January 2011

Three Essays on the Macroeconomics of Information

Yu Zheng
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Department of Economics

Dissertation Examination Committee:
Michele Boldrin, Chair
Philip H. Dybvig
David K. Levine
Juan Pantano
Adrian Peralta-Alva
Ping Wang

Three Essays on the Macroeconomics of Information

by

Yu Zheng

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2011

Saint Louis, Missouri

Abstract

This is a collection of essays on the macroeconomics of information. The first chapter, "The Skill Premium, College Enrollment and Education Signals" explores the quantitative implications of a signaling model of education for the evolution of the skill premium for young workers in the US since the 1970s. I formalize the idea that as college education becomes more affordable for a larger fraction of population, not having a college degree becomes a more precise signal about low ability, increasing the college wage premium. The model, when calibrated, suggests that about 17 percent of the growth in college premium is produced through this signaling channel. In light of the recent financial crisis, in the second essay, I study banks' incentive to produce public information about the return to the loans that they sell to risk-averse investors. Risk-averse banks rely on information production to redistribute risks between themselves and investors. I show that securitization, by eliminating the idiosyncratic component of risk, promises a less risky return, diminishing the marginal benefit of information, hence reducing information production. The third chapter (with Juan Pantano) contributes to the literature of accommodating unobserved heterogeneity in the Hotz-Miller estimation strategies by proposing a new two-step fixed-effect estimation approach. We uncover the type of each observation in the first step by exploiting a consistency requirement of the subjective assessments of a given type reported in subjective expectations data.

Acknowledgements

I owe my deepest gratitude to my advisor, Professor Michele Boldrin, for his continuous encouragement, guidance and support throughout the writing of the dissertation. I have also been a beneficiary of many helpful discussions with the other members of the dissertation examination committee: Professors Philip H. Dybvig, David K. Levine, Juan Pantano, Adrian Peralta-Alva and Ping Wang. Professors Gaetano Antinolfi, Rodolfo Manuelli, John Nachbar, Raul Santaeulilia-Llopis and Yongseok Shin are gracious enough to give me excellent comments on various versions of the papers.

A great deal of learning takes place among my peers and professional friends. I would like to thank Irene Botosaru, Rohan Dutta, Chao Gu, Stephanie Lau, Nicholas Papageorge, Xiaole Wu, Jie Zheng and many more, whose work I will follow and hopefully will cross mine in the future.

Last but not least, none of this would have been possible without the love and support of my parents. Finally, I'm grateful to Mr. Silvian Iticovici, whose aspiration towards perfection has always been a great inspiration.

Table of Contents

List of Figures

List of Tables

Chapter 1 The Skill Premium, College Enrollment and Education Signals

1.1 Introduction

The rise in the college wage premium - defined as the differential between the wages of college and high school graduates - is a well-documented fact. As Card and Lemieux (2001) have shown, the wage premium has evolved differently for different age groups: younger workers account for most of the growth of the premium. In line with the cohort-based perspective, this paper looks at college premium for workers age 23-26 and asks: how much of this evolution can be reasonably explained by the idea that higher education is (also) a signal of talent? The answer is motivated by the observation that the college premium and college enrollment rates have closely tracked each other during the past four decades (Figure 1). The story I submit is the following: as college education becomes more accessible, the lack of a college degree becomes an increasingly clear signal of poor talent; if talent, per se, is useful in the working place but unobservable, the college degree will be rewarded by an increasing premium relative to the high school diploma. The paper provides both a signaling model with closed-form solutions and a robust estimate of the signaling effect for the US economy from 1972 to 2005. Within a broadly defined class of models, the signaling mechanism accounts for about 17% of the growth in college premium.

[Figure 1.1 about here.]

afterwards rises half as quickly as that of college graduates, at 0.8% per year.

[Figure 1.2 about here.]

The model developed below formalizes the intuition for the case of a stationary economy. When I take the model to data, I interpret the increase in the wage to college graduates as due partly to capital accumulation and TFP growth, and partly, as an improvement in college's talent discrimination technology, summarized by the probability of graduating from college given talent. If people of all talents choose to go to college, the discrimination technology boils down to the average college completion rate. If the technology improves - i.e. it becomes easier to complete college if talented and harder if not - the average completion rate increases. Thus an improvement in the discrimination technology leads to a rising wage to college graduates and an increasing college completion rate, as in the data (see Figure 6).

