refer to the specific ability distributions assumed in the experiment. “No CA” denotes no gender comparative advantage in ability endowments, “Uniform” uniform ability distributions, and “Degenerate” degenerate ability distributions.
1.6.3 Occupation Definition

Skill-intensive Occupation

- Managerial and Professional Occupations
  - Managers and Executives, Engineers, Scientists, Teachers, Doctors, Lawyers, Writers and Artists

- Technical, Sales, and Administrative Support Occupations
  - Technicians, Sales and Marketing professionals, Secretaries and Staff

Labor-intensive Occupation

- Service Occupations
  - Police, Waiters, Barbers, Nurses, Janitors, domestic servants

- Precision Production, Craft, and Repair Occupations
  - Tailors, Mechanics, Construction workers, Electricians, Plumbers

- Operators, Fabricators, and Laborers
  - Machine operators, Assemblers, Bus and cab drivers
1.6.4 Moment Conditions

Employment

- Occupational employment rate by year, sex, age
- Occupational employment rate by year, sex, education
- Occupational employment rate by year, sex, age, and last-period activity

Education

- College attainment rate at age 25 by year, sex

Wage

- Mean log annual real wage by occupation, year, sex, age
- Mean log annual real wage by occupation, year, sex, education

1.6.5 Standard Errors

Let $m^D$ and $m^S(\theta)$ denote respectively the vector of sample and simulated moments. Let $g(\theta) = m^D - m^S(\theta)$. Then the estimator is defined by

$$\hat{\theta} = \arg \min_{\theta \in \Theta} g(\theta)' \Omega g(\theta)$$

where $\Omega$ is a positive definite weighting matrix.

Weighting Matrix To obtain the optimal weighing matrix, I follow a two-step procedure. In the first step, calculate the weighing matrix as

$$\tilde{\Omega} = \tilde{V} (m^D)^{-1}$$

where $V(\cdot)$ denotes asymptotic variance, and estimate $\theta$ according to

$$\tilde{\theta} = \arg \min_{\theta \in \Theta} g(\theta)' \tilde{\Omega} g(\theta)$$
In the second step, update the weighting matrix by

\[ \hat{\Omega} = \hat{V}(g(\hat{\theta}))^{-1} \]

, and re-run estimation to get

\[ \hat{\theta} = \arg \min_{\theta \in \Theta} g(\theta)' \hat{\Omega} g(\theta) \]

\( \hat{V}(m^D) \) and \( \hat{V}(g(\hat{\theta}^{(1)})) \) are both computed by bootstrap.

**Standard Errors** The variance-covariance matrix of \( \hat{\theta} \) is given by

\[ \hat{V}(\hat{\theta}) = \left( \hat{G}(\hat{\theta})' \hat{\Omega} \hat{G}(\hat{\theta}) \right)^{-1} \hat{G}(\hat{\theta})' \hat{\Omega} \hat{S}(\hat{\theta}) \hat{\Omega} \hat{G}(\hat{\theta}) \left( \hat{G}(\hat{\theta})' \hat{\Omega} \hat{G}(\hat{\theta}) \right)^{-1} \]

, where

\[ \hat{S}(\hat{\theta}) = \hat{V}(g(\hat{\theta})) \]

, which is similarly computed by bootstrap, and

\[ \hat{G}(\hat{\theta}) = \frac{\partial g(\hat{\theta})}{\partial \theta} = \frac{\partial m^s(\hat{\theta})}{\partial \theta} \]

is the vector of estimated derivatives of the simulated moments with respect to the parameters at \( \hat{\theta} \).
Chapter 2

Going to Graduate School: A Life-cycle Model of Post-College Work and Educational Investment Decisions
Abstract

There has been a continual rise in educational wage differentials in the U.S. over the last thirty years, driven almost entirely by increasing returns to college and graduate education. While college enrollment rate, in response, has increased significantly over the same period, there has been limited growth in the enrollment rate of most graduate programs during the same time. In this paper, I develop and estimate a dynamic model of the educational and employment choices of college-educated workers and investigate possible barriers to enrollment in graduate studies. The model is fitted to a nationally representative sample of college graduates in 1992-93. My results show the existence of substantial transition costs associated with going back to school as well as strong sorting of preferences toward graduate education by mathematical ability.

2.1 Introduction

It has been widely documented that the U.S. wage structure has experienced a significant widening during the last 30 years, driven by dramatic growth at the top of the income distribution, modest growth in the middle, and stagnation at the bottom. From 1979 to 2005, top 1% family income in the U.S. increased from $325,000 to nearly $1.1 million, the middle 60% increased from $42,000 to $51,000, while the poorest 20% of households saw their income increase by only $1,000 from $14,500 to $15,500. The phenomenon has been described by Autor, et al. (2006) as the “polarization” of the U.S. labor market.

Perhaps not surprisingly, the widening and polarization of the wage structure has been associated with significantly expanded educational wage differentials, dominated almost entirely by increased returns to post-secondary schooling. According to estimates by Goldin and Katz (2007), 65% of the increase in

\footnotetext{1}{See, for example, Piketty and Saez (2002), Autor, et al. (2006), Goldin and Katz (2007).}

\footnotetext{2}{All numbers are quoted in 2005 dollars.}
U.S. wage inequality between 1980 and 2005, measured by the variance of log hourly wages, can be accounted for by increased returns to college and post-college education. Most notably, within-group inequality grew substantially during this period of time among college-educated workers, driven largely by increase in the returns to post-college education. Since 1980, the wage gap between male post-college graduates and college-only graduates has risen more than the wage gap between male college-only graduates and high school graduates. Analyses by Mincer (1997) and Deschenes (2001) also show that (log) wage has become an increasingly convex function of years of education.

Post-college education, therefore, appears to be an excellent investment today. Yet unlike college enrollment, enrollment in graduate school has hardly increased in response to the rising wage premium associated with post-college education. Between 1980 and 2003, the percentage of male college-educated workers almost doubled, increasing from 11.6 percent in 1980 to 14.8 percent in 1990 to 20.4 percent in 2003, yet the percentage of those obtaining a graduate degree only increased from 9.4 percent in 1980 to 10.9 percent in 1990 and then decreased to 10.6 percent in 2003. Similarly, the percentage of female college-educated workers increased from 11.0 percent in 1980 to 21.5 percent in 2003, while the percentage of those getting graduate degrees increased at a much more modest pace from 6.9 to 10.9 percent during the same span (Acemoglu and Autor 2010). The same lukewarm increase in post-college education attainment is observed not only in the general population but among graduates of the most elite colleges. In a longitudinal study following the family and career transitions of three cohorts of Harvard college graduates (1970, 1980, 1990), Goldin and Katz (2008) report that the percentage of female graduates completing medical school or getting a doctoral degree as well as the percentage of male graduates completing medical school or law school within 15 years of college education all decreased between the 1970 cohort and the 1990 cohort. The only post-college program that has seen significant increase in popularity is the MBA, with the percentage of MBAs doubled among both male and female graduates. These findings suggest that the tepid growth of post-college education attainment observed in the general
population is not the result of limited supply of graduate schools that could limit enrollment by more selective admission.

While a voluminous literature studies the evolution of the college wage premium or look at the individual’s decision to attend college, few have examined human capital investment in post-college education, despite its increasing importance. In this paper, I build a dynamic discrete choice model to study the post-college educational investment and labor supply decisions of college graduates. The model consists of two stages. In the first stage, individuals enter college and choose a major field of study based on expectations of their GPA, future earnings, and probability of entering and completing different types of graduate programs in the future. In the second stage, individuals are graduated from college and in each period choose whether to enroll in a graduate program, work, or engage in home production. In each case, the individual’s choice is rationalized as the optimal solution to a dynamic discrete choice problem that maximizes her expected lifetime utility.

The choice of a dynamic structural model has several advantages. First, it captures the dynamics of post-college education investment. While most people attend school continuously from grade school all the way through college, years of employment or home production can separate graduate school enrollment from college graduation. Indeed, programs such as the MBA positively requires work experience. In addition, people may drop out of programs, re-enroll in programs, getting multiple degrees, or work and enroll at the same time: post-college education is characterized by frequent entry and exit dynamics as people move between work and school. Such dynamics can only be captured by a dynamic model. In addition, it allows me to model unobserved ability and serially correlated preferences to better control for selection, which has been a central problem for regression studies on human capital investment and returns to schooling.

The model is fit to data from the 1992-1993 Baccalaureate and Beyond Longitudinal Study (B&B:93/03), which follows a cohort of college graduates who earned their bachelor’s degrees in the 1992-93 academic year. Follow-up interviews were conducted in 1994, 1997, and 2003, and information were
collected on each student’s education and employment decisions dating back to the preceding interview. Using data on individual SAT scores, college GPA, major field of study, work, and graduate school attendance decisions after college, I am able to estimate the parameters of interest by matching sample moments to the moments generated by the model.

Several parameters of my central interest concern the existence and magnitude of transition costs between schooling and work. What are the costs faced by college and post-college graduates who wish to look for a job and move into the workplace? What are the costs faced by college-educated workers who wish to go back to school and enroll in a graduate program? The answers to these questions are important in understanding the enrollment and employment dynamics of college graduates, and may explain the observed slow growth in graduate degree attainment rate in the face of dramatically increased wage premiums associated with post-college education over the last 30 years. Indeed, results from previous studies show that such transition costs can be very large. Keane and Wolpin (1997) and Lee (2005), in their study of individual life-cycle schooling, work and occupation choices, both estimate school reentry costs to be in the range of tens of thousands of dollars. Artuç, et al. (2008), in their study of labor mobility, find a moving cost equal to several times average annual wages for workers who move from one broadly aggregated sector of the economy to another. The possible existence of significant transition costs between schooling and work, therefore, may prevent graduate school enrollment to respond effectively to changing wage differentials.

In addition to studying post-college education investment decisions, the paper contributes to the literature on the returns to college majors and on college major choice. Most studies in the literature perform OLS regressions of wages earned by college graduates at some point in their career on their college majors and other characteristics, or use logistic/probit regressions to look at the relationship between major choice and expected future wages - assuming perfect foresight (see, for example, Berger, 1988, James et al., 1989). Not only do these studies use labor market earnings at a single point in time to represent lifetime earnings of a college-educated worker, they do not differ-
entiate between those with terminal Bachelor’s degrees and those who went on to attain graduate degrees. Yet different majors represent different specific human capital investments that could affect the type of graduate programs one could attend after college. For example, law school enrolls a majority of students from Humanities and Social Science majors, while Medical school recruits mostly Biological science majors. The rewards from majoring in a particular field, therefore, should include the option value of continuing to graduate school and enjoying the corresponding returns. Eide and Waehrer (1998) show that the option value is a significant, positive factor in the choice of liberal arts and Science majors. The dynamic model allows me to calculate the returns to each college major, taken into account its associated option value of post-college education. In addition, I allow the returns to each major to differ for each individual according to her comparative advantage, so that each individual chooses the major and the graduate programs to attend that best suit her abilities and preferences.

My model builds most directly on a line of inquiry by Keane and Wolpin (1997), Lee (2005), and Lee and Wolpin (2006), which use dynamic discrete choice models to study individual work, schooling, and occupational choices, as well as their aggregate implications in the allocation of labor across sectors and occupations. I offer, of course, a much more detailed look at determinants of individuals’ educational choices at the college and post-college level. Another related study is Arcidiacono (2004), who builds a dynamic model with learning to study college students’ major choices. Among its findings is the existence of strong sorting across major fields by math ability. Students with the highest math ability, for example, majored in Science, Math and Engineering, while those with the least math ability majored in Education. The rest of the paper is arranged as follows. In section 2.2 I outline the model. Section 2.3 describes the data. Section 2.4 presents the estimation results. Finally, section 2.5 concludes the paper.
2.2 Model

2.2.1 State Space and Timing

Time is discrete. The agents in the model consist of the universe of college students, each of whom faces a series of discrete choice problems as she progresses from college to the work place. Her life in the model consists of two stages. In the first stage, she is a college student. Her main choice in this stage is that of academic major. In the second stage, she is a college graduate and makes a decision, in each period, on whether to work, stay home, or go to graduate school.

When an agent enters the first stage, i.e. when she graduates from high school and enters college, she is differentiated from others by her gender and her endowment of mathematical and verbal abilities, which I proxy by SAT math and verbal test score quartiles. Her state at the beginning of the first stage is represented by $s_{i0} = [am_i, av_i, f_i]$, where $f_i \in \{0, 1\}$ denotes her gender, and $am_i$ and $av_i$ are respectively her SAT math and verbal test score quartiles. During college, the agent has to choose a major among $M$ different fields to study. I aggregate college majors into $M = 4$ fields: Business, Education, Humanities & Social Science, and Science, Math & Engineering.

In the second stage, the agent begins her post-college life. In each period, she chooses whether to work, enroll in a graduate program, or stay home. There are a number of graduate programs to choose from, which I aggregate into $K = 4$ categories: master’s, MBA, professional, and doctoral programs. The choice set $\mathcal{C}$ an agent faces in each period thus contains six alternatives: enrolling in one of the $four$ graduate programs, work, and engaging in home production. The agent is assumed to be “active” for $T$ periods. Each period corresponds roughly to a year in her post-college life. I set $T$ equal to forty, implying a typical “retirement age” of sixty-two for the universe of college graduates. Consequently, the state of an agent at the beginning of each period in the second stage can be characterized by $s_{it} = [d_{i,t-1}, k_{it}, m_i, g_i, f_i]$, where

---

3The aggregation follows that of Arcidiacono (2004), with the criteria for aggregation being the degree of similarity in mean earnings and SAT math and verbal test scores.
is the activity \( i \) has chosen\(^1 \), \( k_{it} \in \{0, \ldots, K\} \) is the most recent type of graduate program for which \( i \) has obtained a degree by time \( t \), with \( k_{it} = 0 \) indicating that \( i \) has not obtained any graduate degrees, \( m_i \in \{1, \ldots, M\} \) is \( i \)'s chosen undergraduate major, and \( g_i \) is \( i \)'s realized undergraduate GPA, which I discretize into \( G = 4 \) levels, corresponding respectively to grade intervals of \([0, 2.5), [2.5, 3.0), [3.0, 2.5), \) and \([3.5, 4.0]\)^5.

### 2.2.2 Labor Market Returns

Let \( w_{it} = w(s_{it}) \) be the wage individual \( i \) receives if she chooses to work in period \( t \). The wage equation is given by

\[
\log w_{it} = \alpha_0 mk + \alpha_1 mk f_i + \alpha_2 mk g_i + \xi_{imkt}
\]

, where \( k \) and \( m \) denote respectively \( i \)'s most recent graduate degree and her undergraduate major.

Equation (2.1) performs a separate wage regression for each combination of graduate degree and undergraduate major. While it is reasonable to suspect that undergraduate major does not play an important role in wage determination for those with graduate degrees, individuals with different undergraduate majors may enroll in different graduate programs. Performing separate regressions for each combination allows me to avoid having to specify a finer categorization of graduate programs. For example, the model does not distinguish between a master's degree in Science and a master's degree in Education. The two degrees may nevertheless have very different labor market returns. Controlling for an individual’s undergraduate major alleviates this problem of aggregation. The inclusion of gender and GPA in the wage equation reflects the idea that different majors and graduate degrees can have different values to

---

4Specifically, I index the alternatives in the choice set as follows: (1) enroll in a master’s program \((d = 1)\); (2) enroll in MBA \((d = 2)\); (3) enroll in a professional program \((d = 3)\); (4) enroll in a doctoral program \((d = 4)\); (5) work \((d = 5)\); (6) stay home \((d = 6)\). For \( t = 1 \), I set \( d_{i,t-1} = 6 \).

5Since different institutions have different grading scales - some colleges give a maximum grade of 4.0 (A) while others grant a maximum grade of 4.3 (A+) - I first normalize GPA values on a scale of \(0.0 - 4.0\).
each member of the sexes and that grades achieved in different undergraduate majors can have different values to individuals with different types of graduate degrees. For example, a female Education major may be able to earn a higher expected wage than a male Education major, a 3.5 GPA achieved in a Science, Math & Engineering major may be more valuable than a 4.0 achieved in a Education major, and undergraduate GPAs may be a more important predictor of wage to master's degree holders than to doctoral degree holders.

2.2.3 Grade Production

The grades an individual receives in college are summarized by her GPA. The GPA serves two purposes in the model. First, it affects the future wage an individual earns through the wage equation (2.1). Second, it is a signal on how likely the individual is going to succeed in a graduate program. Since different subjects require different combinations of abilities, the GPA is assumed to be a function both of individual math and verbal skills and of the major chosen. In addition, I assume the function differs by gender, i.e., it may take different levels of skills for male and female students to achieve comparable grades in a given major. Specifically, let \( g_i \equiv g(s_{it}, m_i) \) be the GPA individual \( i \) achieves in college. The grade production function is specified by

\[
g_i = \gamma_0mf + \gamma_1mfm_i + \gamma_3mfav_i + \eta_imf\tag{2.2}
\]

where \( f \) denotes \( i \)'s gender.

2.2.4 Degree Attainment Probability

To model the process of graduate degree attainment, I assume that for each individual enrolled in a graduate program, successful program completion and degree attainment in each period is an independent bernoulli trial with probability dependent on the individual's characteristics. Specifically, let \( y_{ikt} \equiv J^k(s_{it}) \) be the probability of individual \( i \) attaining a degree in program \( k \) in period \( t \). Let \( adv_{it} = 1 \) if individual \( i \) already has an advanced degree by time \( t \), and \( = 0 \) if not. The degree attainment probability \( y_{ikt} \) is specified by the
following logistic function

\[
\log \frac{y_{ikt}}{1 - y_{ikt}} = \phi_0 m_k + \phi_1 f_i + \phi_2 g_i + \phi_3 adv_{it}
\]  

(2.3)

, i.e. the probability of attaining degree in a given period is affected by one’s gender, undergraduate GPA, whether one already has advanced degrees, and how much one’s undergraduate major matches with the graduate program she is enrolled in.

The degree attainment probability can be thought of equivalently as the inverse of the mean duration of a program. Given \( y_{ikt} \), the expected number of years (or periods) an individual needs to complete a program is \( 1/y_{ikt} \). Since different graduate programs take different lengths of time to finish and may vary significantly from individual to individual, (2.3) presents a way of modeling the process of graduate education, while avoiding to introduce the number of years enrolled in each program as a set of additional state variables.

2.2.5 The First Stage Problem

In the first stage, individuals choose their major in college. As discussed earlier, I emphasize the option value of college majors. The total benefits associated with choosing a particular major include not only the expected future earnings from a terminal Bachelor’s degree in that field, but also the potential expected returns to graduate degrees that one could pursue after graduating from college with the chosen major. Different majors require different skills and characteristics to excel in and represent different specific human capital investments that could affect what kind of graduate programs one could get into after college as well as how likely she is going to do well in them. Thus, different college majors are associated with different option values of future graduate school attendance for individuals with different skills and characteristics. In my model, this is achieved through the mapping function (2.3), which links the performance of an undergraduate in a major field of study to the future likelihood that she could successfully attain a degree in a particular graduate program in a given period of time.
Specifically, each individual $i$ enters college with an initial endowment of mathematical and verbal skills $am_i$ and $av_i$, based on which she could project her expected GPA level for each major field of study, based on the grade production function (2.2). Together, the major of her choice and the realized GPA level will affect both her future wages and her future degree attainment probabilities in different graduate programs. In addition, each individual $i$ draws a taste factor $\nu_{im}$ for each major $m$ that determines how much utility the individual could derive from studying the subject matter of major $m$. $\nu_{im}$ also incorporates the effort cost one must incur in acquiring the major-specific knowledge, which is not separately identifiable.

Once an individual in state $s_{i0} = [am_i, av_i, f_i]$ chooses a major $m_i$, she will enter the second stage in state $s_{i1} = [d_{i0}, k_{i1}, m_i, g_i, f_i]$, where $k_{i1} = 0$, $g_i$ is her realized GPA level upon graduation, and $d_{i0}$ is set to home production ($d_{i0} = 6$) for convenience. The state transition probability function is thus

$$q(s_{i1} | s_{i0}, m_i) = \begin{cases} \Pr (g_i | s_{i0}, m_i) & \text{if } s_{i0} = [am_i, av_i, f_i]\text{and } s_{i1} = [6, 0, m_i, g_i, f_i] \\ 0 & \text{o.w.} \end{cases}$$

, where $\Pr (g_i | s_{i0}, m_i)$ is the probability of achieving a GPA equal to $g_i$ given the state $s_{i0}$ and major choice $m_i$ based on the grade production function (2.2).

Given drawn taste $\nu_i = [\nu_{i1}, \ldots, \nu_{iM}]$, an individual chooses her major by solving the following problem:

$$V(s_{i0}, \nu_i) = \max_{m \in \{1, \ldots, M\}} \left\{ \nu_{im} + \beta \sum_{s_{i1}} EW_1(s_{i1}) q(s_{i1} | s_{i0}, m) \right\}$$

(2.4)

, where $EW_1(s_{i1}) = \int W(s_{i1}, \epsilon_{i1}) dF(\epsilon_{i1})$ is the expected utility of entering the second stage in state $s_{i1}$.

\section{The Second Stage Problem}

In the second stage, individuals are graduated from college and must choose in each period whether to enroll in one of the $K$ graduate programs, work,
or stay home. Although I do not model the application process or duration
of different graduate programs directly and assume that individuals can
enroll in any program of her choice and can continue to be enrolled for as long
as she wishes, the model generates variations in graduate school enrollment
and completion rates through difference in degree attainment probabilities.
Individuals with low degree attainment probability in program \( k \) will on
average take a longer time to finish or with a higher probability drop out of
the program without getting a degree. Thus, ceteris paribus, these individuals
face a higher opportunity cost of enrolling in the program and will more likely
choose not to do so. This mechanism generates the observed differences in both
graduate school enrollment rates and completion rates and in the allocation of
individuals with different college majors and GPA levels among the different
types of graduate programs.