Because the empirical relevance of my theory requires to be plausible the assumption that college has become progressively more affordable because financial constraints were relaxed, I should discuss here the relevant evidence. Baumol and Blackman (1995) and Archibald and Feldman (2008) are two of the very few papers that address the change of college affordability over time directly. Both papers recognize the rise of college price as a cost disease phenomenon. Although the share of income spent on college education has gone up, the relative price of other goods, which have experienced rapid productivity growth, fell so much that given income, one could actually afford more college education and more other goods. Archibald and Feldman (2008) argue that the difference between income and college expense is a better measure of affordability than share of income spent on education. They show that during 1990-2007 the median income left over after paying for college expense increased for both public and private institutions, with larger gains in public institutions. Here I re-construct this measure to cover the period from 1975/76 to 2008/09. Figure 3 plots the time series of the difference between the HP-filtered median household income and the net college price, together with the net college price as a share of median household income. The net college price is obtained by substracting average total aids per full-time-equivalent (FTE) student from average tuition, fees, room and board (TFRB). The total aids include grant aids, federal loans, educaiton tax benefits and federal work-study. The result is broadly consistent with the aforementioned findings. The residual income shows an upward sloping trend, indicating an increase in college affordability, even though the share of income keeps rising too.

[Figure 1.3 about here.]

Micro data tells a similar story. The National Postsecondary Student Aid Study (NPSAS) contains student-level information on financial aid provided by the federal government, the states, postsecondary institutions, employers, and private agencies, along with demographic and enrollment data My sample consists of all students who are dependent and enrolled in a bachelor's degree program in NPSAS 87, 90, 93, 96, 00, 04 and 08. Tables 1 shows the difference between the mean of parents' income and tuition and fees net grants and federal loans, by household income quintile and type of institution. The growth in the residual income is more apparent for 4-year public institutions than for 4-year non-for-profit private institutions. Notably, the increasing trend holds across all income groups for public colleges. In so far as the marginally constrained student is more likely to attend a public school, the evidence is supportive. The case of selective private colleges is examined by Hill, Winston and Boyd (2005). In their sample of 28 highly selective COFHE[1] colleges and universities, the real net price of attending those institutions as share of income fell for all income groups and the most dramatic decline was at the lowest quintile income group. The above analysis assumes a constant family size. If one takes into account that the number of own children under 18 per family has decreased from 1.28 in 1971 to 0.84 in 2009[2], the residual income after paying for children's college expenses should increase even more.

[Tables 1.1 about here.]

Some more evidence is available from the literature on the effect of aid on college enrollment. Federal grants and loans have increased dramatically on a per FTE student basis (Figure 4). The two key questions are how sensitive enrollment is to college price and how effectively the grants and loans programs are in promoting college access. The empirical evidence is mixed (for a review, cf. Kane, 2006). Most of the estimations that exploit cross-sectional variability reach an estimate that $1,000 reduction of college tuition increased the enrollment rate by three to five percentage points. See for example Kane (1994), Dynarski (2003), and Winter (2009). However, those studies that look at the

1 Consortium on the Financing of Higher Education. All private institutions.
2 U.S. Census Bureau, Families and Living Arrangements 2009, Table FM-3.

enrollment of high- and low-income students before and after Pell Grant was launched in 1973 do not find relative increase in attendance in the low-income group, but those models are typically not well identified (Kane, 1995 and Leslie and Brinkman, 1983). Long (2007) finds a positive effect of loans on enrollment for those families who had just become eligible and the effect was concentrated in full-time enrollment. By and large, the evidence seems to favor a positive effect from grants and loans programs.

[Figure 1.4 about here.]

In the line of research that focuses on differential enrollment behaviors across racial/ethnic groups, Cameron and Heckman (2001), Carneiro and Heckman (2002 and 2003) argue that long-run factors that determine the preparedness for college are more important than short-term cash constraints in making schooling decision. Their point can be translated into a high correlation between family income and ability in my model. As long as the correlation is not 1, in which case ability is observable and there is no role for college as a signal, the signaling mechanism in this paper still works, though the college premium would be smaller. In fact, recognizing the positive correlation between family income and talent helps my argument in the sense that the true marginal student, who can benefit from college and is barely financially constrained, is likely from the middle income group instead of the lowest one. I have shown in the preceding paragraphs that the college indeed has become more affordable to the middle income families for both types of institutions. While my model does not aim to provide a theory of enrollment decision per se, the only, realistic, assumption that I need is that college enrollment rates have risen over the years because a bigger and bigger fraction of the population can go to college when they choose to.