If an individual decides to enroll in a graduate program, she pays the ex-
penses associated with the program including tuition, room and board, books,
and other living expenses. Let \( StBudget_k \) be the average student budget mini-
minus all grants for program \( k \). If the individual chooses to work, then she earns
a wage equal to \( w(s_{it}) \). There are no monetary returns or costs associated
with home production. In addition, as discussed earlier, I hypothesize the existence
of significant transition costs between schooling, work and home production
that could explain the observed persistence in individual choices. Let \( C_s \) be
the cost of enrolling in a new graduate program if one is not enrolled in the
program during the last period and \( C_w \) be the cost of entering the workplace
if one is not working during the last period. I do not require \{\( C_s, C_w \)\} to
be positive, allowing the possibility that there are actual benefits rather than
costs involved in moving into work or a new school program. The per-period
payoff function associated with state \( s_{it} \) and choice \( d_{it} \) can thus be written as

\[
U (s_{it}, d_{it}) = \left( -StBudget_{d_{it}} - C_s \mathbb{I} \{d_{i,t-1} \neq d_{it}\} \right) \mathbb{I} \{d_{it} \in \{1, \ldots, 4\}\} + \left( w(s_{it}) - C_w \mathbb{I} \{d_{i,t-1} \neq d_{it}\} \right) \mathbb{I} \{d_{it} = 5\}
\]

The payoff function \( U (s_{it}, d_{it}) \) is time-invariant and common to all indi-
viduals. In addition, I assume each individual in each period draws an idiosyncratic shock $\epsilon_{id}^t$ that represents the time-varying component of the cost or benefit she will incur by choosing alternative $d$ in time $t$. Together, the existence of fixed transition costs and time-varying idiosyncratic shocks give rise to differential choice patterns made by observably similar individuals as well as history-dependent choice behaviors. For example, a one time shock may induce an individual to enroll in a graduate program even though its payoff minus the utility shock is less than those of other options. If in subsequent periods, the shock associated with enrolling in the particular program decreases relative to others, the individual may still find it optimal to stay in the program rather than exit due to the existence of transition costs. The history of choices and not just current and future payoff determinants may thus play an important role in determining the dynamic pattern of individual choices.

Once an individual $i$ in state $s_{it} = [d_{i,t-1}, k_i, m_i, g_i, f_i]$ chooses to work or engage in home production in period $t$, her state at the beginning of $t + 1$ becomes $s_{i,t+1} = [d_{it}, k_{it}, m_i, g_i, f_i]$, where $d_{it} \in \{5, 6\}$ is the alternative she has chosen in. If the individual chooses to enroll in one of the graduate programs, then there are two possibilities. Should she successfully obtain a degree from the program, her next period state will be $s_{i,t+1} = [d_{it}, k_{it}, m_i, g_i, f_i]$, where $k_{it} = d_{it}$ and $d_{it} \in \{1, 2, 3, 4\}$ is the program she has been enrolled in. On the other hand, if she does not attain a degree in period $t$, then $k_{it}$ remains unchanged. The state transition probability function can be written as

$$q(s_{i,t+1}|s_{it}, d_{it}) = \begin{cases} J^{d_{it}}(s_{it}) & \text{if } d_{it} \in \{1, 2, 3, 4\}, s_{it} = [d_{i,t-1}, k_{it}, m_i, g_i, f_i], s_{i,t+1} = [d_{it}, d_{it}, m_i, g_i, f_i] \\ 1 - J^{d_{it}}(s_{it}) & \text{if } d_{it} \in \{1, 2, 3, 4\}, s_{it} = [d_{i,t-1}, k_{it}, m_i, g_i, f_i], s_{i,t+1} = [d_{it}, k_{it}, m_i, g_i, f_i] \\ 1 & \text{if } d_{it} \in \{5, 6\}, s_{it} = [d_{i,t-1}, k_{it}, m_i, g_i, f_i], s_{i,t+1} = [d_{it}, k_{it}, m_i, g_i, f_i] \\ 0 & \text{o.w.} \end{cases}$$
Given drawn utility shock $\epsilon_{it} = [\epsilon^1_{it}, \ldots, \epsilon^6_{it}]$, an individual in period $t$ makes her labor supply and graduate school enrollment decisions by solving the following problem:

$$
W_t(s_{it}, \epsilon_{it}) = \max_{d \in \{1, \ldots, 6\}} \left\{ U(s_{it}, d) + \epsilon^d_{it} + \beta \sum_{s_{i,t+1}} EW_{t+1}(s_{i,t+1}) q(s_{i,t+1}| s_{it}, d) \right\}, \ t = 1, \ldots, T
$$

, where $EW_{t+1}(s_{i,t+1}) = \int W_{t+1}(s_{i,t+1}, \epsilon_{i,t+1}) \, dF(\epsilon_{i,t+1})$ is the expected value of entering $t+1$ in state $s_{i,t+1}$. Since it is assumed that individuals are active for only $T$ periods of time, after which they would retire, I set $EW_{T+1}(s_{i,T+1}) = 0 \forall s_{i,T+1}$.

### 2.2.7 Serial Correlation of Preferences and Unobserved Abilities

The basic model described so far assumes a population that is homogeneous upon entering college except for gender differences and initial math and verbal abilities proxied by SAT test score quartiles. Clearly, SAT test scores are imperfect proxies for true math and verbal abilities and are themselves outcomes of prior human capital investments and innate talents. In addition, individual preferences for different alternatives are generated by random idiosyncratic shocks in each period and are thus not correlated over time. It is reasonable to suspect, however, that individual preferences should be serially correlated, such that an individual who strongly prefers to enroll in doctoral program this period should prefer to do the same next period if she has not yet attained a degree.

To allow for heterogeneous population and serially correlated preferences, I extend the basic model to include $R$ different types of individuals with $\pi_r$ being the proportion of the $r$th type in the population. The individual types
are unobservable to the econometrician but known to each individual. Each type of individual has a different set of preferences for the various choices in stage 1 and 2 that do not change from time to time. Formally, let the first stage problem for type-$r$ individual be

$$V^r(s_{i0}; \nu_i) = \max_{m \in \{1, \ldots, M\}} \left\{ \zeta^r(m) + v_{im} + \beta \sum_{s_{i1}} EW^r_1(s_{i1}) q(s_{i1}|s_{i0}, m) \right\}$$  \hspace{1cm} (2.7)$$

, where $\zeta^r(m)$ is type-$r$ individual’s preference for major $m$.

Let the second stage problem for type-$r$ individual be

$$W^r_t(s_{it}, \epsilon_{it}) = \max_{d \in \{1, \ldots, 6\}} \left\{ U^r(s_{it}, d) + \epsilon_{it}^d + \beta \sum_{s_{i,t+1}} EW^r_{t+1}(s_{i,t+1}) q(s_{i,t+1}|s_{it}, d) \right\}, \ t = 1, \ldots, T$$  \hspace{1cm} (2.8)$$

, where $U^r(s_{it}, d)$ is type-$r$ individual’s non-random payoff associated with state $s_{it}$ and choice $d$ and is specified by

$$U^r(s_{it}, d) = \xi^r(d) + (-StBudget_d - C_s I\{d_{i,t-1} \neq d\}) I\{d \in \{1, 2, 3, 4, 5\}\}
+ (w(s_{it}) - C_w I\{d_{i,t-1} \neq d\}) I\{d = 5\}$$

, where $\xi^r(d)$ is type-$r$ individual’s time-invariant preference for alternative $d$.

It is unlikely that individuals’ initial endowment of math and verbal skills proxied by their SAT test scores are exogenous. Therefore, I condition them on individual types. In other words, I assume each type of individuals have different probabilities of being in each initial state. This requires me to specify the initial state distribution conditional on each unobserved individual type. Equivalently, I can specify the distribution of individual types for the population in each initial state. Let $p(r|s_{i0})$ be the probability of an individual in state $s_{i0}$ being of type $r$. Let $F(s_{i0})$ be the initial distribution of states. Then

$$p(s_{i0}|r) = \frac{p(r|s_{i0}) F(s_{i0})}{\pi_r}$$

, where the conditional type probabilities $\{p(r|s_{i0})\}$ are treated as estimable model parameters.
2.2.8 Model Solution and Estimation

There are two sets of parameters in the basic model that need to be estimated. The first set contains $\alpha$, coefficients of the wage function (2.1), and $\gamma$, coefficients of grade production function (2.2). Both $\alpha$ and $\gamma$ can be estimated directly from the observed wage and GPA data without solving the dynamic programming programs in (2.4) and (2.6). The rest of the model are parametrized by $\theta = \{\phi, C_s, C_w\}$, which are parameters for the degree attainment probability function (2.3) and the fixed transition costs associated schooling and work.

For the extended model with heterogeneous population, I fix the number of unobserved individual types ($R$) at two and condition type proportions on gender and whether one has low (bottom or 2nd quartile) or high (3rd and top quartile) scores on SAT math and verbal tests. Table 2.11 lists the conditional type probabilities treated as estimable parameters. Let $\rho = \{\rho_1, \ldots, \rho_8\}$. For the extended model, $\theta = \{\rho, \phi, C_s, C_w\}$.

Given $\{\hat{\alpha}, \hat{\gamma}\}$ and any value of $\theta$, the individual in the second stage solves the finite horizon dynamic programming problem embodied in (2.6). The solution takes the form of a sequence of decision rules $\{\sigma_t (., \hat{\alpha}, \hat{\gamma}, \theta)\}_{t=1}^T$, such that

$$
\sigma_t (s_{it}, \epsilon_{it} | \hat{\alpha}, \hat{\gamma}, \theta) = \arg \max_{d \in \{1, \ldots, 6\}} \{ U (s_{it}, d | \hat{\alpha}, \theta) + \epsilon_{it}^d 
+ \beta \sum_{s_{i,t+1}} EW_{t+1} (s_{i,t+1} | \hat{\alpha}, \hat{\gamma}, \theta) q (s_{i,t+1} | s_{it}, d, \theta) \}
$$

Similarly, individuals in the first stage solve the optimization problem embodied in (2.4). Let $\sigma_0 (., \hat{\alpha}, \hat{\gamma}, \theta)$ be the first stage optimal decision rule, then

$$
\sigma_0 (s_{i0}, \nu_i | \hat{\alpha}, \hat{\gamma}, \theta) = \arg \max_{m \in \{1, \ldots, M\}} \left\{ \zeta_r (m) + v_{im} + \beta \sum_{s_{i1}} EW_1 (s_{i1} | \hat{\alpha}, \hat{\gamma}, \theta) q (s_{i1} | s_{i0}, m, \hat{\gamma}) \right\}
$$

\footnote{For the extended model, type-$r$ individual solves}

$$
\sigma^r_t (s_{it}, \epsilon_{it} | \hat{\alpha}, \hat{\gamma}, \theta) = \arg \max_{d \in \{1, \ldots, 6\}} \{ U^r (s_{it}, d | \hat{\alpha}, \theta) + \epsilon_{it}^d 
+ \beta \sum_{s_{i,t+1}} EW^r_{t+1} (s_{i,t+1} | \hat{\alpha}, \hat{\gamma}, \theta) q (s_{i,t+1} | s_{it}, d, \theta) \}
$$
\[
\sigma_0 (s_{i0}, \nu_i | \hat{\alpha}, \hat{\gamma}, \theta) = \arg \max_{m \in \{1, \ldots, M\}} \{ \nu_{im} + \beta \sum_{s_{i1}} EW_1 (s_{i1} | \hat{\alpha}, \hat{\gamma}, \theta) dQ (s_{i1} | s_{i0}, m, \hat{\gamma}) \}.
\]

(2.10) and (2.11) are in general not analytic and are obtained numerically by solving the dynamic programing problems in (2.4) and (2.6) via backward recursion. The solution serves as input to the estimation procedure. Estimation of \( \theta \) is by simulated method of moments (SMM). Specifically, from the optimal decision rules I can calculate the choice probabilities \( \Pr (d | s_{it}, \hat{\alpha}, \hat{\gamma}, \theta) = \Pr (\sigma_t (s_{it}, \epsilon_{it}, \hat{\alpha}, \hat{\gamma}, \theta) = d) \) and \( \Pr (m | s_{i0}, \hat{\alpha}, \hat{\gamma}, \theta) = \Pr (\sigma_0 (s_{i0}, \nu_i | \hat{\alpha}, \hat{\gamma}, \theta) = m) \), and generate a simulated panel data set. The parameters in \( \theta \) are then estimated to minimize a weighted average distance between sample moments in B&B:93/03 to moments simulated by the model. The moments I use in the estimation procedure are: the proportion of individuals choosing each of the six alternatives in 1994, 1997 and 2003, by most recent graduate degree, undergraduate major, GPA level, and gender; the proportion of individuals with the most recent degree in each of the four types of graduate programs as well as the proportion of those who have not attained a graduate degree by 1994, 1997 and 2003, by undergraduate major, GPA level, and gender; the proportion of individuals choosing each of the four major fields of study in college, by SAT math test score quartile, SAT verbal test score quartile, and gender; and the proportion of individuals in each employment and enrollment trajectory through 1994, 1997 and 2003 (a trajectory specifies the employment and enrollment status of an individual in each of the three interview years, such as “work, work, work”), by undergraduate major, GPA level, and gender. Table 2.12 provides the list of aggregate moments as well as the number of conditional moments used in the estimation.

The model assumes mutually exclusive enrollment and employment choices in each period. However, within a year, an individual may engage in several different activities such as working and enrolling in school at the same time. Indeed, a large proportion of graduate school enrollment are part-time only. To accommodate this fact and assign a single activity to each individual in
each period, I adopt the following rules: (i) an individual is considered to be enrolled in a graduate program regardless of whether she is enrolled full-time or part-time; (ii) an individual is considered working if she reports to be full-time employed; (iii) an individual is assumed to engage in home production if she is neither working nor enrolled in any graduate program, this includes the possibilities that she is unemployed, or out of the work force, or enrolled in non-certificate programs.

2.3 Data

I use data from the 1992-1993 Baccalaureate and Beyond Longitudinal Study (B&B:93/03) conducted by the National Center for Educational Statistics (NCES). The study follows a nationally representative sample of undergraduates who earned their bachelor’s degrees in the 1992-93 academic year. Follow-up interviews were conducted in 1994, 1997, and 2003, and information were collected on each student’s education and employment decisions dating back to the preceding interview. Education data include SAT/ACT test scores, GPA and major field of study in college, the highest/most recent graduate program attended, the highest/most recent graduate degree attained, and the current enrollment status at each interview date. Employment data include the employment status and annual salaries at each interview date.

2.3.1 Descriptive Statistics

2.3.1.1 College major distribution

Figure 2.1 shows the college major distribution by gender. There are noticeable differences in the choice of major along the gender line. While the majority of female students enroll in Humanities and Social Science, male students are more evenly spread out between Business, Humanities & Social Science, and Science, Math & Engineering majors. The field of Education, however, is almost entirely dominated by females.
Table 2.13 shows the distribution of college majors by SAT math and verbal test score quartiles. It shows that majors in Humanities & Social Science and in Science, Math & Engineering are strongly sorted by SAT math scores. Higher SAT math quartiles are associated with significantly higher percentage of Science, Math & Engineering majors and significantly lower percentage of Humanities & Social Science majors. In addition, individuals with higher verbal quartiles are in general more likely to major in Humanities & Social Science rather than Business and Education, while those with top quartile SAT math scores are significantly less likely to major in Business and Education.

2.3.1.2 Graduate enrollment

Out of the sample of 1992-93 college graduates, 40% had enrolled in a graduate program by 2003. Figure 2.2 shows the percentage of highest graduate enrollment by 2003 by program type and gender. The overall percentages of male and female graduate enrollment are very close. While females are less likely than males to enroll in MBA, professional an doctoral programs, they are much more likely to enroll in master’s programs. This is largely due to the large number of females enrolled in Master’s of Education (M.Ed) programs or certificate programs in education. Figure 2.3 shows the distribution of highest graduate enrollment by GPA. Not surprisingly, higher GPAs are associated with higher percentage of enrollment in all types of graduate programs.

Figure 2.4 shows the highest graduate enrollment by 2003 by undergraduate major. Out of the four major fields of study, Business majors are least likely to have post-college enrollment, while at the same time having the highest enrollment in MBA programs. Education majors are most likely to have enrolled in master’s programs, the majority of which are M.Ed or certificate programs in education, while having low enrollment in the other types of graduate programs. Both Humanities & Social Science and Science, Math & Engineering majors have higher enrollment in all types of graduate programs, including professional and doctoral programs. These patterns are consistent with the observation that Business and Education majors are more “career-oriented,” while Humanities & Social Science and Science, Math & Engineering majors
are more “academic-oriented” (Choy and Bradburn, 2008). They are also consistent with the finding that the option value of post-college education is a more significant factor in the choice of liberal arts and Science majors (Eide and Waehrer, 1998).

Out of the 40% college graduates in the sample who had enrolled in a graduate program by 2003, 25% enrolled in more than one program during the 10-year span. Table 2.14 gives an indication of this pattern. The table tabulates individuals’ highest enrollment by 2003 by their highest degree in 1997. Degree programs are ranked from lowest to highest in the order of Bachelor’s, Master’s, MBA, Professional, and Doctoral programs. It shows that more than 25 percent of master’s degree holders in 1997 had gone on to enroll in MBA, professional, and doctoral programs by 2003, with the majority of them going to doctoral programs, while a smaller but non-negligible proportion of MBA and professional degree holders in 1997 had also gone on to enroll in doctoral programs by 2003.

2.4 Estimation Results

One of the major difficulties in fitting the model to the data has to do with insufficient number of observations of more advanced degree holders. The wage function (2.1) can only be fully estimated for individuals with Bachelor’s and master’s degrees. For MBA, professional, and doctoral degree holders, I opt to use the mean wage for each combination of graduate degree, undergraduate major and gender as the wage prediction function. That is, I instead estimate the following function

\[
\log(w_{it}) = \alpha_{0mk} + \alpha_{1mk} f_i + \xi_{imkt} \tag{2.12}
\]

In addition, the regression and mean estimates are performed using 2003 wage numbers and are thus expected to overestimate the returns to less advanced degree holders as they had had on average more work experience by
Table 2.1: Wage Regressions: Bachelor’s and Master’s Degree

<table>
<thead>
<tr>
<th>Terminal degree</th>
<th>Bachelor’s</th>
<th>Master’s</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business major</td>
<td>24,540 (4,004)</td>
<td>13,520 (3,598)</td>
</tr>
<tr>
<td>Education major (reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humanities &amp; Social Science major</td>
<td>13,667 (3,186)</td>
<td>1,805 (3,392)</td>
</tr>
<tr>
<td>Science, Math &amp; Engineering major</td>
<td>21,450 (2,853)</td>
<td>20,961 (3,568)</td>
</tr>
<tr>
<td>GPA</td>
<td>25.48 (11.58)</td>
<td>16.50 (19.73)</td>
</tr>
<tr>
<td>Intercept</td>
<td>39,553 (4,703)</td>
<td>41,953 (7,330)</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business major</td>
<td>13,577 (1,758)</td>
<td>12,159 (4,092)</td>
</tr>
<tr>
<td>Education major (reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humanities &amp; Social Science major</td>
<td>9,020 (1,416)</td>
<td>3,068 (2,004)</td>
</tr>
<tr>
<td>Science, Math &amp; Engineering major</td>
<td>15,270 (1,515)</td>
<td>14,700 (2,388)</td>
</tr>
<tr>
<td>GPA</td>
<td>1.17 (11.66)</td>
<td>10.54 (8.37)</td>
</tr>
<tr>
<td>Intercept</td>
<td>34,259 (3,895)</td>
<td>37,663 (3,047)</td>
</tr>
</tbody>
</table>

**Note:** Standard error in parentheses. Table entries report coefficients of weighted OLS regressions of 2003 wage on undergraduate major and GPA by gender for individuals with terminal Bachelor’s and Master’s degrees. Wages are in terms of 2003 dollars. GPA is normalized on a 0 to 400 scale. The reference category for undergraduate major is Education.

2003\(^7\). Table 2.1 and 2.2 present the results. In Table 2.1, Education is used as a reference category. For both Bachelor and master’s degree holders, all three other major fields enjoy significant wage premiums against Education. The premiums are smaller for females than for males. For example, male Bachelor degree Business majors on average earn $24,540 more than their counterpart in Education majors, while female Bachelor degree Business majors only earn $13,577 more than Education majors. GPA also plays a less significant role for females. A one point increase in GPA raises the expected wage of male Bach-

\(^7\)The inability to control for experience is a major problem with the wage estimations performed here.
Table 2.2: Mean Wage: MBA, Professional, and Doctoral Degree

<table>
<thead>
<tr>
<th>Terminal degree</th>
<th>MBA</th>
<th>Professional</th>
<th>Doctoral</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business major</td>
<td>81,994</td>
<td>84,286</td>
<td>162,674</td>
</tr>
<tr>
<td>Education</td>
<td>93,776</td>
<td>65,200</td>
<td>50,830</td>
</tr>
<tr>
<td>Humanities &amp; Social Science major</td>
<td>90,997</td>
<td>102,382</td>
<td>67,821</td>
</tr>
<tr>
<td>Science, Math &amp; Engineering major</td>
<td>79,254</td>
<td>95,105</td>
<td>66,760</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business major</td>
<td>59,675</td>
<td>61,512</td>
<td>68,925</td>
</tr>
<tr>
<td>Education</td>
<td>61,442</td>
<td>55,263</td>
<td>29,372</td>
</tr>
<tr>
<td>Humanities &amp; Social Science major</td>
<td>70,835</td>
<td>67,718</td>
<td>57,213</td>
</tr>
<tr>
<td>Science, Math &amp; Engineering major</td>
<td>69,673</td>
<td>82,272</td>
<td>65,348</td>
</tr>
</tbody>
</table>

**Note:** Table entries report 2003 mean wage by gender, undergraduate major, and GPA for individuals with terminal MBA, professional, and doctoral degrees. Wage are in 2003 dollars.

Bachelor degree holders by $2,548, but only by $117 for females. Overall, male Bachelor’s degree holders earn $21,035 more than females\(^8\). Both the wage premiums associated different undergraduate majors and the effect of GPA are across the board smaller for master’s degree holders than for Bachelor’s degree holders, indicating that major field of study and undergraduate academic performance are less important for the labor market returns of master’s degree holders.

Table 2.2 presents the mean wages of MBA, professional, and doctoral degree holders. In general, wages of MBA and professional degree holders are comparable and are both higher than those of doctoral degree holders with the exception of doctoral degree holders with undergraduate majors in Business, whose reported salaries are surprisingly high. Undergraduate majors do not play a significant role on the wage of MBAs, but have a large impact on

---

\(^8\)Male Bachelor degree holders on average earned $65,460 in 2003, compared to $44,425 for females
Table 2.3: GPA Regressions

<table>
<thead>
<tr>
<th>Major Field of Study</th>
<th>Business</th>
<th>Education</th>
<th>Humanities &amp; Social Science</th>
<th>Science, Math &amp; Engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAT math</td>
<td>0.20 (.07)</td>
<td>0.12 (.09)</td>
<td>-0.01 (.05)</td>
<td>0.16 (.04)</td>
</tr>
<tr>
<td>SAT verbal</td>
<td>0.14 (.05)</td>
<td>-0.13 (.12)</td>
<td>0.13 (.05)</td>
<td>0.03 (.04)</td>
</tr>
<tr>
<td>Intercept</td>
<td>117.75 (35.04)</td>
<td>288.79 (54.56)</td>
<td>218.72 (23.30)</td>
<td>188.10 (28.52)</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAT math</td>
<td>0.06 (.15)</td>
<td>0.15 (.05)</td>
<td>0.05 (.05)</td>
<td>0.02 (.04)</td>
</tr>
<tr>
<td>SAT verbal</td>
<td>0.17 (.08)</td>
<td>0.03 (.04)</td>
<td>0.14 (.04)</td>
<td>0.06 (.07)</td>
</tr>
<tr>
<td>Intercept</td>
<td>190.09 (49.77)</td>
<td>231.04 (14.94)</td>
<td>203.83 (22.98)</td>
<td>269.36 (33.91)</td>
</tr>
</tbody>
</table>

**Note:** Standard error in parentheses. Table entries report coefficients of weighted OLS regressions of college GPA on SAT math and verbal test scores by gender and major field of study. GPA is normalized on a 0 to 400 scale.

The wages earned by professional and doctoral degree holders. For example, professional degree holders with undergraduate majors Humanities & Social Science have a $37,182 wage premium against those with undergraduate major in Education. This likely reflects the different types of professional and doctoral programs individuals from different majors are enrolled in.

Table 2.3 reports the estimated coefficients of the grade production function 2.2. For male students, as expected, SAT math quartiles are particularly important for Science, Math & Engineering majors and SAT verbal quartiles are particularly important for Humanities & Social Science majors. It is more surprising that math abilities seem to be more important than verbal abilities for male students to do well in Business and Education as well. For female students, higher SAT verbal quartiles are more correlated with better grades in all but Education majors, where as in the case of male student, math abilities seem to be more important.

Table 2.4 reports the estimated coefficients of the degree attainment prob-
Table 2.4: Estimated Degree Attainment Probability Function

<table>
<thead>
<tr>
<th></th>
<th>Basic</th>
<th>Extended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>-.47</td>
<td>-.5</td>
</tr>
<tr>
<td>GPA level</td>
<td>.59</td>
<td>.84</td>
</tr>
<tr>
<td>Advanced degree</td>
<td>5.67</td>
<td>6.46</td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s ×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business</td>
<td>-3.13</td>
<td>-3.3</td>
</tr>
<tr>
<td>Education</td>
<td>-.51</td>
<td>-3.09</td>
</tr>
<tr>
<td>Humanities &amp; Social Science</td>
<td>-1.18</td>
<td>-2.78</td>
</tr>
<tr>
<td>Science, Math &amp; Engineering</td>
<td>-1.01</td>
<td>-1.9</td>
</tr>
<tr>
<td>MBA ×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business</td>
<td>-2.89</td>
<td>-2.01</td>
</tr>
<tr>
<td>Education</td>
<td>-8.34</td>
<td>1.84</td>
</tr>
<tr>
<td>Humanities &amp; Social Science</td>
<td>-5.88</td>
<td>-2.54</td>
</tr>
<tr>
<td>Science, Math &amp; Engineering</td>
<td>-6.44</td>
<td>-3.54</td>
</tr>
<tr>
<td>Professional ×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business</td>
<td>-7.53</td>
<td>-9.91</td>
</tr>
<tr>
<td>Education</td>
<td>-7.58</td>
<td>9.96</td>
</tr>
<tr>
<td>Humanities &amp; Social Science</td>
<td>-2.98</td>
<td>-1.95</td>
</tr>
<tr>
<td>Science, Math &amp; Engineering</td>
<td>-3.92</td>
<td>-1.39</td>
</tr>
<tr>
<td>Doctoral ×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business</td>
<td>-10</td>
<td>-8.44</td>
</tr>
<tr>
<td>Education</td>
<td>-8.2</td>
<td>-9.94</td>
</tr>
<tr>
<td>Humanities &amp; Social Science</td>
<td>-5.44</td>
<td>-6.66</td>
</tr>
<tr>
<td>Science, Math &amp; Engineering</td>
<td>-4.87</td>
<td>-9.91</td>
</tr>
</tbody>
</table>

The left column reports the estimates obtained from the basic model (without unobserved individual heterogeneity) and the right column reports the estimates from the extended model (with unobserved individual types). Both sets of estimates show that being female is correlated with a lower degree attainment probability. Having a higher undergraduate GPA is correlated with a higher one, while having already attained a graduate degree significantly raises the degree attainment probability.

Without controlling for unobserved heterogeneity, estimates from the basic model show that among those enrolled in master’s programs, Education majors are most likely to get a degree in any period (−.51). Among those enrolled in MBA programs, Business majors are most likely to get a degree in any period (−2.89), while Education majors are least likely to do so (−8.34). Both Humanities & Social Science majors and Science, Math & Engineering majors
Table 2.5: Estimated transition Costs

<table>
<thead>
<tr>
<th></th>
<th>Basic</th>
<th>Extended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost of entering a new graduate program</td>
<td>169,370</td>
<td>22,500</td>
</tr>
<tr>
<td>Cost of moving into work</td>
<td>-765,000</td>
<td>-808,750</td>
</tr>
</tbody>
</table>

Note: estimates are in $.

are much more likely to obtain degrees in professional and doctoral programs compared with the other two majors, with Humanities & Social Science majors having an advantage in professional programs (−2.98) and Science, Math & Engineering majors having an advantage in doctoral programs (−4.87). After controlling for unobserved heterogeneity, Science, Math & Engineering majors become the most likely to obtain master’s degrees (−1.9). On the other hand, Education majors become the most likely to obtain MBA (1.84) and professional (9.96) degrees. Overall, degree attainment probability decreases from master’s programs to doctoral programs, indicating that higher degree programs on average take longer to finish.

Table 2.5 reports the estimated transition costs. The cost of entering a new graduate program is large. The basic model estimates the cost to be $169,370 while the extended model gives a smaller value at $22,500. The extended model estimate is more in line with Lee (2005), who estimates that school re-entry costs are $34,449 for males and $28,083 for females. The cost of moving into work, on the other hand, is negative and very large according to both the basic model and extended model estimates. Thus there appears to be a huge “non-pecuniary” benefit associated with moving into working, whether it is from schooling or home production. That such benefit can amount to close to a million dollars is surprising and perhaps reflect the fact that my model does not include experience and age effects in the earnings equation

\[ \text{9} \text{The estimates should be taken with a grain of salt. For example, for a male student with an undergraduate GPA between 3.0 and 3.5, who majored in Humanities & Social Science in college, his degree attainment probabilities estimated by the basic model are 0.64, 0.016, 0.23 and 0.025 respectively for master’s MBA, professional and doctoral programs, which translate into mean completion times of 1.55, 61.95, 4.35 and 40.25 years.} \]
### Table 2.6: Estimated Type-specific Preferences

<table>
<thead>
<tr>
<th></th>
<th>Type 1</th>
<th>Type 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business major</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education major</td>
<td>100,000</td>
<td>-431,250</td>
</tr>
<tr>
<td>Humanities &amp; Social Science major</td>
<td>156,250</td>
<td>12,500</td>
</tr>
<tr>
<td>Science, Math &amp; Engineering major</td>
<td>43,750</td>
<td>-6,250</td>
</tr>
<tr>
<td><strong>Second Stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s program enrollment</td>
<td>43,750</td>
<td>-6,250</td>
</tr>
<tr>
<td>MBA enrollment</td>
<td>-481,250</td>
<td>18,750</td>
</tr>
<tr>
<td>Professional program enrollment</td>
<td>-281,250</td>
<td>-37,500</td>
</tr>
<tr>
<td>Doctoral program enrollment</td>
<td>-143,750</td>
<td>-43,750</td>
</tr>
<tr>
<td>Work</td>
<td>743,750</td>
<td>893,750</td>
</tr>
<tr>
<td>Home production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(reference)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Estimates are in $. and assumes each individual earn the same wage in each period. Since wage typically grow with labor market experience, the model assumptions can lead to under-estimated lifetime earnings. The large estimated benefit of moving into work may thus reflect the magnitude of underestimated lifetime payoffs associated with work.

Table 2.6 reports type-specific preferences obtained from the extended model. There are large difference in preferences between the two types. The type-1 individual strongly prefers to major in Education or Humanities & Social Science in college. The type-2 individual, on the other hand, has a strong disutility associated with majoring in Education and has similar preference toward the other three majors while most preferring Humanities & Social Science. The utility differences between the most favored subject and the least favored one are in the range of $100,000 for both types. After college, the type-1 individual strongly prefers to work. Among graduate programs, she most prefers to enroll in master’s programs and least prefers to enroll in MBA. Like the type-1 individual, the type-2 individual strongly prefers to work after col-
Table 2.7: Estimated Type Proportions

<table>
<thead>
<tr>
<th>SAT math</th>
<th>SAT verbal</th>
<th>Gender</th>
<th>Probability of being Type 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>low</td>
<td>male</td>
<td>.719</td>
</tr>
<tr>
<td>low</td>
<td>high</td>
<td>male</td>
<td>.688</td>
</tr>
<tr>
<td>high</td>
<td>low</td>
<td>male</td>
<td>.087</td>
</tr>
<tr>
<td>high</td>
<td>high</td>
<td>male</td>
<td>.278</td>
</tr>
<tr>
<td>low</td>
<td>low</td>
<td>female</td>
<td>.766</td>
</tr>
<tr>
<td>low</td>
<td>high</td>
<td>female</td>
<td>.996</td>
</tr>
<tr>
<td>high</td>
<td>low</td>
<td>female</td>
<td>.277</td>
</tr>
<tr>
<td>high</td>
<td>high</td>
<td>female</td>
<td>.389</td>
</tr>
</tbody>
</table>

However, out of the four types of graduate programs, she most prefers to enroll in MBA programs and least prefers to enroll in doctoral programs. For both types, staying home is a more favorable option than enrolling in either professional or doctoral programs. The type-1 individual, however, values home production more in comparison with work than the type-2 individual. The type-1 individual also has a stronger disutility associated with attending doctoral and professional programs than the type-2 individual in comparison with staying home. The utility differences between the most favored alternative and the least favored one in the second stage are close to $1 million for both types. The very high utility associated with working may again reflect the magnitude of underestimated lifetime earnings from work.

Table 2.7 reports the estimated proportions of the two types of individuals conditional their gender and SAT test scores. The results show a strong sorting of types according math abilities. Individuals with low SAT math scores are much more likely to be type-1 individuals, while those with high SAT math scores are predominantly type-2. Given math ability, high SAT verbal scores raise the probability of being type-1 slightly by about 10–20 percentage points. Thus, looking back at type preferences, we can see that type-1 individuals,
Table 2.8: Goodness of Fit (a): Major Choice

<table>
<thead>
<tr>
<th></th>
<th>Business</th>
<th>Education</th>
<th>Humanities &amp; Social Science</th>
<th>Science, Math &amp; Engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td>.259</td>
<td>.132</td>
<td>.358</td>
<td>.251</td>
</tr>
<tr>
<td>Basic</td>
<td>.278</td>
<td>.201</td>
<td>.244</td>
<td>.278</td>
</tr>
<tr>
<td>Extended</td>
<td>.230</td>
<td>.113</td>
<td>.403</td>
<td>.254</td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td>.300</td>
<td>.060</td>
<td>.328</td>
<td>.312</td>
</tr>
<tr>
<td>Basic</td>
<td>.294</td>
<td>.189</td>
<td>.242</td>
<td>.275</td>
</tr>
<tr>
<td>Extended</td>
<td>.276</td>
<td>.082</td>
<td>.378</td>
<td>.264</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td>.226</td>
<td>.191</td>
<td>.383</td>
<td>.201</td>
</tr>
<tr>
<td>Basic</td>
<td>.264</td>
<td>.210</td>
<td>.246</td>
<td>.280</td>
</tr>
<tr>
<td>Extended</td>
<td>.193</td>
<td>.139</td>
<td>.423</td>
<td>.245</td>
</tr>
</tbody>
</table>

who have relatively low math ability, strongly prefer to major in Education or Humanities & Social Science in college, derive higher utility from home production in comparison with work after college, and, out of the four types of graduate programs, prefer to enroll in master’s programs and have strong disutility associated with attending the other three types of programs. Type-2 individuals, who have relatively high math abilities, prefer non-Education majors in college, and, out of the four types of graduate programs, most prefer to enroll in MBA after college and, unlike type-1, do not have a strong disutility associated with attending the other three types of programs.

Table 2.8-2.10 shows the goodness of fit for both the basic and the extended models. Table 2.8 compares the predicted and the observed distribution of major choices. We can see that the extended model provides a better fit. In particular, the basic model tends to over-predict the percentage of individuals majoring in Education and under-predict the percentage of those in Humanities & Social Science. Table 2.9 compares the predicted and the observed percentages of individuals choosing each of the six alternatives in the interview year 1997 and 2003. Again, the extended model provides a better fit. In particular, the basic model gives almost identical predictions of the percent-
ages of individuals enrolling in each of the four graduate programs. This leads to, among other things, a significant under-prediction of the master’s programs enrollment rate. The extended model, however, is also less successful in matching the the master’s programs enrollment rate and the home production rate in 2003, over-predicting the former and under-predicting the latter. Table 2.10 compares the predicted and the observed percentages of individuals with no graduate degree and those with most recent degree in each of the four types of graduate programs in 1997 and 2003. Both models are quite successful in matching the observed degree distribution, although the extended model under-predicts the master’s degree attainment rate and over-predicts that of

### Table 2.9: Goodness of Fit (b): Activity

<table>
<thead>
<tr>
<th></th>
<th>1997</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>2003</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Master’s</td>
<td>MBA</td>
<td>Professional</td>
<td>Doctoral</td>
<td>Work</td>
<td>Home</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td>.077</td>
<td>.016</td>
<td>.022</td>
<td>.020</td>
<td>.805</td>
<td>.059</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>.011</td>
<td>.010</td>
<td>.010</td>
<td>.010</td>
<td>.899</td>
<td>.061</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended</td>
<td>.064</td>
<td>.016</td>
<td>.012</td>
<td>.015</td>
<td>.828</td>
<td>.065</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.044</td>
<td>.015</td>
<td>.008</td>
<td>.016</td>
<td>.799</td>
<td>.119</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>.021</td>
<td>.021</td>
<td>.019</td>
<td>.019</td>
<td>.795</td>
<td>.125</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended</td>
<td>.063</td>
<td>.015</td>
<td>.012</td>
<td>.016</td>
<td>.830</td>
<td>.065</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2.10: Goodness of Fit (c): Degree Attainment

<table>
<thead>
<tr>
<th></th>
<th>1997</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>2003</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bachelor’s</td>
<td>Master’s</td>
<td>MBA</td>
<td>Professional</td>
<td>Doctoral</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td>.885</td>
<td>.089</td>
<td>.007</td>
<td>.017</td>
<td>.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>.884</td>
<td>.080</td>
<td>.014</td>
<td>.017</td>
<td>.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended</td>
<td>.895</td>
<td>.066</td>
<td>.020</td>
<td>.019</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td>.739</td>
<td>.152</td>
<td>.051</td>
<td>.039</td>
<td>.020</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>.737</td>
<td>.139</td>
<td>.046</td>
<td>.053</td>
<td>.026</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended</td>
<td>.733</td>
<td>.156</td>
<td>.053</td>
<td>.048</td>
<td>.010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
the MBA degree in 1997.

Finally, to visualize the evolution of the choice and degree distributions of the entire cohort, Figure 2.5 and 2.6 plot respectively the enrollment, work and home production rates and the degree attainment rates of the cohort over the 40 years after college, simulated by the extended model. Figure 2.6 shows a smooth increasing rate of attainment for each type of graduate degree, culminating in about 35 percent of the cohort attaining a master’s degree, 13 percent attaining an MBA degree, 11 percent attaining a professional degree, and 7 percent attaining a doctoral degree by the time they retire. Figure 2.5 shows almost zero program enrollment rates and home production rate in year 1, a big jump in year 2, a much smaller drop in year 3, and a smooth declining trend over the next 36 years (except the doctoral program enrollment rate, which does not seem to decline), before a slight jump again in year 39 and then a drop to zero in year 40. For work rates, the pattern is reversed. In year 1, almost 100 percent of the cohort choose to work. The rate drops in year 2, followed by a near constant trend, before jumping back to almost 100 percent in year 40.

The evolution of the enrollment, work and home production rates shown in Figure 2.5 is unusual and reveals several limitations of the model and data. The slow decline of program enrollment rates over the years reflect the lack of an age or time component in the utility function. Although the finite horizon by itself leads to a declining enrollment patterns with time, with i.i.d. utility shocks, the model generates persistently high school enrollment rates. The unusual jumps and drops in enrollment, work and home production rates at the beginning and the end of the periods are due to the artifact of the school and work transition costs. Because the model estimated work transition “benefit” is very large ($808,750), and much higher than the estimated school transition cost ($22,500), individuals have the incentive to exit work in year 39 just to transition back to work in year 40 to reap the work transition benefits, hence the observed jump in program enrollment and home production rates in year 39. Similarly, due to the large work transition benefits, the majority of individuals who are enrolled or staying home in a period will transition
into work the next period, hence the observed almost 100 percent work rate in year 1 (since everybody is in school in the year before). To rectify such irregularities, better data are needed to better estimate the wage function and account for the effects of work experience and age, so that the estimated work transition cost does not reflect underestimated lifetime earnings from work.