1.2 A Brief Literature Review

I briefly review the related literature. The evolution of the aggregate skill premium is described, among others, by Autor, Katz and Kearney (2008). Katz and Murphy (1992) provide a supply and demand framework to account for the dynamics of wages. Autor, Katz and Krueger (1998) rely on skill-biased technological change to rationalize the demand for skilled labor outpacing the supply. While their model involves assumptions on the unobservable quality of labor, Krusell, Ohanian, Rios-Rull and Violante (2000) show that the capital-skill complementarity can account for almost all of the growth in aggregate skill premium without any change in the trend of the unobservable.

While all of the papers above look at wage differentials by education attainment across all age groups, Card and DiNardo (2002) point out that the skill premium does not grow at the same rate across age groups. Further, Card and Lemieux (2001) estimate a production model with imperfect substitution between workers from different age groups and attribute the rising college premium for younger workers to the slowdown in the rate of growth of educational attainment starting with the 1950 cohorts. My paper shares with their work this cohort-based perspective. Guvenen and Kuruscu (2009) calibrate a overlapping generations model of human capital accumulation with skill-biased technical change and heterogeneous agents differing in the ability to accumulate human capital. Their model generate behaviors of the overall wage inequality and college premium for young workers that are consistent with the data. This paper differs from all of the above papers in that I abstract away the technological progress in the production process that change the labor demand. Instead, I focus on the implication of the signaling effect of education in an environment in which the suppliers of labor are less and less financially constrained in their schooling decision.

While the application of signaling theory to the college wage premium is relatively new, the idea of education-as-a-signal is obviously not: it dates back to Spence (1973). Hendel, Shapiro and Willen (2005) argue that decreasing interest rates on borrowing or decreasing tuition has the unintended consequence of widening the wage gap for similar reasons to the ones in this paper. They develop a model with imperfect capital markets and look at a separating equilibrium with two types, in which only the high ability type can benefit from college. The presence of the wedge between the borrowing and lending rates of interest enriches the dynamics of the skill premium and college attendance and allow them to discuss policies such as college loans. In contrast, this paper looks at a pooling equilibrium where all agents having continuously distributed abilities can benefit from college as long as they can afford it, while shutting down credit markets completely. The change in affordability, which depends on the availability of financial aids and loans, is governed by the speed with which the budget constraints are relaxed, a parameter which is calibrated to match the observed enrollment rates. The convenience of a pooling equilibrium is technical. The equilibrium dynamics has a closed form which facilitates the calibration. However, it is plausible to me that a high school graduate believes that he can benefit from college given the option of dropping out. Bedard (2001) lends some support to this by showing that high school dropout rates are higher in areas with greater university access. When more high school graduates have access to college, being a high school graduate without college enrollment is not worthy of the effort to complete the high school. While both my model and Hendel et al. (2005) predict no variation over time in the wage offer to college graduates, Balart (2010) specifies conditions on the wealth distribution under which more access to higher education decreases earnings for all education groups within the framework of Hendel et. al.

This paper also contributes to the literature which quantifies the relative importance of college education as a process of human capital enhancement and as a signaling device. Riley (2001) summarizes a large body of empirical research that tests the educational screening hypothesis against the human capital accumulation hypothesis, reaching mixed conclusions. I refer the reader to the references therein. Recognizing both roles of college education in generating college premium, Fang (2006) estimates a structural static model of endogenous education choices and wage determination and finds that productivity enhancement accounts for at least two-thirds of the college wage premium. On the other hand, Taber (2001) develops a dynamic programming selection model and finds evidence that the change in college premium in the 1980s was more plausibly driven by increasing demand for unobservable abilities than for skills acquired at school. While Taber (2001) suggests that the educational signal was likely to play a big role, he does not model the education signaling explicitly. He assumes the within-cohort ability differential between college graduates and high school graduates to be constant over time, eliminating the cohort effect on the evolution of college premium. It is precisely this cohort effect that is the focus of this paper. More specifically, to borrow from Taber's terminology, the change of college premium has three potential sources: the change in the payoff to skills acquired in college, the change in the payoff to unobservable ability, the change in the ability differential conditioning on education outcome. Fang (2006) suggests that the first source is important, because in a static setting the college premium is determined mostly by the payoff to skills learned in school. Taber's (2001) argument is that the second source seems to play a larger role than the first, ignoring the third possibility. In contrast, my paper argues, roughly, that regardless of the relative importance of the first two roles, the third source, the cohort effect, accounts for around 17% of the growth in college premium.