2.5 Conclusion

In this paper, I develop a dynamic discrete choice model to estimate the post-college educational investment and labor supply decisions of college graduates. The model is fitted to a nationally representative sample of college graduates in 1992-93. My results find strong sorting by math ability in individuals’ taste for college majors and post-college choices. Individuals with high math ability favor non-Education majors in college, prefer MBA (out of the four types of graduate programs) after college, and have higher utility toward enrolling in professional and doctoral programs than individuals with low math ability, who most favor Education majors in college, prefer master’s programs (out of the four types of graduate programs) after college, and value home production in comparison with work more than individuals with high math ability. The findings regarding the degree attainment probabilities associated with different graduate programs have a less clear pattern. The results, however, do show unambiguously that being female reduces the probability of attaining a degree in any given period, while having a higher GPA and having already attained a graduate degree significantly raises the probability of degree attainment. Finally, my results reveal a significant positive cost of enrolling in a new graduate program and a significant benefit associated with entering the workplace. The workplace transition benefit, however, is likely the result of underestimated lifetime earnings associated with work, the correction of which awaits better data in future work. Overall, my results suggest that significant transition costs and individual preferences may be the key elements that prevent graduate program enrollment to respond to the large wage premiums associated with post-college education as observed over the last 30 years.
2.6 Appendix

2.6.1 Figures

Figure 2.1: Major Distribution by Gender

Figure 2.2: Highest Graduate Enrollment by 2003 by Gender
Figure 2.3: Highest Graduate Enrollment by 2003 by GPA

Figure 2.4: Highest Graduate Enrollment by 2003 by Undergraduate Major
Figure 2.5: Simulated Activities By Year
Figure 2.6: Simulated Degree Attainment By Year
### 2.6.2 Tables

Table 2.11: Conditional Type Probability Parameters for the Extended Model

<table>
<thead>
<tr>
<th>SAT math</th>
<th>SAT verbal</th>
<th>Gender</th>
<th>Probability of being Type 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>low</td>
<td>male</td>
<td>$\rho_1$</td>
</tr>
<tr>
<td>low</td>
<td>high</td>
<td>male</td>
<td>$\rho_2$</td>
</tr>
<tr>
<td>high</td>
<td>low</td>
<td>male</td>
<td>$\rho_3$</td>
</tr>
<tr>
<td>high</td>
<td>high</td>
<td>male</td>
<td>$\rho_4$</td>
</tr>
<tr>
<td>low</td>
<td>low</td>
<td>female</td>
<td>$\rho_5$</td>
</tr>
<tr>
<td>low</td>
<td>high</td>
<td>female</td>
<td>$\rho_6$</td>
</tr>
<tr>
<td>high</td>
<td>low</td>
<td>female</td>
<td>$\rho_7$</td>
</tr>
<tr>
<td>high</td>
<td>high</td>
<td>female</td>
<td>$\rho_8$</td>
</tr>
</tbody>
</table>
Table 2.12: Aggregate Moments

<table>
<thead>
<tr>
<th>Aggregate Moment</th>
<th>Number of Conditional Moments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Master’s program enrollment rate</td>
<td>$3 \times 5 \times 4 \times 4 \times 2$</td>
</tr>
<tr>
<td>MBA program enrollment rate</td>
<td>$3 \times 5 \times 4 \times 4 \times 2$</td>
</tr>
<tr>
<td>Professional program enrollment rate</td>
<td>$3 \times 5 \times 4 \times 4 \times 2$</td>
</tr>
<tr>
<td>Doctoral program enrollment rate</td>
<td>$3 \times 5 \times 4 \times 4 \times 2$</td>
</tr>
<tr>
<td>Employment rate</td>
<td>$3 \times 5 \times 4 \times 4 \times 2$</td>
</tr>
<tr>
<td>Staying home rate</td>
<td>$3 \times 5 \times 4 \times 4 \times 2$</td>
</tr>
<tr>
<td>Master’s degree attainment rate</td>
<td>$3 \times 4 \times 4 \times 2$</td>
</tr>
<tr>
<td>MBA degree attainment rate</td>
<td>$3 \times 4 \times 4 \times 2$</td>
</tr>
<tr>
<td>Professional degree attainment rate</td>
<td>$3 \times 4 \times 4 \times 2$</td>
</tr>
<tr>
<td>Doctoral degree attainment rate</td>
<td>$3 \times 4 \times 4 \times 2$</td>
</tr>
<tr>
<td>Undergraduate major distribution</td>
<td>$4 \times 4 \times 2$</td>
</tr>
<tr>
<td>Employment and enrollment trajectory</td>
<td>$3 \times 3 \times 3 \times 4 \times 4 \times 2$</td>
</tr>
</tbody>
</table>

**Note:** 1. The first six choice moments are conditioned on year (3), most recent graduate degree (5), undergraduate major (4), GPA level (4), and gender (2); 2. The next four moments are conditioned on year (3), undergraduate major (4), GPA level (4), and gender (2); 3. Undergraduate major distribution is conditioned on SAT math quartile (4), SAT verbal quartile (4), and gender (2); 4. Employment and enrollment trajectory (the 27 possible combinations of being enrolled, employed, or staying home in the three interview years) is conditioned on undergraduate major (4), GPA level (4), and gender (2); Not all listed moments are available. A number of them are missing due to insufficient observations.
Table 2.13: Major Distribution by SAT scores

<table>
<thead>
<tr>
<th>Major Field of Study</th>
<th>SAT verbal=1</th>
<th>SAT verbal=2</th>
<th>SAT verbal=3</th>
<th>SAT verbal=4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Business</td>
<td>Education</td>
<td>Humanities &amp; Social Science</td>
<td>Science, Math &amp; Engineering</td>
</tr>
<tr>
<td>SAT math=1</td>
<td>25%</td>
<td>15%</td>
<td>44%</td>
<td>15%</td>
</tr>
<tr>
<td>SAT math=2</td>
<td>30%</td>
<td>18%</td>
<td>35%</td>
<td>17%</td>
</tr>
<tr>
<td>SAT math=3</td>
<td>37%</td>
<td>17%</td>
<td>22%</td>
<td>24%</td>
</tr>
<tr>
<td>SAT math=4</td>
<td>35%</td>
<td>7%</td>
<td>15%</td>
<td>43%</td>
</tr>
<tr>
<td>SAT math=1</td>
<td>25%</td>
<td>13%</td>
<td>46%</td>
<td>16%</td>
</tr>
<tr>
<td>SAT math=2</td>
<td>30%</td>
<td>14%</td>
<td>33%</td>
<td>23%</td>
</tr>
<tr>
<td>SAT math=3</td>
<td>30%</td>
<td>13%</td>
<td>31%</td>
<td>26%</td>
</tr>
<tr>
<td>SAT math=4</td>
<td>16%</td>
<td>13%</td>
<td>27%</td>
<td>44%</td>
</tr>
<tr>
<td>SAT math=1</td>
<td>25%</td>
<td>12%</td>
<td>43%</td>
<td>20%</td>
</tr>
<tr>
<td>SAT math=2</td>
<td>21%</td>
<td>9%</td>
<td>52%</td>
<td>18%</td>
</tr>
<tr>
<td>SAT math=3</td>
<td>27%</td>
<td>10%</td>
<td>34%</td>
<td>30%</td>
</tr>
<tr>
<td>SAT math=4</td>
<td>23%</td>
<td>8%</td>
<td>33%</td>
<td>36%</td>
</tr>
<tr>
<td>SAT math=1</td>
<td>16%</td>
<td>18%</td>
<td>55%</td>
<td>13%</td>
</tr>
<tr>
<td>SAT math=2</td>
<td>17%</td>
<td>13%</td>
<td>61%</td>
<td>9%</td>
</tr>
<tr>
<td>SAT math=3</td>
<td>19%</td>
<td>10%</td>
<td>50%</td>
<td>21%</td>
</tr>
<tr>
<td>SAT math=4</td>
<td>10%</td>
<td>6%</td>
<td>45%</td>
<td>38%</td>
</tr>
</tbody>
</table>

Note: Table entries report undergraduate major distribution by combination of SAT math and verbal test score quartiles. SAT math quartiles are: 1st quartile=[0, 430), 2nd quartile=[430, 510), 3rd quartile=[510, 600), 4th quartile=[600, 800]. SAT verbal quartiles are: 1st quartile=[0, 390), 2nd quartile=[390, 460), 3rd quartile=[460, 520), 4th quartile=[520, 800].
Table 2.14: Highest Graduate Enrollment 2003 by Highest Degree Attainment 1997

<table>
<thead>
<tr>
<th>Highest Degree by 1997</th>
<th>Highest Graduate Enrollment by 2003</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Enrollment</td>
</tr>
<tr>
<td>Bachelor’s</td>
<td>67.1%</td>
</tr>
<tr>
<td>Master’s</td>
<td>74.2%</td>
</tr>
<tr>
<td>MBA</td>
<td>97.9%</td>
</tr>
<tr>
<td>Professional</td>
<td></td>
</tr>
<tr>
<td>Doctoral</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Degree programs are ranked from lowest to highest in the following order: Bachelor’s, Master’s, MBA, Professional, Doctoral
Chapter 3

Bayesian Estimation of Dynamic Models of Imperfect Competition
Abstract

Estimation of dynamic models of imperfect competition is computationally challenging due to the need of repeatedly solving the dynamic programming problems of forward-looking agents and finding the equilibria of dynamic multi-agent games. This paper develops a Bayesian Markov chain Monte Carlo procedure to handle the computational burden using an algorithm developed by Imai, Jain, and Ching (2009) for inference in single-agent dynamic discrete choice models. Three algorithms are proposed to extend the Imai-Jain-Ching procedure to the estimation of dynamic models of imperfect competition. The performance of the algorithms is tested using Monte Carlo simulations of an entry and exit model.

3.1 Introduction

Structural estimations of dynamic models have been known for their computational demands, which have to date limited their use to relatively simple exercises or models heavily tailored for their convenience. One of the major sources of computational burden in doing structural estimation of dynamic models lies in the necessity of solving the dynamic programming (DP) problems of forward-looking agents. In the context of Maximum Likelihood estimation (MLE), the likelihood functions are based on the explicit solution of these DP problems. Estimations, therefore, have to proceed by iterating on a set of Bellman equations until convergence for all values of model parameters considered by the estimator. Various estimators have been proposed hitherto to lessen the computational burden in this process, but with the notable exception of Bajari, Benkard, and Levin (2007), most of them remain MLE estimators with limited improvement in computational efficiency.

This paper tries to build on the exciting new development in the estimation of dynamic models made by Imai, Jain, and Ching (2007). In their paper, the authors have proposed a Bayesian method for estimating dynamic discrete choice models. The method, dubbed by the authors as "Bayesian Dynamic
Programming” (BDP), is a combination of the DP solution algorithm and the Markov Chain Monte Carlo (MCMC) algorithm, and solves DP problems and estimates model parameters simultaneously. The authors have demonstrated significant increases in computational efficiency using the BDP estimator as compared to conventional Bayesian MCMC and MLE estimators. The key is to iterate on Bellman equations only once while proposing new parameter values during each estimation iteration, and use local nonparametric regression to smooth value functions across different parameter values, so that the values of value functions in previous iterations can be used to infer those in later iterations, thereby effectively keeping value functions continuously updated while the algorithm moves around the parameter space. Another useful feature of the BDP algorithm, as a result of the same technique, is that it allows the number of grid points on the state space to effectively increase with the number of estimation iterations, even though value functions only need to be calculated on a small number of them each iteration, thus making it a powerful method in dealing with models with continuous state space or, potentially, tackling the ”curse of dimensionality,” i.e. the problem that computational burden increases exponentially with the size of a discrete state space.

The BDP algorithm, as proposed in Imai, Jain, and Ching (2007), is for solving single-agent dynamic discrete choice problems\footnote{or for non-interacting multi-agent environments, in which each agent’s choice does not depend on (her expectation of) other agents’ states and actions.}. In this paper, we extend their procedure to estimating dynamic models of imperfect competition. Compared to single agent problems, dynamic models of imperfect competition involve DP problems that are more complex (each agent’s choice of action depends on her belief about other agents’ private information and choice probabilities) and require both value functions and choice probability functions to converge in equilibrium (what Aguirregabiria and Mira (2007) call a ”coupled fixed-point problem”). In this paper, we provide three different algorithms that extend the BDP procedure to the estimation of the dynamic models of imperfect competition. The algorithms differ from each other in the way equilibrium choice probability functions are calculated. We also describe how
to modify them to allow random grid generation. Finally, to test our proposed algorithms and compare their performance, we perform Monte Carlo simulations on a model of entry and exit adopted from Dunne, Roberts, and Xu (2006).

The rest of the paper proceeds as follows. In section 2, we lay out the basic structure of a dynamic model and discuss the traditional MLE based methods of estimation, providing motivation for the BDP estimator. Section 3 and 4 summarize the BDP algorithm and an extension of the algorithm that allows for random grid generation. Section 5 conducts several toy experiments to illustrate the workings of the BDP algorithms as well as discussing in detail some practical issues in their implementation. Section 6 provides our extensions of the BDP algorithm to estimating dynamic models of imperfect competition. Section 7 conducts a Monte Carlo simulation to assess the performance of our extension algorithms. Section 8 concludes the paper.

### 3.2 Motivation

Consider first a single agent dynamic programming problem:

\[
V(s, \varepsilon|\theta) = \max_{a \in A} \left\{ \pi(s, a, \varepsilon_a|\theta) + \beta \sum_{s' \in \Psi} \left( \int V(s', \varepsilon'|\theta)dF_{\varepsilon'}(.)|\theta \right) \cdot g(s'|s, a, \theta) \right\}
\]

(3.1)

where \(s\) is the observed state of the agent and belongs to a finite state space \(\Psi = \{1, ..., S\}\). \(A = \{1, ..., J\}\) is a finite set of actions the agent chooses from. Each period, the agent receives a private shock \(\varepsilon_a\) depending on her choice of action \(a\) and receives a return of \(\pi(s, a, \varepsilon_a, \theta)\). The vector of choice-contingent shocks, \(\varepsilon = (\varepsilon_1, ..., \varepsilon_J)\), are i.i.d. distributed according to \(F_{\varepsilon}(.,|\theta)\), and state transition follows a Markov process with transition probability \(g(s'|s, a, \theta)\). Let the model be parametrized by \(\theta \in \Theta\), where \(\Theta\) is the parameter space.

Consider a time series data \(Z^d = \{s^d_t, a^d_t\}_{t=1}^T\) where \(s^d_t, a^d_t\) are respectively the state the agent is in and the choice of action she makes in period \(t\). (The superscript \(d\) denotes a data variable). We would like to estimate the true parameter \(\theta_0\) that generates \(Z^d\).
To this end the MLE approach is to solve the following problem:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \log L(Z^d|EV(.|\theta), \theta)$$ (3.2)

where

$$L(Z^d|EV(.|\theta), \theta) = \Pi_t \Pr (a^t = \sigma (s^t, \varepsilon|EV(.|\theta), \theta))$$

and we define

$$EV (s|\theta) \equiv E_{\varepsilon} [V(s, \varepsilon|\theta)]$$

$$\sigma (s, \varepsilon|EV(.|\theta), \theta) \equiv \arg \max_{a \in A} \left\{ \pi (s, a, \varepsilon_a|\theta) + \beta \sum_{s' \in \Psi} EV(s'|\theta)g(s'|s, a, \theta) \right\}$$

When the choice probability function $\Pr (a = \sigma (s, \varepsilon|EV(.|\theta), \theta))$ has a closed-form expression and can be differentiated with respect to $\theta$ and $EV(.|\theta)$, (3.2) can be solved as a constrained optimization problem. Let $f (EV, \theta) \equiv L(Z^d|EV(.|\theta), \theta)$, the problem (3.2) is equivalent to

$$\max_{\theta \in \Theta} f (EV, \theta) - \lambda^T (EV - T_\theta (EV))$$ (3.3)

where $\lambda$ is the Lagrangian multiplier and $T_\theta$ is the contraction mapping defined by

$$[T_\theta (EV)] (s) = \int \sum_{a \in A} \left\{ \pi (s, a, \varepsilon_a|\theta) + \beta \sum_{s' \in \Psi} EV(s'|\theta)g(s'|s, a, \theta) \right\} \Pr (a = \sigma (s, \varepsilon|EV, \theta)) dF_{\varepsilon} (.|\theta)$$

Standard constrained optimization algorithm turns (3.3) into a set of first-order conditions:

$$\nabla_\theta f (EV, \theta) - \nabla_\theta (EV - T_\theta (EV))^T \lambda = 0$$

$$\nabla_{EV} f (EV, \theta) - \nabla_{EV} (EV - T_\theta (EV))^T \lambda = 0$$

$$EV - T_\theta (EV) = 0$$

For example, if $\pi (s, a, \varepsilon_a|\theta) = \pi (s, a|\theta) + \varepsilon_a$, and $F_{\varepsilon}$ is a type I extreme value distribution: $\varepsilon \sim \exp (-\exp (-\varepsilon))$, then

$$\Pr (a = \sigma (s, \varepsilon|EV, \theta)) = \frac{\exp \{ \pi (s, a|\theta) + \beta \int EV(s'|\theta)g(s'|s, a, \theta)ds' \}}{\sum_{a' \in A} \exp \{ \pi (s, a'|\theta) + \beta \int EV(s'|\theta)g(s'|s, a', \theta)ds' \}}$$

which can be differentiated with respect to both $\theta$ and $EV$. 

115
and applies the Newton–Raphson method to find their solution\(^3\). Doing so achieves a quadratic rate of convergence and avoids the separate calculations of the expected value function \(EV (.|\theta)\).

When a closed-form solution to \(Pr \left( a_t^d = \sigma \left( s_t^d, \varepsilon | EV (.|\theta), \theta \right) \right)\) is not available, however, (3.2) has to be solved as an unconstrained optimization problem, in which case, \(EV(s|\theta)\) has to be calculated for all points on the state space and for each \(\theta\) at which the likelihood function \(L(Z^d|EV(.|\theta), \theta)\) is evaluated (such as at each Newton–Raphson iteration in an unconstrained optimization algorithm). For each \(s\) and \(\theta\), the calculation of \(EV(s|\theta)\) involves iterating on the Bellman equation (3.1) until convergence. It is this need to repeatedly solve the dynamic programming problem that has become the computational albatross on structural estimation of dynamic models.

### 3.3 Bayesian Dynamic Programming (BDP)

Instead of the MLE approach, the Bayesian Dynamic Programming (BDP) method proposed by Imai, Jain, and Ching (2007) offers a way of doing Bayesian inference on dynamic models that combines the dynamic programming (DP) solution algorithm (i.e. iteration on the Bellman equation) with the Markov Chain Monte Carlo (MCMC) algorithm to effectively sample the model parameter \(\theta\) from its posterior distribution:

\[
Pr \left( \theta | Z^d \right) = \rho(\theta) Pr \left( Z^d | \theta \right) \\
\propto \rho(\theta)L(Z^d | EV(.|\theta), \theta)
\]

where we assume the prior distribution of \(\theta\) is \(\rho(\theta)\).

The basic algorithm of BDP works as follows: A sequence of parameters \(\{\theta^{(r)}\}_{r=1}^t\) is constructed through iteration. In each iteration \(t\), given \(\theta^{(t-1)}\), a candidate parameter vector \(\theta^{*(t)}\) is drawn from a proposal density \(q \left( \theta^{(t-1)}, \theta^{*(t)} \right)\). Then, we update the value function \(V(s, \varepsilon | \theta)\) at \(\theta^{*(t)}\) and

\(^3\)For more discussion on the use of constrained optimization algorithm to estimate dynamic models, see Su and Judd (2007)
thereby effectively keeping the value function continuously updated while the function in preceding iterations can be used to infer those in later iterations, \( \theta \) the value function with respect to different regression techniques, here a Nadaraya–Watson Kernel estimator, to evaluate the proposed candidate problem for each \( K \), where

\[
\{ \lambda, V_s \}_{s \in \Psi} \text{ the Bellman equation (3.1) for } s = 1, \ldots, S:
\]

\[
V^{(t)} (s, \varepsilon^{(t)} | \theta^{(t)}) = \max_a \left\{ \pi (s, a, \varepsilon_a^{(t)} | \theta^{(t)}) + \beta \sum_{s' \in \Psi} EV^{(t)} (s'| \theta^{(t)}) g(s'|s, a, \theta^{(t)}) \right\} \tag{3.4}
\]

Since we do not have the value of \( EV^{(t)} (\cdot | \theta^{(t)}) \), we approximate it by averaging over the value functions of \( N(t) \) most recent iterations. Let \( H(t) \equiv \{ \theta^{(r)}, V(r) (\cdot, \varepsilon^{(r)} | \theta^{(r)}) \}_{r=1}^{t-1} \) be the history of candidate parameters and value functions up to iteration \( t \), we set

\[
EV^{(t)} (s | \theta^{(t)}) = \sum_{j=1}^{N(t)} V^{(t-j)} (s, \varepsilon^{(t-j)} | \theta^{(t-j)}) \cdot \frac{K_h (\theta^{(t)} - \theta^{(t-j)})}{\sum_{k=1}^{N(t)} K_h (\theta^{(t)} - \theta^{(t-k)})} \tag{3.5}
\]

, where \( K_h (\cdot) \) is a multivariate kernel with bandwidth \( h \).

Then we calculate the Metropolis-Hastings acceptance probability

\[
\lambda (\theta^{(t-1)}, \theta^{(t)}) = \min \left\{ 1, \frac{\rho (\theta^{(t-1)}) L (Z^d | EV^{(t)} (\cdot | \theta^{(t-1)}) , \theta^{(t)} ) q (\theta^{(t-1)}, \theta^{(t-1)})}{\rho (\theta^{(t-1)}) L (Z^d | EV^{(t)} (\cdot | \theta^{(t-1)}) , \theta^{(t-1)}) q (\theta^{(t-1)}, \theta^{(t)})} \right\} \tag{3.6}
\]

, where \( EV^{(t)} (\cdot | \theta^{(t-1)}) \) is approximated in the same way as in (3.5)\(^4\), and accept the candidate parameter \( \theta^{(t)} \) with probability \( \lambda (\theta^{(t-1)}, \theta^{(t)}) \)\(^5\).

Step (3.5) is the key to the BDP algorithm. Instead of fully solving the DP problem for each \( s \) and \( \theta^{(t)} \), we iterate on the Bellman equation only once at the proposed candidate \( \theta^{(t)} \) during each iteration, and use local nonparametric regression techniques, here a Nadaraya–Watson Kernel estimator, to evaluate the value function with respect to different \( \theta \), so that the values of the value function in preceding iterations can be used to infer those in later iterations, thereby effectively keeping the value function continuously updated while the

\[
EV^{(t)} (s | \theta^{(t-1)}) = \sum_{j=1}^{N(t)} V^{(t-j)} (s, \varepsilon^{(t-j)} | \theta^{(t-j)}) \cdot \frac{K_h (\theta^{(t-1)} - \theta^{(t-j)})}{\sum_{k=1}^{N(t)} K_h (\theta^{(t-1)} - \theta^{(t-k)})}
\]

\(^4\)i.e.,

\[
EV^{(t)} (s | \theta^{(t-1)}) = \sum_{j=1}^{N(t)} V^{(t-j)} (s, \varepsilon^{(t-j)} | \theta^{(t-j)}) \cdot \frac{K_h (\theta^{(t-1)} - \theta^{(t-j)})}{\sum_{k=1}^{N(t)} K_h (\theta^{(t-1)} - \theta^{(t-k)})}
\]

\(^5\)i.e. let \( \theta^{(t)} = \theta^{(t)} \) with probability \( \lambda (\theta^{(t-1)}, \theta^{(t)}) \) and let \( \theta^{(t)} = \theta^{(t-1)} \) with probability \( 1 - \lambda (\theta^{(t-1)}, \theta^{(t)}) \).
algorithm moves around the parameter space. In this way, the DP problem is solved at the same time as the posterior distribution of \( \theta \) is generated.

What if, instead of (3.5), we calculate \( EV(s|\theta^*(t)) \) by duly solving the DP program? Then

\[
EV(t)(s|\theta^*(t)) = \int V(t)(s, \varepsilon^{(t)}|\theta^*(t)) \, dF_\varepsilon(\cdot|\theta^*(t)) \tag{3.7}
\]

In most cases, there is no ready analytical solution to the integration problem in (3.7). Assume \( F_\varepsilon(\cdot|\theta^*(t)) \) has a known functional form and can be easily sampled from, then the integral in (3.7) is typically evaluated by Monte Carlo integration: We draw, independently, \( \varepsilon^{(t)}_m \sim F_\varepsilon(\cdot|\theta^*(t)) \) for \( m = 1, \ldots, M \) and then let

\[
EV(t)(s|\theta^*(t)) = \frac{1}{M} \sum_{m=1}^{M} V(t)(s, \varepsilon^{(t)}_m|\theta^*(t)) \tag{3.8}
\]

Equations (3.4) and (3.8) would then have to be iterated until convergence.

Therefore, compared to a conventional Bayesian MCMC estimation of the dynamic model, in which the DP problem is solved for every \( s \) and \( \theta^*(t) \), the BDP algorithm as presented above represents a two-fold saving in computation by obviating both repeated iterations of the Bellman equation, and the need of doing Monte Carlo integration in (3.8). The latter saving is due to the mechanism that, in BDP, we only need to draw a single \( \varepsilon^{(t)} \) in each iteration. In this sense, we can also think of (3.5) as conceptually a combination of two steps: first, it is a Nadaraya-Watson Kernel regression that maps \( \{V(r)(\cdot, \varepsilon^{(r)}|\theta^*(r))\}_{r=t-N(t)}^{t-1} \) into \( \{V(t)(\cdot, \varepsilon^{(t)}|\theta^*(t))\}_{r=t-N(t)}^{t-1} \); second, it is a Monte Carlo integration that averages \( \{V(t)(\cdot, \varepsilon^{(r)}|\theta^*(t))\}_{r=t-N(t)}^{t-1} \) over \( \varepsilon \) to generate \( EV(t)(s|\theta^*(t)) \); though the two cannot be separated in the actual algorithm. The added computational cost of BDP is the need to run more iterations for \( \{\theta^{(r)}\}_{r=1}^{t} \) to converge to its posterior distribution due to the DP program being dynamically solved\(^6\), and the need to do local nonparametric

---

\(^6\)So, in a sense, what the BDP algorithm truly saves is not the need of repeatedly iterating the Bellman equation *per se*, but the need of doing so on \( \theta \) far away from \( \theta_0 \). By dynamically updating the value function, the algorithm can quickly locate the area close to \( \theta_0 \).
regressions to calculate $EV^{(t)}(s|\theta^{*(t)})$ and $EV^{(t)}(s|\theta^{(t-1)})$ for $s = 1, \ldots, S$ in each iteration. These regressions can take up the bulk of the computational time of BDP and are, collectively, the bottleneck of the speed of the algorithm.

The chain of parameter values, $\{\theta^{(t)}\}$, generated by the BDP algorithm, is not a Markov chain, since its state transition probabilities depend on calculations involving a variable, $N(t)$, number of proceeding states. Imai, Jain, and Ching (2007) have proved, however, that under the condition that $N(t) \to \infty$ and $t - N(t) \to \infty$, $\theta^{(t)} \to P \tilde{\theta}^{(t)}$, where $\tilde{\theta}^{(t)}$ is a Markov chain generated by $q(\theta^{(t-1)}, \theta^{*})$ and $\lambda (\theta^{(t-1)}, \theta^{*})$. The distribution of the Markov chain $\{\tilde{\theta}^{(t)}\}$ then converges, under regularity conditions, in total variation to the true posterior distribution by standard MCMC asymptotic results. The BDP algorithm, therefore, can also be aptly, though not tersely, named as a ”Dynamic Programming Embedded, Pseudo-Markov Chain Monte Carlo (DPE-PMCMC)” algorithm.

### 3.4 Bayesian Dynamic Programming with Random Grid Generation

When the state space is continuous, a common practice is to discretize it into a finite number of grid points (Rust (1997)). As Su and Judd (2007) demonstrate in their estimation of the Zucher engine replacement model, however, coarser discretizations can produce nontrivial errors in the estimated parameters.

Another useful feature of the BDP is that, much like it allows value functions to be continuously updated as it roams the parameter space, it allows the number of grid points on the state space at which value functions are evaluated to increase at the same time, achieving finer and finer discretization of the state space with each successive iteration. In each iteration, we can generate a new grid point $s^{(t)}$ on the state space and iterate on Bellman equations once only at the proposed candidate $\theta^{*}$ and the new grid point $s^{(t)}$, while using local nonparametric regression to estimate value functions at other grid points and parameter values. Such an algorithm can be called: ”BDP with
Random Grid Generation."

Specifically, if state transition is stochastic and \( g(.|s, a, \theta) \) is continuous, we replace step (3.4) and (3.5) in the basic BDP algorithm with\(^7\):

\[
V^{(t)}(s^{(t)}, \varepsilon^{(t)}|\theta^{*(t)}) = \max_a \left\{ \pi \left( s^{(t)}, a, \varepsilon^{(t)}_a | \theta^{*(t)} \right) + \beta W^{(t)}(s^{(t)}, a|\theta^{*(t)}) \right\} \tag{3.9}
\]

\[
W^{(t)}(s^{(t)}, a|\theta^{*(t)}) = \sum_{j=1}^{N(t)} V^{(t-j)}(s^{(t-j)}, \varepsilon^{(t-j)}|\theta^{*(t-j)}) \cdot \frac{K_h(\theta^{*(t)} - \theta^{*(t-j)}) \cdot g(s^{(t-j)}|s^{(t)}, a, \theta^{*(t)})}{\sum_{k=1}^{N(t)} K_h(\theta^{*(t)} - \theta^{*(t-k)}) \cdot g(s^{(t-k)}|s^{(t)}, a, \theta^{*(t)})} \tag{3.10}
\]

, where

\[
W(s, a|\theta) \equiv E_{s',\varepsilon'}[V(s, \varepsilon|\theta)] = \int \int V(s', \varepsilon'|\theta) g(s'|s, a, \theta) dF_{\varepsilon'}(\cdot|\theta) ds'
\]

In addition, we need to calculate \( W^{(t)}(s, a|\theta^{*(t)}) \) and \( W^{(t)}(s, a|\theta^{(t-1)}) \) for all unique states in \( \{ s^d_t \}_{t=1}^T \) in order to evaluate \( \lambda(t^{(t-1)}, \theta^{*(t)}) \).

If state transition is deterministic, i.e. \( \exists \) a function \( r \) such that \( s' = r(s, a, \theta) \), then (3.4) and (3.5) can be replaced by

\[
V^{(t)}(s^{(t)}, \varepsilon^{(t)}|\theta^{*(t)}) = \max_a \left\{ \pi \left( s^{(t)}, a, \varepsilon^{(t)}_a | \theta^{*(t)} \right) + \beta EV^{(t)}(r(s^{(t)}, a, \theta^{*(t)}) |\theta^{*(t)}) \right\} \tag{3.11}
\]

\[
EV^{(t)}(s^{(t)}|\theta^{*(t)}) \equiv \sum_{j=1}^{N(t)} V^{(t-j)}(s^{(t-j)}, \varepsilon^{(t-j)}|\theta^{*(t-j)}) \cdot \frac{K_{h_s}(s - s^{(t-j)}) \cdot K_{h_\varepsilon}(\theta^{*(t)} - \theta^{*(t-j)})}{\sum_{k=1}^{N(t)} K_{h_s}(s - s^{(t-k)}) \cdot K_{h_\varepsilon}(\theta^{*(t)} - \theta^{*(t-k)})} \tag{3.12}
\]

, where \( h_s \) and \( h_\varepsilon \) denote different Kernel windows. \( EV^{(t)}(s, a|\theta^{*(t)}) \) and \( EV^{(t)}(s, a|\theta^{(t-1)}) \) are likewise calculated for all unique states in \( \{ s^d_t \}_{t=1}^T \).

In this way, value functions can be updated continuously at more and more grid points as the algorithm runs on, gradually mapping out their entire shapes across the state and parameter space.

Imai, Jain, and Ching (2007) make the claim that, when the state space is not continuous, BDP with random grid generation can also help to tackle the

\(^7\)now we keep a history \( H^{(t)} \equiv \{ \theta^{(r)}, s^{(r)}, V^{(r)}(., \varepsilon^{(r)}|\theta^{*(r)}) \}_{r=1}^{t} \) of candidate parameters, value functions and state grid points generated in each successive iteration.
“curse of dimensionality,” i.e. the problem that the number of points in the state space increase exponentially with the dimensionality of the state space, due to the algorithm’s ability to update value functions on as few as a single grid point each iteration. The claim, however, is subject to qualifications.

Although the BDP algorithm with random grid generation only need to iterate on Bellman equations once on a single grid point each iteration, it computes a large number of local nonparametric regressions to infer the values of value functions at other grid points observed in the data in order to calculate the likelihood function. In contrast, the standard BDP algorithm without random grid generation iterates Bellman equations on each point in the state space during each iteration, but does not need to use local nonparametric regression to estimate value functions at any other point in the state space. The computational cost of doing a single iteration on Bellman equation is usually smaller than performing a local nonparametric regression. Therefore, BDP with random grid generation only decreases computational time if the number of states observed in the data is significantly smaller than the number of states in the state space.

3.5 Toy Experiments

We construct several toy experiments to illustrate the workings of the BDP algorithm (with and without random grid generation), as well as consider several niceties in the implementation of the algorithm. In particular, we visualize the process of how value functions are continuously updated at new candidate parameters (and grid points), and how their shapes (contours) are gradually mapped out through iterations. We then compare several nonparametric estimators and test how they impact the performance of the algorithm. Finally, we consider several special cases in which the specifications of the likelihood or the value functions might pose problems for estimation and discuss how to deal with them.
3.5.1 Experiment 1: \( V \) as a linear function of \( \theta \)

3.5.1.1 The Experiment

We first construct a value function that is simply a linear function of \( \theta \). Let the Bellman equation be

\[
V(\theta) = \theta + 0.9 \times V(\theta)
\]

(3.13)

, where \( \theta \in \Theta \equiv \mathbb{R}_+ \). Then \( V(\theta) = 10\theta \). We suppose, however, that we do not know this solution and have to iterate on (3.13) to find out the value of \( V(\theta) \) for every \( \theta \).

We generate a data set \( Z^d \equiv \{ s^d_t \}_{t=1}^{T=1000} \) as follows: Define \( p(\theta) \equiv \min\{ \frac{V(\theta)}{100}, 1 \} \). Let \( \theta_0 = 5 \), such that \( p(\theta_0) = 0.5 \). Then let \( s^d_t \sim_{i.i.d.} \text{Bernoulli}(p(\theta_0)) \). We run the BDP program to estimate \( \theta_0 \) with an initial guess of \( \theta = 1 \). Figure 1(a) - (d) illustrate the dynamic updating process of the value function \( V(\theta) \) during this implementation.

In each plot, the blue circles form a scatter plot of \( \{ V^{(t)}(\theta^{*(t)}) \}_{t=1}^{it} \), where \( it \) is the iteration number at which the plot is made. They are the values of \( V \) generated in each iteration up to \( it \). The red line performs nonparametric smoothing on the \( N(it) \) most recent \( V^{(t)} \)'s using the same nonparametric estimator that is used to estimate the value of \( V^{(t)}(\theta^{*(t)}) \) in each iteration.

8 The value functions we use in our toy experiments are constructed such that they are functions of \( \theta \) and are solutions to a dynamic programming problem. They do not have economic interpretations and are not supposed to represent the value function of a dynamic, forward-looking agent.

9 i.e. \( s^d_t = 1 \) with probability 0.5, and = 0 with probability 0.5

10 Other details of the estimation program: we use a log-transformation of the random walk proposal in generating candidate parameter \( \theta^{*(t)} \) during each iteration:

\[
\theta^{*(t)} \sim \exp\left( \log\left( N(\theta^{(t-1)}, \sigma) \right) \right)
\]

The standard deviation of the random walk is set to \( \sigma = 0.1 \). The number of past iterations used for nonparametric estimation is set to \( N(t) = \min(t - 1, 100) \). We use an uninformative prior, so that the Metropolis-Hastings acceptance probability is

\[
\lambda(\theta^{(t-1)}, \theta^{*(t)}) = \min\left\{ 1, \frac{L(Z^d|\theta^{*(t)})}{L(Z^d|\theta^{(t-1)})} \right\}
\]

122
Therefore, we can view the red line as drawing the "current" estimated value function at iteration \( it \). Finally, the green dashed line plots the true value function \( V(\theta) = 10\theta \).

Let \( \hat{V}(\theta) \) be the function whose shape is delineated by the red line. The domain of \( \hat{V}(\theta) \) spans the set of candidate parameters generated in the \( N(it) \) most recent iterations, i.e. \( \{\theta^{*}(t)\}^{it}_{t=it-N(it)} \). We can see that by iteration 1500, \( \hat{V}(\theta) \) has largely converged to \( V(\theta) \) on a domain centered at the true parameter \( \theta_0 = 5 \), which means that a. the DP problem is largely solved for \( \theta \) around \( \theta_0 \) by iteration 1500; b. the estimation procedure has successfully narrowed the support for the posterior distribution of \( \theta \) to an area around \( \theta_0 \). Taking the initial 1500 values as the "burn-in" of the generated pseudo-markov chain, Figure 2 plots the time-series and the densities of the posterior mean of \( \theta \).
3.5.1.2 Comparing Nonparametric Estimators

The nonparametric estimator we have used to generate Figure 1 and 2 is the local linear estimator (with a normal Kernel). The rationale for the choice of the estimator will be made clear in the ensuing discussion. In this subsection, we consider several other nonparametric estimators including Loess, Spline, and the Nadaraya–Watson Kernel (henceforth N-W Kernel) estimator used in Imai, Jain, and Ching (2007), and discuss their respective properties as well as how they affect the performance of the BDP algorithm. Among the regression methods considered, N-W Kernel, local linear, and Loess regression can all be viewed as local polynomial methods, with N–W Kernel regression equivalent to a degree 0 local polynomial method, while local linear regression uses a degree 1 polynomial and Loess uses a quadratic polynomial. Comparing these estimators is akin to finding the optimal degree of local nonparametric regression for the performance of the BDP program.

The Nadaraya–Watson Kernel estimator We have used the N–W Kernel estimator in our outline of the BDP algorithm following Imai, Jain, and Ching (2007). To see how the estimator works, we plug it into our toy model estimation program and plot the dynamic updating process of $\hat{V}(\theta)$ à la Figure 1. Figure 3 displays the results.

We can see that by iteration 1500, the estimated value function $\hat{V}(\theta)$ has a domain centered at $\theta_0$, has a shape that crosses the true value function $V(\theta)$ at around $\theta_0$, but has a slope flatter than that of $V(\theta)$ and is far from convergence. In fact, if we increase the iteration number to 10,000, the plot still looks almost identical to Figure 3(d). Therefore, upon first sight, the Kernel estimated value function does not converge well.

The reason may be due to the following: 1. we have set $N(t) = \min(t - 1, 100)$, i.e. $N(t)$ goes up to only 100 in our estimation program; 2. we have been using a variable window width suggested by Bowman and Azzalini (1997), which, for a sample size of 100 in our toy data, generates a window width $h$ of around 6.5. The very small sample size and the wide window width pose obvious problems for N-W Kernel regression.
If we let $N(t)$ go up to 1000 and set the window width $h$ manually to 0.01, however, things do not improve. Figure 4(a) - (d) show the updating process of $V(\theta)$ with this configuration at iteration 500, 1000, 1500 and 10,000.

By iteration 10,000, $\hat{V}(\theta)$ is still far from converging. Even worse, $\theta_0$ is not yet in the domain of $\hat{V}(\theta)$. Looking more closely, we can see that the
concave, bell-like contour of the scatter plot of \( \{ V^{(t)} (\theta^*(t)) \}_{t=1}^{it} \) in Figure 4(d) matches the contour of the part of the scatter points that lie at the southwest corner of Figure 3. Obviously, with a larger \( N(t) \) and narrower window width, the estimated value function \( \hat{V}(\theta) \) converges much slower to the true function. One observation, therefore, is in line: because initial estimations of value functions can be very deviant from their true values, we need to keep \( N(t) \) small during initial iterations so that only the latest updated values are used to estimate \( V^{(t)} (\theta^*(t)) \). We should then let \( N(t) \) grow after the value functions have largely converged and the support of the posterior distribution of \( \theta \) approximately located, in order for the estimations to be fine tuned.

Therefore, although we expect the N-W Kernel estimator to eventually "get it," if we run the sampler long enough and allow \( N(t) \) to keep increasing (regardless of the computational cost), it is clear that "in the short run," the local linear estimator outperforms the N-W Kernel. The advantage the N-W Kernel estimator does have over the local linear, however, is its speed. Regression on a sample of a modest size of 1000, for example, is around 10 times faster using the N-W Kernel than the local linear estimator. To see which one is better "in the long run," we run our BDP program using the two, along with the Loess estimator we are about to discuss, until convergence, and compare their estimation results and convergence time. For our purpose, we set a loose convergence criterion, requiring only the standard deviation of the generated pseudo-markov chain to be driven down to 1. Again, we let \( N(t) \) go up to 1000\(^1\) and use the variable window width recommended by Bowman and Azzalini (1997). Table 3.1 display the results.

Thus, the N-W Kernel estimator does converge faster in the long run, due to its significant speed advantage when \( N(t) \) is large. We therefore recommend that the BDP program be run with the local linear estimator during initial iterations for faster convergence of value functions, and then switch to the N-W Kernel estimator for speed when \( N(t) \) grows large and value functions

\(^{11}\text{Specifically, we let } N(t) = \begin{cases} 
  t - 1, & t \leq 100 \\
  100, & t \in [101, 1000] \\
  500, & t \in [1001, 2000] \\
  1000, & t \geq 2001 
\end{cases}\)
Table 3.1: Convergence Time*

<table>
<thead>
<tr>
<th></th>
<th>Local Linear</th>
<th>N-W Kernel</th>
<th>Loess</th>
</tr>
</thead>
<tbody>
<tr>
<td>estimate $\left( \hat{E} [\theta] \right)^*$</td>
<td>4.73</td>
<td>4.91</td>
<td>5.22</td>
</tr>
<tr>
<td>iterations needed</td>
<td>5654</td>
<td>12599</td>
<td>3701</td>
</tr>
<tr>
<td>time</td>
<td>8 min 57 sec</td>
<td>41 sec</td>
<td>36 min 21 sec</td>
</tr>
</tbody>
</table>

*convergence criterion: s.t.d. of the generated pseudo-markov chain < 1; $\theta_0 = 5$

have largely converged.

**Loess** Loess is a smoothing technique that uses a tricube weight function and a second degree polynomial to perform locally weighted linear least square regression for the purpose of smoothing data\(^{12}\). We can perform nonparametric regression using a Loess estimator based on Loess smoothing.

As stated in the introduction of 5.1.2, comparing Loess, local linear and N–W Kernel regression is akin to finding the optimal degree of local nonparametric regression for the performance of BDP. It stands to reason that if local linear regression, which uses a first degree polynomial, is better in the short run than N–Watson Kernel regression, which is a 0 degree local polynomial method, then Loess regression, which uses a second degree polynomial, may perform even better. To see this, we plug the Loess estimator into our toy model estimation program and plot the dynamic updating process of $V(\theta)$. Figure 5 displays the results.

By iteration 1500, the Loess estimator does not produce better convergence than the local linear estimator. What is more, even on a sample size of 100, the Loess estimator is about 5 times slower than the local linear. Table 3.1 also displays the result of running the BDP program using the Loess estimator, which shows clearly that the estimator is not a good choice for our purpose.

\(^{12}\)Loess and lowess are often used without distinction. The distinction we make here is that lowess uses a a first degree polynomial for locally weighted least square regression, while loess uses a second degree polynomial.
Spline  For curiosity, we also test using Spline smoothing to perform non-parametric regression in the BDP program. We use the cubic Spline and Figure 6 displays the results. Obviously, Spline is not a good method for our purpose.
3.5.1.3 Conclusion

Nonparametric estimation is the bottleneck for the speed of executing the BDP program. However, due to the necessity of keeping \( N(t) \) small during initial updating of the value functions, it can be worthwhile to use a higher degree local polynomial method than N-W Kernel regression for better and faster convergence of value functions. In the toy model that we have experimented on, the optimal degree of local polynomial to use for local nonparametric regression during initial iterations is 1, pointing to local linear regression. The experimenter, of course, needs to adjust the regression method and its window-width according to the complexity of the model she is estimating.

It should be pointed out, however, that in later iterations, when value functions have largely converged, it should be possible to use even faster methods than N-W Kernel regression to estimate value functions at new parameter values. For example, we could simply use linear interpolation rather than any local polynomial based nonparametric regression techniques. For future work, we would like to test this possibility.

3.5.2 Experiment 2: The likelihood function having a narrow support on the range of \( V \), and other anomalies

Since in dynamic models, the likelihood is a function of \( V \), and in BDP estimation, the movement of the pseudo-markov chain hinges on the evaluation of the likelihood function, we consider what happens when the likelihood function has a narrow support on the range of \( V \), i.e. the likelihood function is only positive for a narrow range of \( V \) values.