The rest of the paper is organized as follows. Section 2 presents the theory, while Section 3 simulates the model and provides a measurement of the effect of signals on the growth of skill premium. Section 4 concludes. All proofs are in the Appendix.

1.3 Model

1.3.1 A Static Model: the Working of the Education Signal

A static model may help the reader's intuition. Assume personal talent is private information that is nevertheless useful in production. Firms can base their wage offer only on the observable signal, which consists of having attained, or not, a college degree. Everyone is born with a high school diploma.

The population has size one, half is endowed with high talent, $\overline{\theta}$, and half with low talent $\underline{\theta}$. Let the distribution of wealth in the population be $F(\Omega)$. College education has a fixed cost of Q. Assume that all those with wealth $\Omega > Q$ go to college, hence, the fraction of people who goes to college is $F(Q)$. Assume there is randomness in successfully completing college. The probability of a high (low) talent person to complete college is $\overline{p}$ ($\underline{p}$), with $\overline{p} > \underline{p}$. The wage offer is simply the expected talent conditional on the signal received.

With some algebra, we have the wage offer to college graduates $\overline{W}$ and to high school graduates $\underline{W}$,

$$
\begin{aligned}
\overline{W} &= \frac{\overline{p}}{\overline{p}+\underline{p}}\overline{\theta} + \frac{\underline{p}}{\overline{p}+\underline{p}}\underline{\theta}, \\
\underline{W} &= \frac{1-\overline{p}[1-F(Q)]}{2-(\overline{p}+\underline{p})[1-F(Q)]}\overline{\theta} + \frac{1-\underline{p}[1-F(Q)]}{2-(\overline{p}+\underline{p})[1-F(Q)]}\underline{\theta}.
\end{aligned}
$$

While $\overline{W}$ is a constant, $\underline{W}$ depends on the fraction of people that can afford to go to college. Write $x = 1 - F(Q)$, we have $\underline{W}'(x) < 0$, implying that the wage differential increases together with college attendance. Next we embed this simple mechanism in a dynamic model of production.

1.3.2 Embedding the Signals in a Dynastic Model

This is a continuous time discrete-choice problem. Each agent is indexed by the pair (θ, k_0), where θ denotes talent, distributed in $[0, \overline{\theta}]$ according to a cumulative distribution function $G(\theta)$, and k_0 is the initial endowment of capital from a distribution $F(k_0)$ over $[0, \overline{k}_0]$. The distributions $G(\cdot)$ and $F(\cdot)$ are independent. Each agent is endowed with 1 unit of labor. In each instant, an agent faces a discrete choice of whether going to college or not. There are two implicit assumptions in this formulation. One, the offspring of the high (low) type remains high (low); since our main concern is not about social mobility, this assumption seems innocuous. Two, firms cannot, through repeated interaction with an agent from the same dynasty, infer her type. Agents save a constant fraction of their income in each instant. Saving must be positive, i.e. agents cannot borrow against future income. We will relax this assumption later. College education requires a fixed cost $Q > 0$. The rest is the same as in the static model, with $p(\theta)$, a monotone increasing function, representing the probability of completing college for type θ.

1.3.2.1 The Agents' Problem

At each instant of time, an agent (θ, k_0) decides whether to go to college or directly to the labor market. If he decides to go to college, he pays the fixed cost Q, after which one of the two possible states of nature is realized: he either completes college or not. After finding a job, he works, consumes and saves a fraction σ of his income. Agents are risk neutral and maximize the discounted sum of future consumption taking the rental rate of capital $R(t)$ and the wages $\overline{W}(t)$, $\underline{W}(t)$ as given: $U(c(t)) = \int_0^\infty c(t)e^{-rt}dt$.

Since there is no disutility from labor, all agents supply 1 unit of labor inelastically. There is no capital depreciation. For ease of exposition, the time argument is suppressed when it does not cause confusion.