Suppose the likelihood \( L \) is a function of \( V \) only. Let \( \Psi \) be the support of \( L \) and let \( \Theta \equiv V^{-1}(\Psi) \). Then \( \Theta \) is the set of \( \theta \) on which \( L \) is positive. If \( \Theta \) is a very small set, or if we start the estimation program from an initial guess far away from \( \Theta \), then it becomes more unlikely for any proposal \( \theta^{*}(t) \) in subsequent iterations to fall in \( \Theta \). Since \( L = 0 \) as long as \( \theta^{*}(t) \notin \Theta \), the
pseudo-markov chain would stay in its initial state and be unable to move on.

To visualize this, we run a toy experiment and let our set-up be the same as in experiment 1, except that \( \Pr(s^d_t|V,\theta) = \begin{cases} 0.5, & \text{if } |V(\theta) - 50| < 0.1 \\ 0, & \text{otherwise} \end{cases} \), where 50 is the value of \( V(\theta_0) \), so that the likelihood function \( L(Z^d|V,\theta) \equiv \Pi_t \Pr(s^d_t|V,\theta) \) is not 0 only for \( V \) belonging to a small interval around \( V(\theta_0) \).

Like in experiment 1, we let \( \theta_0 = 5 \), start the estimation with an initial guess of \( \theta = 1 \), and use a log-transformation of the random walk proposal with standard deviation equal to 0.1. Figure 7 shows the dynamic updating of \( V(\theta) \) at iteration 100, 1000, 200 and 10,000. Figure 8 shows the time-series plot of the pseudo-markov chain.

The chain can never move, even though \( \hat{V}(\theta) \) has completely converged to \( V(\theta) \) in a range of \( \theta \) around the initial value of \( \theta = 1 \). This is due to the likelihood function being 0 everywhere near \( \theta = 1 \). We can remedy the problem by increasing the standard deviation of our (log transformed) random walk proposal, or using another proposal function such as one that draws from a uniform distribution. Figure 9 and 10 show the results once the standard deviation of the proposal function is increased from 0.1 to 0.5.
Therefore, in models where the likelihood function has a narrow support on the range of $V$, and in turn, on the space of $\theta$, there is a risk of the pseudo-Markov chain never being able to move from the initial state. Increasing the standard deviation of the proposal function can help to find $\tilde{\Theta}$, the region of the parameter space on which the likelihood function is positive. Since the support of the posterior distribution of $\theta$ is always a subset of $\tilde{\Theta}$, the posterior distribution must be narrow as well. Therefore, we need to decrease the standard deviation of the proposal function once $\tilde{\Theta}$ has been located. In order to know when that is true, the experimenter needs to monitor the evaluation of likelihood function in each iteration. In more complicated models with multi-dimensional parameter space, however, this strategy of increasing then decreasing the proposal standard deviation may not work either.

Similar to this situation, the problem of a stalling pseudo-Markov chain can occur for the following situations: 1. $L$ being a constant function of $V$ on a continuous area; 2. $V$ being a constant function of $\theta$ on a continuous area; 3. $V$ having a small support on the parameter space $\Theta$, the reason being the same:
in any one case, it is difficult for the pseudo-markov chain to move away from
the parameter space that corresponds to the area on which the likelihood is a
constant (with 0 being a particular case), if that area is large. More generally,
MLE estimation as well as Bayesian MCMC estimation pose a smoothness
requirement on the likelihood as a function of model parameters, as well as a
uniqueness requirement on its local and global maximum. As an extension,
Bayesian estimation of dynamic models poses additional requirements on the
smoothness of $L$ as a function of $V$ and $V$ as a function of $\theta$, as well as requiring
them to both have a healthy support for the estimation program to perform
without malaise.

### 3.5.3 Experiment 3: $V$ as a function of $\theta$ and a
continuous state variable $k$

In this sub-section, we introduce a state variable $k$ with a continuous state
space and implement the BDP algorithm with random grid generation. Specifically, we construct a value function $V(k|\theta)$ that is the solution to the following
Bellman equation

$$V(k|\theta) = k + \beta \theta V(k|\theta)$$

(3.14)

, where $\beta = 0.9, k \in K \equiv (0,10), \theta \in \Theta \equiv \left(1, \frac{1}{\beta}\right)$.

The solution to the DP problem is $V(k|\theta) = \frac{k}{1-\beta \theta}$, but suppose we do not
know the solution and have to iterate on (3.14) until convergence to solve for
$V(k|\theta)$. The data set consists of $Z^d \equiv \{k^d_t, y^d_t\}_{t=1}^{T=1000}$, where $y^d_t$ is either 1 or 0. The likelihood function is $L = \prod_t \text{Pr}(y^d_t|k^d_t, V, \theta)$, where

$$\text{Pr}(y^d_t = 1|k^d_t, V, \theta) \equiv \min\left(V(k^d_t|\theta)/100, 1\right)$$

$$\text{Pr}(y^d_t = 0|k^d_t, V, \theta) = 1 - \text{Pr}(y^d_t = 1|k^d_t, V, \theta)$$

We generate the data set with $\theta_0 = 1$, such that $V(k|\theta_0) \in (0,100) \forall k \in K,$
and $\text{Pr}(y^d_t = 1|k^d_t, V, \theta) = V(k^d_t|\theta)/100 \in (0,1) \forall Z^d$. We run a program
that implements the BDP algorithm with random grid generation to estimate
$\theta_0$, starting with an initial guess of $\theta = 0.5$. More specifically, we draw uni-
formly a random state $k^{(t)} \in K$ in each iteration $t$, and estimate the value of
by nonparametric regression on \( \{ (k^{(r)}, \theta^{*^{(r)}}), V^{(r)} (k^{(r)} | \theta^{*^{(r)}}) \}^{t-1} \). Figure 11(a) - (d) demonstrate the dynamic updating process of \( V(k|\theta) \).

In each plot, the colored surface represents a plot of the \( V(t) \)s generated in the \( N(t) \) most recent iterations, where \( N(t) \) is set to \( \min(t-1, 100) \) in our implementation, while the blue dots represent a scatter plot of \( \{ V^{(t)} (k^{(t)} | \theta^{*^{(t)}}) \}^{it} \), where \( it \) is the current iteration number. We can see how state space points \( k^{(t)} \) are randomly and uniformly generated for different values of \( \theta^{*^{(t)}} \), and that by iteration 500, the range of \( \theta \) on which the surface plot sojourns has moved to an area around the true parameter \( \theta_0 = 1 \). Discarding the \( V(t) \)s in earlier iterations, Figure 11(e) takes a closer look at the surface representing the latest values of \( V(t) (k^{(t)} | \theta^{*^{(t)}}) \) by iteration 1500.

Finally, Figure 12 shows the time-series plot of the pseudo-markov chain generated by our estimation program.

### 3.6 Dynamic Models of Imperfect Competition

The BDP algorithm, as proposed in Imai, Jain, and Ching (2007), is for solving single-agent dynamic discrete choice problems. In this section, we extend the algorithm to estimating dynamic models of imperfect competition. Compared to single agent dynamic decision problems, dynamic models of imperfect competition involve multiple DP problems that are more complex than their counterpart in single agent problems. In a dynamic game of imperfect compe-

\[ V^{(t)} (k^{(t)} | \theta^{*^{(t)}}) \]

\[ \{ (k^{(r)}, \theta^{*^{(r)}}), V^{(r)} (k^{(r)} | \theta^{*^{(r)}}) \}^{t-1} \]

\[ \lambda \left( \theta^{(t-1)}, \theta^{*^{(t)}} \right) = \min \left\{ 1, \frac{L(Z^d|\theta^{*^{(t)}})}{L(Z^d|\theta^{(t-1)})} \right\} \]

---

\(13\) Other details of the estimation program include the use a transformation of the random walk proposal in generating candidate parameters: in each iteration \( t \), given \( \theta^{(t-1)} \), let \( \gamma = \log \left( \frac{\theta^{(t-1)} - a}{\theta^{(t-1)} - b} \right) \), where \( a = 0, b = 1/b \) are, respectively, the lower and upper bound of the parameter space \( \Theta \); Draw \( \gamma^* \sim N(\gamma, \sigma) \). Then let \( \theta^{*^{(t)}} = \frac{(a + b\gamma^*)}{(1 + e^{\gamma^*})} \). This guarantees that \( \theta^{*^{(t)}} \in [a, b] \).

The standard deviation \( \sigma \) used in the proposal function is set to \( 0.1 \). The nonparametric estimator used is the local linear estimator. \( N(t) = \min(t-1, 100) \). We use an uninformative prior, so that the Metropolis-Hastings acceptance probability is
Figure 11(a): Iteration = 100

- Surface plot of 100 most recent points
- \( \phi(\theta_i, \theta_{ij}) \)

Figure 11(b): Iteration = 500

- 3D scatter plot with increasing density of points
partition, each player’s strategy (choice of action) depends, in part, on her belief about other players’ private information and their strategies, thus weaving a net of interdependent strategies. The equilibrium concept usually applied is that of a Markov perfect equilibrium (MPE), which states that in equilibrium, 1. each player’s behavior depends only on the current state and her current private shock; 2. for each player, given other players’ equilibrium strategies, her optimal strategy is the equilibrium strategy. The second equilibrium condition is equivalent to requiring that every player in the game has the correct belief about other players’ strategies. Solution to a set of DP problems defined by the MPE of a dynamic model of imperfect competition, therefore, requires both the value functions and the strategy functions in the DP problems to converge, constituting what Aguirregabiria and Mira (2007) describe as ”coupled fixed-point problem[s].”

Specifically, consider the following set-up. Let there be $N$ players in a game. Let $\Psi$ be the state space and $s \in \Psi$ be the vector of commonly observed states. First suppose $\Psi$ is discrete. Let $A \equiv \{1, ..., J\}$ be a finite set of actions each player chooses from and $a \equiv (a_1, ..., a_N) \in A^N$ be the vector of actions chosen. The model is parametrized by $\theta \in \Theta$. Let $\varepsilon \equiv (\varepsilon_1, ..., \varepsilon_N) \sim \text{i.i.d.} F_{\varepsilon}$. $F_{\varepsilon}(\cdot | \theta)$ be the vector of private shocks, with each $\varepsilon_i \equiv (\varepsilon_{i1}, ..., \varepsilon_{iJ})$ being a vector of choice-dependent shocks, and with $\Xi \equiv \Xi_1 \times ... \times \Xi_N$ being the support of $F_{\varepsilon}$. Let $g(s'|s, a, \theta)$ be the state transition function, and $P_i(a_i|s, \theta)$ be the choice probability function of player $i$, i.e. the probability of player $i$ choosing action $a_i$ when the commonly observed state is $s$ and the model parameter is $\theta$. Define $P(a|s, \theta) \equiv (P_1(a_1|s, \theta), ..., P_N(a_N|s, \theta))$ and $P_{-i}(a_{-i}|s, \theta) \equiv (P_1(a_1|s, \theta), ..., P_{i-1}(a_{i-1}|s, \theta), P_{i+1}(a_{i+1}|s, \theta), ..., P_N(a_N|s, \theta))$. Let the markov strategy for player $i$ be a function $\sigma_i : \Psi \times \Xi \rightarrow A$, and let the strategy profile be $\sigma : \Psi \times \Xi \rightarrow A^N, \sigma \equiv (\sigma_1, ..., \sigma_N)$. The MPE of the game is then
characterized by the following set of DP problems: For \( i = 1, \ldots, N \),

\[
V_i(s, \varepsilon_i|\theta) = \max_{a \in A} \sum_{a_{-i}} \left\{ \pi_i(s, a, a_{-i}, \varepsilon_{ia}|\theta) + \beta \sum_{s'} E V_i(s'|\theta) \cdot g(s'|s, a, a_{-i}, \theta) \right\} \cdot \Pr(a_{-i}|s, \theta) \tag{3.15}
\]

\[
P_i(a_i|s, \theta) = \Pr\{a_i = \arg \max_{a \in A} \sum_{a_{-i}} \left\{ \pi_i(s, a, a_{-i}, \varepsilon_{ia}|\theta) + \beta \sum_{s'} E V_i(s'|\theta) \cdot g(s'|s, a, a_{-i}, \theta) \right\} \cdot \Pr(a_{-i}|s, \theta) \} \tag{3.16}
\]

where \( EV_i(s|\theta) \equiv \int V_i(s, \varepsilon_i|\theta)dF_{\varepsilon_i} \), \( \Pr(a_{-i}|s, \theta) \equiv \Pi_{j \neq i} P_j(a_j|s, \theta) \).

Together, (3.15) and (3.16) define a coupled fixed point problem that can be re-expressed as

\[
V(s, \varepsilon|\theta) = \mathbf{Y}(s, \varepsilon|E V(.|\theta), P(.|., \theta), \theta) \tag{3.17}
\]

\[
P(a|s, \theta) = \mathbf{A}(a|s, E V(.|\theta), P(.|., \theta), \theta) \tag{3.18}
\]

where \( V(s, \varepsilon|\theta) \equiv (V_1(s, \varepsilon|\theta), \ldots, V_N(s, \varepsilon|\theta)) \)

\( E V(s|\theta) \equiv (E V_1(s|\theta), \ldots, E V_N(s|\theta)) \)

\( \mathbf{Y}(s, \varepsilon|E V(.|\theta), P(.|., \theta), \theta) \equiv \{Y_1(s, \varepsilon) = E V_1(s, \varepsilon|\theta), P_1(.|., \theta, \theta), \ldots, Y_N(s, \varepsilon) = E V_N(s, \varepsilon|\theta), P_N(.|., \theta, \theta) \} \)

\( \mathbf{A}(a|s, E V(.|\theta), P(.|., \theta), \theta) \equiv \{A_1(a_1|s, E V_1(.|\theta), P_1(.|., \theta, \theta), \ldots, A_N(a_N|s, E V_N(.|\theta), P_N(.|., \theta, \theta) \} \)

and \( \{Y_i\}_{i=1}^N, \{A_i\}_{i=1}^N \) are sets of contraction mappings defined by:

\[
Y_i(s, \varepsilon|E V_i(.|\theta), P_{-i}(., ., \theta, \theta) \equiv \max_{a \in A} \sum_{a_{-i}} \left\{ \pi_i(s, a, a_{-i}, \varepsilon_{ia}|\theta) + \beta \sum_{s'} E V_i(s'|\theta) \cdot g(s'|s, a, a_{-i}, \theta) \right\} \cdot \Pr(a_{-i}|s, \theta)
\]

\[
\Lambda_i(a_i|s, E V_i(.|\theta), P_{-i}(., ., \theta, \theta) \equiv \Pr(a_i = \sigma_i(s, \varepsilon_i|E V_i(.|\theta), P_{-i}(., ., \theta, \theta))
\]

\[
\sigma_i(s, \varepsilon_i|E V_i(.|\theta), P_{-i}(., ., \theta, \theta) \equiv \arg \max_{a \in A} \sum_{a_{-i}} \left\{ \pi_i(s, a, a_{-i}, \varepsilon_{ia}|\theta) + \beta \sum_{s'} E V_i(s'|\theta) \cdot g(s'|s, a, a_{-i}, \theta) \right\} \cdot \Pr(a_{-i}|s, \theta)
\]

\[
3.6.1 \quad \text{Traditional Methods}
\]

Consider a time series data \( Z^d \equiv \{ s^d_t, a^d_t \}_{t=1}^T \), where \( s^d_t \) is the commonly observed state and \( a^d_t \equiv (a_{1t}^d, \ldots, a_{Nt}^d) \) specifies the choice of action made by each
player in period $t$. The likelihood is a function of $P$ and $EV$:

$$L(Z^d|EV(.|\theta), P(.|., \theta), \theta) = \Pi_{t=1}^{T} \Pi_{i=1}^{N} P_i(a_{i,t}^{d}|s_t, \theta) = \Pi_{t=1}^{T} \Pi_{i=1}^{N} A_i(a_i|s_t, EV_i(.|\theta).P_{-i}(.|., \theta, \theta)) = \Pi_{t=1}^{T} \Lambda(a_t^{d}|s_t, EV(.|\theta), P(.|., \theta, \theta)) \cdot 1$$

, where $1 \equiv (1, ..., 1)$ is an $N$–tuple.

The traditional methods of estimating the true parameter $\theta_0$ that generates $Z^d$ use MLE or conventional Bayesian MCMC estimators. In either case, $EV(s|\theta)$ and $P(a|s, \theta)$ have to be solved for every $s \in \Psi$, $a \in A^N$, and every $\theta$ at which the likelihood function is evaluated in the search for $\theta_0$. For each $\theta$, the calculation of $EV(s|\theta)$ and $P(a|s, \theta), s \in \Psi, a \in A^N$, requires solving the coupled fixed point problem of (3.17) and (3.18), for which we need to iterate on the Bellman equations (3.17) and (3.18) recursively until convergence. The procedure is detailed below.

1. Draw $M$ random shocks: $\varepsilon^{(m)} \sim F_\varepsilon(.|\theta), m = 1, ..., M$

2. In any iteration $t$, given $EV^{(t-1)}, P^{(t-1)}$, 

$$V^{(t)}(s, \varepsilon^{(m)}|\theta) = \Upsilon(s, \varepsilon^{(m)}|EV^{(t-1)}(.|\theta), P^{(t-1)}(.|., \theta, \theta)), m = 1, ..., M$$

$$EV^{(t)}(s|\theta) = \frac{1}{M} \sum_{m=1}^{M} V^{(t)}(s, \varepsilon^{(m)}|\theta)$$

$$P^{(t)}(a|s, \theta) = \Lambda(a|s, EV^{(t)}(.|\theta), P^{(t-1)}(.|., \theta, \theta))$$

3. Repeat until convergence: $\|EV^{(t)}(s|\theta) - EV^{(t)}(s|\theta)\| < \delta \forall s \in \Psi$

### 3.6.2 Bayesian Dynamic Programming

The BDP algorithm can be extended naturally to estimate dynamic models of imperfect competition while avoiding the computational cost of repeatedly solving the coupled fixed point problem of (3.17) and (3.18). The idea is the same as in the estimation of single agent problems. Instead of fully solving the DP problem for each $\theta$, we iterate on the Bellman equations once during each
iteration, and use local nonparametric regression to keep the value functions continuously updated using their values in past iterations.

Because of the existence of the choice probability function $P$, there are several different algorithms we can propose as extensions of the BDP. The algorithms differ from each other in how they handle $P$. We detail them below.

### 3.6.2.1 Algorithm 1

In each iteration $t$, given $\theta^{(t-1)}$ and history $H^{(t)} = \{\theta^{(r)}, \varepsilon^{(r)}, V^{(r)}, P^{(r)}\}_{r=1}^{t-1}$,

1. Generate $\theta^{(t)} \sim q(\theta^{(t-1)}, \cdot)$, where $q$ is any proposal density.

2. Draw a random shock $\varepsilon^{(t)} \sim F_\varepsilon (\cdot | \theta^{(t)})$.

3. $\forall s \in \Psi, a \in A^N$, let

$$EV^{(t)}(s | \theta^{(t)}) = \sum_{j=1}^{N(t)} V^{(t-j)}(s, \varepsilon^{(t-j)} | \theta^{(t-j)}) \cdot \frac{K_h (\theta^{(t)} - \theta^{(t-j)})}{\sum_{k=1}^{N(t)} K_h (\theta^{(t)} - \theta^{(t-k)})}$$

$$\hat{P}^{(t)}(a | s, \theta^{(t)}) = \sum_{j=1}^{N(t)} P^{(t-j)}(a | s, \theta^{(t-j)}) \cdot \frac{K_h (\theta^{(t)} - \theta^{(t-j)})}{\sum_{k=1}^{N(t)} K_h (\theta^{(t)} - \theta^{(t-k)})}$$

4. $\forall s \in \Psi, a \in A^N$, iterate on the Bellman equations to update $V^{(t)}, P^{(t)}$

$$V^{(t)}(s, \varepsilon^{(t)} | \theta^{(t)}) = Y(s, \varepsilon^{(t)} | EV^{(t)}(\cdot | \theta^{(t)}), \hat{P}^{(t)}(\cdot | \cdot, \theta^{(t)}), \theta^{(t)})$$

$$P^{(t)}(a | s, \theta^{(t)}) = \Lambda(a | s, EV^{(t)}(\cdot | \theta^{(t)}), \hat{P}^{(t)}(\cdot | \cdot, \theta^{(t)}), \theta^{(t)})$$

5. Given diffuse prior and symmetric proposal, calculate the Metropolis-Hastings acceptance ratio

$$\lambda(\theta^{(t-1)}, \theta^{(t)}) = \min \left\{ 1, \frac{L(Z^d | \theta^{(t)})}{L(Z^d | \theta^{(t-1)})} \right\}$$

, where

$$L(Z^d | \theta^{(t)}) = \Pi_{t=1}^{T} \hat{P}^{(t)}(a_t^d | s_t^d, \theta^{(t)}) \cdot 1 \quad (3.20)$$

$$L(Z^d | \theta^{(t-1)}) = \Pi_{t=1}^{T} \hat{P}^{(t)}(a_t^d | s_t^d, \theta^{(t-1)}) \cdot 1$$
and
\[
\hat{P}^{(t)}(a|s, \theta^{(t-1)}) = \sum_{j=1}^{N(t)} P^{(t-j)}(a|s, \theta^{(t-j)}) \frac{K_h \left( \theta^{(t-1)} - \theta^{(t-j)} \right)}{\sum_{k=1}^{N(t)} K_h \left( \theta^{(t-1)} - \theta^{(t-k)} \right)} \forall s \in \Psi, a \in A
\] (3.21)

6. Let
\[
\theta^{(t)} = \theta^{(t)} \text{ with probability } \lambda
\]
\[
\theta^{(t)} = \theta^{(t-1)} \text{ with probability } 1 - \lambda
\]

In addition to (3.20), the experimenter has the freedom to calculate the likelihood functions differently at additional computational costs. For example, we can let
\[
L \left( Z^d | \theta^{(t)} \right) = \prod_{t=1}^T P^{(t)}(a_t^d|s_t^d, \theta^{(t)}) \cdot 1
\]
\[
L \left( Z^d | \theta^{(t-1)} \right) = \prod_{t=1}^T \hat{P}^{(t)}(a_t^d|s_t^d, \theta^{(t-1)}) \cdot 1
\]
, where we use the choice probability function \( P \) that has been updated by an iteration on the Bellman equation (3.18). Doing so incurs additional computational costs in calculating \( P^{(t)}(.|., \theta^{(t-1)}) \). In addition to (3.21), we need to estimate \( EV^{(t)}(.|\theta^{(t-1)}) \) by
\[
EV^{(t)}(s|\theta^{(t-1)}) = \sum_{j=1}^{N(t)} V^{(t-j)}(s, \varepsilon^{(t-j)}) \left( \theta^{(t-j)} \right) \frac{K_h \left( \theta^{(t-1)} - \theta^{(t-j)} \right)}{\sum_{k=1}^{N(t)} K_h \left( \theta^{(t-1)} - \theta^{(t-k)} \right)} \forall s \in \Psi
\] (3.22)
, so that
\[
P^{(t)}(a|s, \theta^{(t-1)}) = \Lambda \left( a|s, EV^{(t)}(.|\theta^{(t-1)}), \hat{P}^{(t)}(.|., \theta^{(t-1)}, \theta^{(t-1)}) \right) \forall s \in \Psi, a \in A
\]

We can go even a step further and let
\[
L \left( Z^d | \theta^{(t)} \right) = \prod_{t=1}^T \Lambda \left( a_t^d|s_t^d, EV^{(t)}(.|\theta^{(t)}), P^{(t)}(.|., \theta^{(t)}, \theta^{(t)}) \right) \cdot 1
\]
\[
L \left( Z^d | \theta^{(t-1)} \right) = \prod_{t=1}^T \Lambda \left( a_t^d|s_t^d, EV^{(t)}(.|\theta^{(t-1)}), P^{(t)}(.|., \theta^{(t-1)}, \theta^{(t-1)}) \right) \cdot 1
\]
where we plug $\mathbf{EV}(t)$ and $\mathbf{P}(t)$ into the contraction mapping $\Lambda$, and effectively do another iteration on the Bellman equation (3.18) in calculating $L$. These choices of calculating the likelihood functions are warranted if the computational cost of (3.22) and the cost of iterating on (3.18) are relatively small.