Lemma 1 *If it is optimal for an agent with talent θ to go to college at t, then it is optimal for any agent who has talent greater than θ to go to college at t as long as his current capital holding $k \geq Q$.*

Intuitively, for an agent with talent θ attending college is convenient if $p(\theta)(\overline{W} - \underline{W}) - RQ$ is positive. Because $p(\theta)$ is increasing, this implies the result.

1.3.2.2 Production

In each period the representative firm rents capital from the households and hires workers. I will look at two different classes of production functions. The first class, call it $P1$, is

$$Y(K, L_H, L_C) = [\lambda L_H^\rho E(\theta|HSG) + \upsilon K^\rho + (1-\lambda-\upsilon) L_C^\rho E(\theta|CG)]^{1/\rho},\ \rho \leq 1, \quad \text{(P1)}$$

where L_H is the number of high school graduates and L_C is the number of college graduates. Here high school graduates and college graduates are perceived as different inputs, i.e. they are assigned different jobs. The productivity of each group is its average talent, by Law of Large Numbers. Implicitly, college education here is productive in the sense that successfully completing college equips the college graduates with a particular set of skills that allow them to undertake a particular task. The elasticity of substitution between two types of labor is the same as their elasticity with capital. In contrast, the second class of production functions only employs aggregate labor and capital as its inputs, that is, skilled and unskilled labor are perfect substitutes:

$$Y(K, L) = A[\alpha K^\rho + \beta(L \cdot E(\theta))^\rho]^{1/\rho}. \quad \text{(P2)}$$

In both cases, markets are competitive and the high school (or college) graduates will be paid by their marginal product conditional on the signal. Later, in the calibration section, I will explore the different quantitative implications of the two production functions. The total stock of capital is $K(t) = \int_0^{\overline{k_0}} k(t) dF(k_0)$ and the total labor supply $L(t) = 1, \forall t$. Following the tradition, skilled labor (or, unskilled) and college graduates (or, high school graduates) are used interchangeably.

1.3.2.3 Equilibrium

Definition 1 Equilibrium without credit markets

An equilibrium without credit markets of this economy is a list $(c(t), k(t), sh(t))$ for each agent (θ, k_0) and a list of prices $(R(t), \overline{W}(t), \underline{W}(t))$ given initial capital distribution $F(\cdot)$ and distribution of talent $G(\cdot)$, the exogenous positive saving rate σ and the production technology, so that
(i) Agents optimally make schooling decision $sh(K(t))$, given $R(t)$, $\overline{W}(t)$, $\underline{W}(t)$;
(ii) Firm maximizes period profit;
(iii) Factor Markets clear.

To provide an analytically convenient environment, we will look at a special class of the equilibrium defined above, the pooling equilibria in which all agents optimally go to college as soon as they can afford it. More discussion on equilibrium selection can be found at the end of this section. Before proving the existence of the pooling equilibria, I will prove the monotonicity of the wage differential in enrollment under the proposed strategy profile, which will be useful in the construction of the equilibrium later. Let x be the fraction of agents who go to school and we have $x = 1 - F(Q)$. The theoretical results here are presented mainly for $P1$. An analogous characterization of equilibria with $P2$ can be obtained from the author upon request.

Lemma 2 *For $P1$, under the strategy profile that all types of agents go to college as soon as their current capital holdings $k \geq Q$, for high ρ and low Q, $\ln(\overline{W}/\underline{W})$ is increasing in the fraction, x, of agents going to college.*

To facilitate interpretation, the wage differential has the form of $\frac{\overline{W}}{\underline{W}} = \frac{1-\lambda-\upsilon}{\lambda}(\frac{L_C}{L_H})^{\rho-1}\frac{E[\theta|CG]}{E[\theta|HSG]}$. An increase in the attendance will unambiguously lead to a higher ratio of expected talents, $\frac{E[\theta|CG]}{E[\theta|HSG]}$, by exactly the same logic as in the static model. Imagine $\rho = 1$, then the wage differential will unambiguously go up. However, for $\rho < 1$, the general equilibrium effect kicks in. Since college graduates become more abundant, its marginal productivity decreases relative to that of high school graduates, and this mitigates the effects of the signals. For every Q, I can find a $\widehat{\rho} \leq 1$, such that for all $\rho \geq \widehat{\rho}$, this monotonicity property of the wage gap holds. In general, the monotonicity of wage differential rely also on small Q and high ρ. Consider a separating equilibrium, in which higher types opt for school and lower types don't. Suppose that college is very expensive, hence few people can afford it. Then a college degree is more correlated with wealth than with talent and the signal it contains is weak. The marginal productivity of skilled labor is high, hence skilled labor would be receiving a high payment, if identifiable. But holding a college degree is not such a clear signal of talent, as only the rich can afford it. If college enrollment increases while its cost is constant the signal's quality does not improve as the high cost of attending college implies we are scrapping the "bottom of the barrel" among wealthy people. More generally, this is true also when the cost of attending college decreases as long as it is high and the distribution of wealth is not concentrated at high values of wealth. The marginal productivity of skilled labor decreases, though, relative to that of unskilled labor and, as a result, we may have a range in which increasing college attendance brings about a decrease of the wage premium.