In many models, it is possible to define a function $\Phi(s,a|\theta)$ and a function $f: \mathbb{R}^N_+ \times [0,1]^N \rightarrow \mathbb{R}^N_+$ such that

\[ \Phi(s,a|\theta) = f(\mathbf{EV}(s|\theta),\mathbf{P}(a|s,\theta)) \] (3.23)
\[ V(s,\varepsilon|\theta) = \Upsilon(s,\varepsilon|\Phi(\ldots|\theta),\theta) \] (3.24)
\[ \mathbf{P}(a|s,\theta) = \Lambda(a|s,\Phi(\ldots|\theta),\theta) \] (3.25)

For example, if each player's per period return does not depend on her or other players' current period actions, such as when the actions in the model are of the nature of entry, exit or investment decisions that do not impact current period returns, then the DP problem defined by (3.15) and (3.16) can be re-written as

\[ V_i(s,\varepsilon_i|\theta) = \max_{a \in A} \{ \pi_i(s,\varepsilon_{ia}|\theta) + \beta \Phi_i(s,a|\theta) \} \] (3.26)
\[ P_i(a_i|s,\theta) = \Pr \left\{ a_i = \arg \max_{a \in A} \{ \pi_i(s,\varepsilon_{ia}|\theta) + \beta \Phi_i(s,a|\theta) \} \right\} \] (3.27)

, where

\[ \Phi_i(s,a_i|\theta) \equiv \sum_{a_{-i}} \sum_{s' \in \Psi} \mathbf{EV}_i(s'|\theta) \cdot g(s'|s,a_i,a_{-i},\theta) \cdot \Pr(a_{-i}|s,\theta) \] (3.28)

(3.26), (3.27) and (3.28) can then be vectorized into the form of (3.23), (3.24) and (3.25).

To estimate this form of dynamic model of imperfect competition, we can, in addition to Algorithm 1, use a slightly different procedure. Instead of estimating $\hat{\mathbf{P}}(t)$ nonparametrically in each iteration, we can estimate $\Phi(t)$ non-parametrically and then use $\Phi(t)$ to generate $\mathbf{V}(t)$ and $\mathbf{P}(t)$ through iteration on the Bellman equations (3.24) and (3.25). Specifically, we can replace step 3 and 4 in Algorithm 1 with
3. \( \forall s \in \Psi, a \in A^N \), let

\[
\mathbf{EV}^{(t)}(s|\theta^{*t}) = \sum_{j=1}^{N(t)} V^{(t-j)}(s, \varepsilon^{(t-j)}|\theta^{*t-j}) \cdot \frac{K_h (\theta^{*(t)} - \theta^{*(t-j)})}{\sum_{k=1}^{N(t)} K_h (\theta^{*(t)} - \theta^{*(t-k)})}
\]

\[
\hat{\Phi}^{(t)}(s, a|\theta^{*t}) = \sum_{j=1}^{N(t)} \Phi^{(t-j)}(s, a|\theta^{*(t-j)}) \cdot \frac{K_h (\theta^{*(t)} - \theta^{*(t-j)})}{\sum_{k=1}^{N(t)} K_h (\theta^{*(t)} - \theta^{*(t-k)})}
\]

4. \( \forall s \in \Psi, a \in A^N \),

\[
V^{(t)}(s, \varepsilon^{(t)}|\theta^{*t}) = \Upsilon(s, \varepsilon^{(t)}|\hat{\Phi}^{(t)}(\ldots, \theta^{*t}) , \theta^{*t})
\]

\[
\mathbf{P}^{(t)}(a|s, \theta^{*t}) = \Lambda(a|s, \hat{\Phi}^{(t)}(\ldots, \theta^{*t}) , \theta^{*t})
\]

\[
\Phi^{(t)}(s, a|\theta^{*t}) = f(\mathbf{EV}^{(t)}(s|\theta^{*t}), \mathbf{P}^{(t)}(a|s, \theta^{*t}))
\]

In addition, we keep a history of \( H^{(t)} \equiv \{\theta^{*(r)}, \varepsilon^{(r)}, \mathbf{V}^{(r)}, \Phi^{(r)}\}_{r=1}^{t-1} \) instead of \( \{\theta^{*(r)}, \varepsilon^{(r)}, \mathbf{V}^{(r)}, \mathbf{P}^{(r)}\}_{r=1}^{t-1} \) for each iteration \( t \).

Conceivably, there can be advantages of using this variant of Algorithm 1 over its original version if the true values of \( P_t(.|., \theta_0) \) are close to their boundary values of 0 and 1, in which case local nonparametric regression would tend to be less accurate. In most cases though, the experimenter can choose either of the two variants depending on which of \( \Phi \) and \( P \) is a smoother function and more fit to be estimated by local nonparametric regression.

### 3.6.2.2 Algorithm 2

Instead of solving for the equilibrium choice probability \( P \) by iterating on the Bellman equation (3.18), we can approximate it by the empirical distribution of each player’s choices, i.e. we can solve the DP problem defined by (3.17) and (3.18) with the following routine: for each \( \theta \),

1. Draw \( M \) random shocks: \( \varepsilon^{(m)} \sim F_{\varepsilon}(., \theta), m = 1, \ldots, M \)

2. In any iteration \( t \), given \( \mathbf{EV}^{(t-1)}, \mathbf{P}^{(t-1)} \),
∀s ∈ Ψ, a ∈ A^N, let

\[ V^{(t)}(s, \varepsilon^{(m)}|\theta) = \mathbb{Y}\left(s, \varepsilon^{(m)}|EV^{(t-1)}(.|\theta), P^{(t-1)}(.|., \theta), \theta\right), m = 1, ..., M \]

\[ P^{(t)}(a|s, \theta) = \frac{1}{M} \sum_{m=1}^{M} I\left(a = \sigma\left(s, \varepsilon^{(m)}|EV^{(t-1)}(.|\theta), P^{(t-1)}(.|., \theta), \theta\right)\right) \]

\[ EV^{(t)}(s|\theta) = \frac{1}{M} \sum_{m=1}^{M} V^{(t)}(s, \varepsilon^{(m)}|\theta) \]

, where \( \sigma (s, \varepsilon|EV (.|\theta), P(.|., \theta), \theta) \equiv \begin{pmatrix} \sigma_1 (s, \varepsilon_1|EV_1 (.|\theta), P_{-1}(.|., \theta), \theta) \\ \vdots \\ \sigma_N (s, \varepsilon_N|EV_N (.|\theta), P_{-N}(.|., \theta), \theta) \end{pmatrix} \)

and \( \sigma_i (s, \varepsilon_i|EV_i (.|\theta), P_{-i}(.|., \theta), \theta) \) is defined as in (3.19).

3. Repeat until convergence: \( \|EV^{(t)}(s|\theta) - EV^{(t)}(s|\theta)\| < \delta \forall s \in \Psi \)

In other words, for each player \( i \), we let \( P_i(a_i|s, \theta) = 0.1 \), for example, if 10% of the generated shocks lead player \( i \) to choose action \( a_i \). This method of approximating \( P \) by the empirical distribution of \( \sigma \) is essentially a Monte Carlo technique. As discussed in section 3, the BDP algorithm, by drawing a single \( \varepsilon^{(t)} \) in each iteration, can save the computational cost of doing Monte Carlo integration. Below we detail how to apply the BDP algorithm.

In each iteration \( t \), given \( \theta^{(t-1)} \) and history \( H^{(t)} \equiv \{\theta^{(r)}, \varepsilon^{(r)}, V^{(r)}, \sigma^{(r)}\}_{r=1}^{t-1} \).

1. Generate \( \theta^{*r} \sim q \left(\theta^{(t-1)}, .\right) \), where \( q \) is any proposal density

2. Draw a random shock \( \varepsilon^{(t)} \sim F_{\varepsilon} (.|\theta^{*r}) \)

3. ∀s ∈ Ψ, a ∈ A^N, let

\[ EV^{(t)}(s|\theta^{*r}) = \sum_{j=1}^{N(t)} V^{(t-j)}(s, \varepsilon^{(t-j)}|\theta^{*r}) \cdot \frac{K_h \left(\theta^{*r} - \theta^{*r(t-j)}\right)}{\sum_{k=1}^{N(t)} K_h \left(\theta^{*r} - \theta^{*r(t-k)}\right)} \]

\[ P^{(t)}(a|s, \theta^{*r}) = \sum_{j=1}^{N(t)} I\left(a = \sigma^{(t-j)}(s, \varepsilon^{(t-j)}|\theta^{*r(t-j)})\right) \cdot \frac{K_h \left(\theta^{*r} - \theta^{*r(t-j)}\right)}{\sum_{k=1}^{N(t)} K_h \left(\theta^{*r} - \theta^{*r(t-k)}\right)} \]
4. \( \forall s \in \Psi, a \in A^N, \)

\[
V^{(t)}(s, \varepsilon^{(t)}|\theta^{*}) = \Upsilon \left( s, \varepsilon^{(t)}|EV^{(t)}(.|\theta^{*}), P^{(t)}(.|.), \theta^{*} \right)
\]

\[
\sigma^{(t)}(s, \varepsilon^{(t)}|\theta^{*}) = \begin{pmatrix}
\sigma_1 \left( s, \varepsilon_1|EV_1^{(t)}(.|\theta^{*}), P_1^{(t)}(.|.), \theta^{*} \right) \\
\ldots \\
\sigma_N \left( s, \varepsilon_N|EV_N^{(t)}(.|\theta^{*}), P_N^{(t)}(.|.), \theta^{*} \right)
\end{pmatrix},
\]

where \( \sigma_i(s, \varepsilon_i|EV_i(.|\theta), P_{-i}(., \theta), \theta) \) is defined as in (3.19).

5. Given diffuse prior and symmetric proposal, calculate the Metropolis-Hastings acceptance ratio

\[
\lambda(\theta^{(t-1)}, \theta^{*}) = \min \left\{ 1, \frac{\prod_{t=1}^{T} P^{(t)}(a^d|s^d, \theta^{*}) \cdot 1}{\prod_{t=1}^{T} P^{(t)}(a^d|s^d, \theta^{(t-1)}) \cdot 1} \right\}
\]

, where

\[
P^{(t)}(a|s, \theta^{(t-1)}) = \sum_{j=1}^{N(t)} I(a = \sigma^{(t-j)}(s, \varepsilon^{(t-j)}|\theta^{*})), \frac{K_h(\theta^{(t-1)} - \theta^{*(t-j)})}{\sum_{k=1}^{N(t)} K_h(\theta^{(t-1)} - \theta^{*(t-k)})} \forall s \in \Psi, a \in A
\]

6. Let

\[
\theta^{(t)} = \theta^{*} \text{ with probability } \lambda \\
\theta^{(t)} = \theta^{(t-1)} \text{ with probability } 1 - \lambda
\]

Like step (3.29), step (3.30) is a combination of local nonparametric estimation and Monte Carlo integration and is the key to this algorithm. Empirically, this algorithm works best when the action set \( A \) is small. When \( A \) is large, we need to increase \( N(t) \) or draw more \( \varepsilon^{(t)} \) in each iteration in order to increase the sample size on which the empirical distribution of \( \sigma \) is calculated. When \( A \) is small, however, algorithm 2 can potentially be better than algorithm 1, as it obviates the need to iterate on equation (3.18).

3.6.2.3 Algorithm 3

In this algorithm, we avoid the computation of the equilibrium choice probability function \( P \) altogether and instead, estimate \( P \) from data and plug the
estimated function directly into the Bellman equation for $V$ (3.17). The idea follows from Bajari, Benkard, and Levin (2007) (henceforth BBL), who propose a two-stage simulation-based estimator for dynamic models of imperfect competition. In stage 1, players’ policy functions (choice probability functions) and the state transition function are estimated from data; Value functions are then simulated using the estimated transition function and policy functions for every $\theta$ considered in the second stage estimation. In stage 2, a minimum distance estimator is used to find a set of structural parameters that rationalize the observed policies as a set of optimal decisions, thereby recovering the model parameters.

The underlying assumption of the BBL two-stage estimator is that observed player behaviors are consistent with Markov perfect equilibrium. Therefore, we can recover players’ equilibrium beliefs and strategies from their observed choices of action, with which we can then generate value functions by forward simulation at little computational cost. As BBL asserts, ”the biggest advantage of the approach is that it avoids the need for equilibrium computation.” On the flip side of the coin, however, this lack of equilibrium computation can also be a weakness of the estimator, for it is difficult to prove identification for its estimation results, as there is no guarantee that the value functions simulated by the BBL estimator are truly the fixed point solutions of the model’s DP problems.

In this algorithm, we follow the idea of BBL and estimate rather than compute the choice probability function from data and plug it into the BDP algorithm. However, we still compute $V$ by solving for the fixed point of the DP problem (3.17). Doing so present two advantages: 1. the reason BBL choose to simulate rather than compute $V$ is precisely due to the computational cost of solving DP problems. By using the BDP algorithm, however, this cost is lessened and we can afford to solve $V$ and enjoy the computational savings of not having to solve $P$ at the same time; 2. the BBL estimator, for all its computational savings, still have to simulate $V$ for every $\theta$ considered in its second stage estimator. By using the BDP algorithm, we eliminate any need of repeated simulations or Bellman equation iterations by the use of local
nonparametric regression. Therefore, this algorithm should, in principle, be more robust than the BBL estimator and faster than either Algorithm 1 or 2. Below we describe the detailed workings of the algorithm.

In stage 1, estimate the choice probability function from data to obtain \( \hat{P}(a|s) \). The estimation can either take the form of parametric regression of \( \{a_t^d\}_{t=1}^T \) on \( \{s_t^d\}_{t=1}^T \), or be done nonparametrically such as using the observed choice frequencies as \( \hat{P} \) at observed states and use local nonparametric regression to infer the values of \( \hat{P} \) at other points in the state space.

In stage 2, given \( \hat{P} \), iterate on the following routine:

In each iteration \( t \), given \( \theta^{(t-1)} \) and history \( H^{(t)} = \{\theta^{(r)}, \varepsilon^{(r)}, V^{(r)}\}_{r=1}^{t-1} \),

1. Generate \( \theta^{(t)} \sim q(\theta^{(t-1)}, .) \), where \( q \) is any proposal density
2. Draw a random shock \( \varepsilon^{(t)} \sim F_{\varepsilon}(\cdot|\theta^{(t)}) \)
3. \( \forall s \in \Psi, a \in A^N \), let
   \[
   EV^{(t)}(s|\theta^{(t)}) = \sum_{j=1}^{N(t)} V^{(t-j)}(s, \varepsilon^{(t-j)}|\theta^{(t-j)}) \cdot \frac{K_h(\theta^{(t)} - \theta^{(t-j)})}{\sum_{k=1}^{N(t)} K_h(\theta^{(t)} - \theta^{(t-k)})}
   \]
4. \( \forall s \in \Psi, a \in A^N \),
   \[
   V^{(t)}(s, \varepsilon^{(t)}|\theta^{(t)}) = \Upsilon \left( s, \varepsilon^{(t)}|EV^{(t)}(\cdot|\theta^{(t)}), \hat{P}, \theta^{(t)} \right)
   \]
5. Given diffuse prior and symmetric proposal, calculate the Metropolis-Hastings acceptance ratio
   \[
   \lambda(\theta^{(t-1)}, \theta^{(t)}) = \min \left\{ 1, \frac{\Pi_{t=1}^T \Lambda \left( a_t^d | s_t^d, EV^{(t)}(\cdot|\theta^{(t)}), \hat{P}, \theta^{(t)} \right) \cdot 1}{\Pi_{t=1}^T \Lambda \left( a_t^d | s_t^d, EV^{(t)}(\cdot|\theta^{(t-1)}), \hat{P}, \theta^{(t-1)} \right) \cdot 1} \right\}
   \]
   where
   \[
   EV^{(t)}(s|\theta^{(t-1)}) = \sum_{j=1}^{N(t)} V^{(t-j)}(s, \varepsilon^{(t-j)}|\theta^{(t-j)}) \cdot \frac{K_h(\theta^{(t-1)} - \theta^{(t-j)})}{\sum_{k=1}^{N(t)} K_h(\theta^{(t-1)} - \theta^{(t-k)})} \forall s \in \Psi
   \]
6. Let
   \[
   \begin{align*}
   \theta^{(t)} &= \theta^{(t)} \text{ with probability } \lambda \\
   \theta^{(t)} &= \theta^{(t-1)} \text{ with probability } 1 - \lambda
   \end{align*}
   \]
3.6.2.4 Conclusion

We have proposed 3 algorithms - call them Algorithm 1, 2, and 3 - that extend the BDP algorithm to the estimation of dynamic models of imperfect competition. The differences between these 3 algorithms lie in the way they treat the DP problem defined by (3.17) and (3.18), and in particular, they way they solve for the equilibrium choice probability function $P$. Algorithm 1 treat the DP problem as a coupled fixed point problem and duly solve it by iterating on both Bellman equations (3.17) and (3.18) in each iteration, while Algorithm 2 and 3 avoids iteration on (3.18) - thus avoids treating $P$ as the solution to a fixed point problem - and approximate $P$ either by Monte Carlo methods or estimation from data.

3.6.3 Bayesian Dynamic Programming with Random Grid Generation

Like in the estimation of single agent problems, we can use random grid generation to approximate the state space if the state space is continuous or much larger than the set of observed states. We describe below how to modify Algorithm 1, 2, and 3 to allow for random grid generation in the face of a continuous state space, depending on whether state transition is stochastic or deterministic.

3.6.3.1 Stochastic Transition

If state transition is stochastic, define a function $W(s, a|\theta) \equiv (W_1(s, a|\theta), ..., W_N(s, a|\theta))$, where

$$W_i(s, a|\theta) \equiv \int EV_i(s' | \theta) \cdot g(s'|s, a, \theta)ds'$$
, such that the DP problem defined by (3.15) and (3.16), in the case of a continuous state space, can be written as

\[ V_i(s, \varepsilon|\theta) = \max_{a_0} \sum_{a_{-i}} \{ \pi_i(s, a, a_{-i}, \varepsilon_{ia}|\theta) + \beta W_i(s, a, a_{-i}|\theta) \} \cdot \Pr(a_{-i}|s, \theta) \]

\[ P_i(a_i|s, \theta) = \Pr \left\{ a_i = \arg \max_{a} \sum_{a_i} \{ \pi_i(s, a, a_{-i}, \varepsilon_{ia}|\theta) + \beta W_i(s, a, a_{-i}|\theta) \} \cdot \Pr(a_{-i}|s, \theta) \right\} \]

, which, in vectorized form, become

\[ \mathbf{V}(s, \varepsilon|\theta) = \Upsilon(s, \varepsilon|\mathbf{W}(., .|\theta), \mathbf{P}(., ., \theta), \theta) \]

\[ \mathbf{P}(a|s, \theta) = \Lambda(a|s, \mathbf{W}(., .|\theta), \mathbf{P}(., ., \theta), \theta) \]

To implement the BDP algorithm with random grid generation, we replace step 3 and 4 of Algorithm 1 with:

3. Generate a random grid \( s^{(t)} \in \Psi \); \n
\[ \forall a \in A^N, \text{estimate } \mathbf{W} \text{ and } \mathbf{P} \text{ at the new grid point } s^{(t)} \text{ and the proposed parameter } \theta^{*}(t) \]

\[ \mathbf{W}^{(t)}(s^{(t)}, a|\theta^{*}(t)) = \sum_{j=1}^{N(t)} \mathbf{V}^{(t-j)}(s^{(t-j)}, \varepsilon^{(t-j)}|\theta^{*}(t-j)) \cdot \frac{K_{h_\alpha}(\theta^{*}(t) - \theta^{*}(t-j)) \cdot g(s^{(t-j)}|s^{(t)}, a, \theta^{*}(t))}{\sum_{k=1}^{N(t)} K_{h_\alpha}(\theta^{*}(t) - \theta^{*}(t-k)) \cdot g(s^{(t-k)}|s^{(t)}, a, \theta^{*}(t))} \]

\[ \mathbf{P}^{(t)}(a|s^{(t)}, \theta^{*}(t)) = \sum_{j=1}^{N(t)} \mathbf{P}^{(t-j)}(a|s^{(t-j)}, \theta^{*}(t-j)) \cdot \frac{K_{h_\alpha}(s - s^{(t-j)}) \cdot K_{h_\beta}(\theta^{*}(t) - \theta^{*}(t-j))}{\sum_{k=1}^{N(t)} K_{h_\alpha}(s - s^{(t-k)}) \cdot K_{h_\beta}(\theta^{*}(t) - \theta^{*}(t-k))} \]

4. Update \( \mathbf{V} \) and \( \mathbf{P} \) at \( s^{(t)} \) and \( \theta^{*}(t) \)

\[ \mathbf{V}^{(t)}(s^{(t)}, \varepsilon^{(t)}|\theta^{*}(t)) = \Upsilon(s^{(t)}, \varepsilon^{(t)}|\mathbf{W}^{(t)}(s^{(t)}, .|\theta^{*}(t)), \mathbf{P}^{(t)}(.|s^{(t)}, \theta^{*}(t)), \theta^{*}(t)) \]

\[ \mathbf{P}^{(t)}(a|s^{(t)}, \theta^{*}(t)) = \Lambda(a|s^{(t)}, \mathbf{W}^{(t)}(s^{(t)}, .|\theta^{*}(t)), \mathbf{P}^{(t)}(.|s^{(t)}, \theta^{*}(t)), \theta^{*}(t)), \forall a \in A^N \]
In addition, we need to estimate \( \hat{P}(t|a,s,\theta^*(t)) \) and \( \hat{P}(t|a,s,\theta^{(t-1)}) \) for all observed states in order to calculate the Metropolis-Hastings acceptance ratio \( \lambda(\theta^{(t-1)}, \theta^*(t)) \).

Algorithm 2 and 3 can be modified in a similar way. In particular, we apply the same changes of (3.31) and (3.33) to their step 3 and 4, while in addition, for Algorithm 2, replace (3.32) with

\[
\hat{P}(t|a,s(t-1),\theta^*(t)) = N(t) \sum_{j=1}^{N(t)} I(a = \sigma^{(t-j)}(s^{(t-j)}, \varepsilon^{(t-j)}|\theta^*(t-j))) \cdot K_h(s - s^{(t-j)}) \cdot K_h(\theta^*(t) - \theta^*(t-j))
\]

### 3.6.3.2 Deterministic Transition

When transition is deterministic, let \( r \) be the transition function such that \( s' = r(s,a,\theta) \). The DP problem defined by (3.15) and (3.16) becomes

\[
V_i(s,\varepsilon_i|\theta) = \max_{a \in A} \sum_{a_{-i}} \{ \pi_i(s,a,a_{-i},\varepsilon_{ia}|\theta) + \beta EV_i(r(s,a,a_{-i},\theta)|\theta) \} \cdot Pr(a_{-i}|s,\theta)
\]

\[
P_i(a_i|s,\theta) = \Pr \left\{ a_i = \arg \max_{a \in A} \sum_{a_{-i}} \{ \pi_i(s,a,a_{-i},\varepsilon_{ia}|\theta) + \beta EV_i(r(s,a,a_{-i},\theta)|\theta) \} \cdot Pr(a_{-i}|s,\theta) \right\}
\]

which, in vectorized form, can still be expressed as

\[
V(s,\varepsilon|\theta) = \Upsilon(s,\varepsilon|EV(.,|\theta),P(.,.|\theta),\theta)
\]

\[
P(a|s,\theta) = \Lambda(a|s,\varepsilon|EV(.,|\theta),P(.,.|\theta),\theta)
\]

To implement the BDP algorithm with random grid generation for (3.35) and (3.36), we replace step 3 and 4 of Algorithm 1 with:
dynamics: each period, each incumbent firm draws a scrap value and compares

\[ N \]

In each period the model is as follows. Let there be one market and \( N \) and exit model adopted from Dunne, Roberts and Xu (2006). The set-up of dynamic models of imperfect competition, we apply them to a simple entry

3.7 Simulation

To assess the performance of the algorithms we have proposed for estimating dynamic models of imperfect competition, we apply them to a simple entry and exit model adopted from Dunne, Roberts and Xu (2006). The set-up of the model is as follows. Let there be 1 market and \( N \) (homogeneous) firms. In each period \( t \), the market structure consists of \( n_{1t} \) incumbent firms and \( N - n_{1t} \) potential entrants, and evolves over periods according to the following dynamics: each period, each incumbent firm draws a scrap value and compares

3. Generate a random grid \( s^{(t)} \in \Psi \) and let

\[
EV^{(t)}(s^{(t)}|\theta^{(t)}) = \sum_{j=1}^{N(t)} V^{(t-j)}(s^{(t-j)}, \varepsilon^{(t-j)}|\theta^{(t-j)}) \]

\[
\hat{P}^{(t)}(a|s^{(t)}, \theta^{(t)}) = \sum_{j=1}^{N(t)} P^{(t-j)}(a|s^{(t-j)}, \theta^{(t-j)}) \]

4.

\[
V^{(t)}(s^{(t)}, \varepsilon^{(t)}|\theta^{(t)}) = Y\left(s^{(t)}, \varepsilon^{(t)}|EV^{(t)}(s^{(t)}|\theta^{(t)}), \hat{P}^{(t)}(.|s^{(t)}, \theta^{(t)}), \theta^{(t)}\right) \]

\[
P^{(t)}(a|s^{(t)}, \theta^{(t)}) = \Lambda\left(a|s^{(t)}, EV^{(t)}(s^{(t)}|\theta^{(t)}), \hat{P}^{(t)}(.|s^{(t)}, \theta^{(t)}), \theta^{(t)}\right) \]

, while \( \hat{P}^{(t)}(a|s, \theta^{(t)}) \) and \( \hat{P}^{(t)}(a|s, \theta^{(t-1)}) \) are also estimated for all observed states in order to calculate \( \lambda(\theta^{(t-1)}, \theta^{(t)}) \).

Algorithm 2 and 3 can be modified in a similar way. In particular, for Algorithm 2, we apply the changes of (3.37), (3.38) and (3.39), while for Algorithm 3, we apply (3.37) and (3.39) only.

### 3.7 Simulation

To assess the performance of the algorithms we have proposed for estimating dynamic models of imperfect competition, we apply them to a simple entry and exit model adopted from Dunne, Roberts and Xu (2006). The set-up of the model is as follows. Let there be 1 market and \( N \) (homogeneous) firms. In each period \( t \), the market structure consists of \( n_{1t} \) incumbent firms and \( N - n_{1t} \) potential entrants, and evolves over periods according to the following dynamics: each period, each incumbent firm draws a scrap value and compares
its expected discounted value of remaining in the market with the scrap value of liquidation to decide whether to stay in or exit the market. Meanwhile, each potential entrant draws an entry cost and weighs its expected discounted value of entering the market against the sunk cost of entry to decide whether or not to enter the market.

More specifically, let the state variable be the number of incumbent firms, \( n_1 \), and the state space be \( \Psi = \{0, 1, \ldots, N\} \). Let \( \pi_t(n_1) = R/n_1 \) be the in-period profit of an incumbent firm when there are \( n_1 \) firms in the market. Each period, the incumbent firm draws a scrap value \( \text{scrap} \sim \chi^2(\phi) \), where \( \chi^2 \) denotes a chi-square distribution, and the potential entrant draws an entry cost \( \text{entry} \sim \exp(\xi) \), where \( \exp \) denotes an exponential distribution. Let \( \phi \) and \( \xi \) be the model parameters that we want to estimate. Denote \( \theta \equiv (\phi, \xi) \).

Let \( p_1(n_1|\theta) \) be the probability of an incumbent firm to remain in the market when there are \( n_1 \) incumbent firms and \( p_2(n_1|\theta) \) be the probability of a potential entrant to enter the market when there are \( n_1 \) incumbent firms. Let \( V_1 \) and \( V_2 \) be, respectively, the value functions of an incumbent firm and a potential entrant\(^{14}\). Then \( V_1 \) and \( V_2 \) satisfy

\[
V_1(n_1, \text{scrap}|\theta) = \pi(n_1) + \max \{\text{scrap}, \beta \sum_{y_1=1}^{N-n_1} \sum_{y_2=0}^{n_1} \text{EV}_1(y_1 + y_2|\theta) \cdot f(y_1 - 1; n_1 - 1, p_1(n_1|\theta)) \cdot f(y_2; N - n_1, p_2(n_1|\theta)) \}
\]

\[
V_2(n_1, \text{entry}|\theta) = \max \{\text{entry}, \beta \sum_{y_1=0}^{n_1} \sum_{y_2=1}^{N-n_1} \text{EV}_1(y_1 + y_2|\theta) \cdot f(y_1; n_1, p_1(n_1|\theta)) \cdot f(y_2 - 1; N - n_1 - 1, p_2(n_1|\theta)) \}
\]

, where \( \text{EV}_1(n|\theta) \equiv E_{\text{scrap}}[V_1(n, \text{scrap}|\theta)] \), \( f \) is the Binomial probability function: \( f(y; n, p) \equiv \binom{n}{y} p^y (1 - p)^{n-y} \), and the choice probability functions \( p_1 \) and

\(^{14}\)Since all firms are homogeneous in this model, all incumbent firms have value function \( V_1 \) and all potential entrants have value function \( V_2 \).
p_2 \text{ satisfy} \\
p_1(n_1|\theta) = \Pr\{\text{scrap} < \beta \sum_{y_1=1}^{n_1} \sum_{y_2=0}^{N-n_1} EV_1(y_1 + y_2|\theta) \cdot f(y_1 - 1; n_1 - 1, p_1(n_1|\theta)) \cdot f(y_2; N - n_1, p_2(n_1|\theta))\}

p_2(n_1|\theta) = \Pr\{\text{entry} < \beta \sum_{y_1=0}^{N-n_1} \sum_{y_2=1}^{n_1} EV_1(y_1 + y_2|\theta) \cdot f(y_1; n_1, p_1(n_1|\theta)) \cdot f(y_2 - 1; N - n_1 - 1, p_2(n_1|\theta))\}

Together, (3.41), (3.43) and (3.44) form the coupled-fixed problem that characterize the equilibrium of this model. Notice that the potential entrant is not playing a dynamic game since it does not time its entry based on expectations of future entry costs, i.e. it does not solve an "optimal stopping problem." Therefore, \( V_2 \) does not enter into our DP problem.

For our simulation, we let \( N = 10, R = 0.1 \) and set the data generating parameter \( \theta_0 \equiv (\phi_0, \xi_0) = (1.5, 4.5) \). We compute the equilibrium expected value function \( EV_1(., \theta) \) and choice probability functions \( p_1(n_1|\theta), p_2(n_1|\theta) \) at \( \theta_0 \). We then generate a data set of 100 periods: \( Z^d \equiv \{(n_{1t}^d, y_{1t}^d, y_{2t}^d)\}_{t=1}^{T=100} \), where \( y_{1t}^d, y_{2t}^d \) are, respectively, the observed number of remaining incumbents and entries in period \( t \), i.e. \( n_{1,t+1}^d = y_{1t}^d + y_{2t}^d \). The log likelihood function for the observed data is

\[
\log L(\theta) = \sum_{t=1}^{100} \sum_{i=1}^{2} \left\{ y_{it}^d \log p_i(n_{1t}^d|\theta) + (n_{it}^d - y_{it}^d) \log (1 - p_i(n_{1t}^d|\theta)) \right\}
\]

, where \( n_{2t}^d \equiv N - n_{1t}^d \ \forall t \)

### 3.7.1 Applying the BDP Algorithms

We first apply Algorithm 1 and 2 to the estimation of the model. We start all estimation with the initial guess of \( \theta = (3, 3) \), and run each algorithm for 10 times, each time for a duration of 30 minutes (CPU time\textsuperscript{15}). In other words, we run 10 replications of each algorithm for a fixed amount of time.

\textsuperscript{15} All estimation exercises are done on a 1.60 GHz Pentium 4 Windows laptop machine
To compare the performance of the two algorithms, we record their estimation results (averaged over 10 replicas) at fixed time intervals, say, at 1, 5, 10 and 15 minutes. Table 3.2 presents the results.

In the table, MSE stands for Mean Squared Error: Given $n$ estimates $\hat{\theta}_1, \ldots, \hat{\theta}_n$ of $\theta_0$, $MSE(\hat{\theta}) \equiv \frac{1}{n} \sum_{i=1}^{n} (\hat{\theta}_i - \theta_0)^2$.

Recall that in Algorithm 1, during each iteration, the choice probability
function $P$ is first estimated using its values in past iterations and then updated by an iteration on the Bellman equation, while in Algorithm 2, $P$ is estimated by the empirical distribution of decisions made in past iterations without going through the Bellman. From Table 3.2, we can see that both algorithms work well and achieve MSE < 0.5 for all their estimates in 15 minutes. We can also see that the time needed to complete 1 iteration is about the same for both algorithms. But the table clearly indicate, however, that Algorithm 1 is superior to Algorithm 2 for the purpose of estimating the current model. At all points of time, Algorithm 1 is able to produce estimates with smaller MSE and standard deviations than Algorithm 2, and achieves MSE < 0.05 for all her estimates in as quickly as 10 minutes.

In Figure 13(a), we plot the time series of the poster mean of $\phi$ and $\xi$, and lay the graphs generated by the two algorithms upon each other. In Figure 13(b), we plot the posterior densities of the estimates. Again we can see, the estimates of Algorithm 1 are consistently closer to the true parameter values than those of Algorithm 2, and their posterior distributions have smaller support as well as being more centered on the true parameter values.

### 3.7.2 Applying the BDP Algorithms with Random Grid Generation

We next apply the modified Algorithm 1 and 2 with random grid generation to the estimation of the model. The state space of the model, however, is both discrete and small, and by random grid generation, we basically draw a state $n_1^{(t)}$ out of $\Psi = \{0, 1, ..., 10\}$ every iteration. Although the model itself does not warrant the use of random grid generation, the estimation exercise, however, can still test the performance of our modified algorithms and shed light on their differences.

Like in the previous estimation exercise, we start our estimation with the initial guess of $\theta = (3, 3)$ and run each algorithm with 10 replications. For this exercise, each replica is run for 3 hours and 30 minutes, due to the much longer time each iteration takes in algorithms with random grid generation,
and we only stop at the 3 hour 30 minutes mark due to time and computing resource constraints. Table 3.3 presents the estimation results at select time intervals, while Figure 14 plot the the time series of the poster mean of $\phi$ and $\xi$ as well as their posterior densities.

We can see that neither algorithms produce satisfactory estimates at the end of 3 hours and 30 minutes. Algorithm 2, in particular, registers huge MSE values for its early estimates of $\xi$, indicating the unreliability of the algorithm for small number of iterations. In Figure 14, the posterior means of $\phi$ and $\xi$ generated by Algorithm 1 are consistently closer to their true values than those produced by Algorithm 2 after 1500 iterations. Their posterior densities are also more centered around $\phi_0$ and $\xi_0$ than those produced by Algorithm 2, again indicating the superiority of Algorithm 1 over 2 (even when they are modified to incorporate random grid generation).

We also note, however, that by the end of 3 hours and 30 minutes, both algorithms have completed on average only a little more than 4000 iterations, and the MSE of their estimates are comparable to - looking back at Table 3.2 - those produced by the unmodified Algorithm 1 and 2 after 3000 iterations (at the 5 minutes mark). Therefore, the reason we do not get as satisfactory estimates from the modified algorithms may simply be that we haven’t run them long enough.

Recall our discussion on the merit of doing BDP with random grid generation in Section 4. Although the algorithm only needs to iterate on Bellman equations once on a single point each iteration, it computes a large number of local nonparametric regressions to estimate the value functions at other grid points observed in the data in order to calculate the likelihood function. These computations of local nonparametric regressions can be much more costly than single iterations on Bellman equations when $N(t)$, the sample on which the regressions are computed, grows large with iterations, thus giving the BDP algorithm with random grid generation a distinct computational disadvantage that is more pronounced the more number of Bellman equations there are in the model. Here we see that, for dynamic models of imperfect competition, which have more Bellman equations to iterate on than single agent problems,
Table 3.3: BDP estimation of dynamic entry and exit model with random grid generation

<table>
<thead>
<tr>
<th></th>
<th>Algorithm 1</th>
<th>Algorithm 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>MSE</td>
</tr>
<tr>
<td>CPU time: 1 min</td>
<td># iterations: 207</td>
<td># iterations: 203</td>
</tr>
<tr>
<td>$\phi$</td>
<td>2.23 (0.65)</td>
<td>1.13</td>
</tr>
<tr>
<td>$\xi$</td>
<td>5.77 (1.73)</td>
<td>4.71</td>
</tr>
<tr>
<td>CPU time: 30 min</td>
<td># iterations: 985</td>
<td># iterations: 990</td>
</tr>
<tr>
<td>$\phi$</td>
<td>1.77 (0.48)</td>
<td>0.33</td>
</tr>
<tr>
<td>$\xi$</td>
<td>5.38 (1.17)</td>
<td>2.60</td>
</tr>
<tr>
<td>CPU time: 1 h</td>
<td># iterations: 1416</td>
<td># iterations: 1438</td>
</tr>
<tr>
<td>$\phi$</td>
<td>1.71 (0.43)</td>
<td>0.27</td>
</tr>
<tr>
<td>$\xi$</td>
<td>5.29 (1.09)</td>
<td>2.25</td>
</tr>
<tr>
<td>CPU time: 2 h</td>
<td># iterations: 2362</td>
<td># iterations: 2339</td>
</tr>
<tr>
<td>$\phi$</td>
<td>1.64 (0.38)</td>
<td>0.20</td>
</tr>
<tr>
<td>$\xi$</td>
<td>5.14 (0.98)</td>
<td>1.77</td>
</tr>
<tr>
<td>CPU time: 3 h 30 min</td>
<td># iterations: 4322</td>
<td># iterations: 4296</td>
</tr>
<tr>
<td>$\phi$</td>
<td>1.56 (0.34)</td>
<td>0.12</td>
</tr>
<tr>
<td>$\xi$</td>
<td>4.93 (0.90)</td>
<td>1.10</td>
</tr>
</tbody>
</table>

*standard deviation in parenthesis; true value of $(\phi, \xi)$ is (1.5, 4.5)
even for a state space of 11 elements, algorithms with random grid generation take an inordinately more amount of time to execute than those without. Therefore, this estimation exercise reiterates the claim we make in Section 4 that random grid generation is only useful as means of approximating a continuous state space, or when the state space is significantly larger than the set of observed states in data.

3.8 Conclusion

This paper explores the Bayesian Dynamic Programming method proposed by Imai, Jain, and Ching (2007), which is used to estimate single agent dynamic discrete choice problems, and extends the method to estimating dynamic models of imperfect competition. Three different algorithms are proposed for this purpose, along with discussions of how to modify each to incorporate a random grid generation mechanism that is useful for estimating models with a continuous state space or a discrete state space that is much larger than the set of observed states in data. Monte Carlo simulations are performed to test the efficacy of these algorithms. Results indicate that the first proposed algorithm might be the best. For future work, we hope to apply the proposed algorithms to empirical data with dynamic models that have proved difficult for existing (Bayesian or non-Bayesian) methods to estimate.
Bibliography


Jiaming Mao

Address: Department of Economics
The Pennsylvania State University
University Park, PA 16802

Email: jmao@psu.edu
Web: http://sites.google.com/site/jiamingmao

EDUCATION:
- Ph.D. Economics, Penn State University, 2013
- B.A., Physics, Bard College, 2005

Ph.D. THESIS:
- Essays in Firm and Labor Market Dynamics
  Thesis Advisor: Prof. James Tybout

FIELDS:
- Primary: Applied Microeconomics, Labor Economics
- Secondary: Industrial Organization, Econometrics

PAPERS:
- “Going to Graduate School: A Life-cycle Model of Post-College Work and Educational Investment Decisions,” 2013
- “Bayesian Estimation of Dynamic Models of Imperfect Competition,” 2009

TEACHING EXPERIENCE:
- Instructor
  - International Economics (ECON 333, Summer 2008)
- Teaching Assistant
  - Econometrics (ECON 490, 1 semester)
  - Monetary Theory and Policy (ECON 451, 2 semesters)
  - Labor Economics (ECON 315, 2 semesters)
  - Macroeconomic Analysis (ECON 304, 2 semesters)
  - Microeconomic Analysis (ECON 302, 1 semester)

AWARDS:
- Distinguished Scientist Scholarship, Bard College (2001-2005)

OTHER RESEARCH EXPERIENCE:
- Research assistant for Prof. Henry J. Castejon, Department of Physics, Notre Dame University, 2002

ACTIVITIES:
- Vice President, Penn State Chinese Students and Scholars Association, 2007