Proposition 1 *Under some assumptions, for Q sufficiently small, there exists a pooling equilibrium where all types of agents choose to go to college as soon as $k \geq Q$.*

To guarantee that the net benefit of college attendance, $p(\theta)(\overline{W}(t) - \underline{W}(t)) - R(t)Q$ is positive for all t, Q cannot be too high. A sufficient upper bound, $\widehat{Q}$, is the solution (which exists) to $p(0)(\overline{W}(0) - \underline{W}(0)) = \upsilon(K(0) - \widehat{Q})^{\rho-1}\widehat{Q}$.

Remark 1 The bound of admissible Q, $\widehat{Q}$, is (i) increasing in x_0; (ii) increasing in $p(0)$; necessarily $p(0) > 0$; (iii) increasing in $K(0)$.

The above proposition has nice implications about the trends of college enrollment rate and of skill premium.

Corollary 1 *There is a cut-off level of the initial wealth for a given $t, k_0(t)$, so that for all agents whose endowment $k_0 \geq k_0(t)$, they will choose college education at t. That is, the college enrollment rate is increasing over time.*

Observe that all agents who haven't attended college accumulate capital in exactly the same fashion: $\dot{k^i} = \sigma[R(t)k^i + \underline{W}(t)]$. Therefore, $k_0(t)$ satisfies $k_0(t) = Q - \int_0^t \dot{k}(s)ds$,where the evolution of $k(s)$ follows $\dot{k} = \sigma[R(s)k(s) + \underline{W}(s)]$, $0 \leq s \leq t$. Obviously, $k_0(t)$ is decreasing over time along the equilibrium path.

Corollary 2 *The wage gap is widening over time along the equilibrium path.*

The equilibrium path is completely characterized in terms of the aggregate capital, $K(t)$, and the cut-off wealth level, $k_0(t)$:

$$\begin{cases} \dot{K}(t) = \sigma Y(K(t) - x(t)Q, 1 - x(t)\int pdG, x(t)\int pdG) \\ \dot{k_0}(t) = -\sigma[R(t)Q + \underline{W}(t)] \end{cases} \tag{1}$$

$$s.t.\ k_0(t) \geq 0$$

$$\text{with } K(0) = \int_0^{\overline{k_0}} k_0 dF(k_0) \text{ and } k_0(0) = Q;$$

where $Y(K, L_H, L_C)$ is given by (P1), $R(t)$ given by (A1), $\underline{W}(t)$ given by (A2) and $x(t) = 1 - F(k_0(t))$,

I will use this dynamic system to simulate the model in Section 1.4.

For $P2$, $\frac{\overline{W}}{\underline{W}} = \frac{E(\theta|CG)}{E(\theta|HSG)}$. I can establish the existence of the pooling equilibrium under even weaker assumptions.

Lemma 2' *For $P2$, under the strategy profile that all types of agents go to college as soon as their current capital holdings $k \geq Q$, $\ln(\overline{W}/\underline{W})$ is increasing in the fraction,*

x, of agents going to college.

Proposition 1' *For $P2$, under the assumption that $Q < K(0)$, for Q sufficiently small, there exists a pooling equilibrium where all types of agents choose to go to college as soon as $k \geq Q$.*

The two corollaries continue to hold and the dynamic system that characterizes the equilibrium path remains valid with modified production technology and prices.

In general, there may exist separating equilibria in the sense that only a fraction of agents who can benefit from college self-select to attend college. In this case, Lemma 1 continues to hold, so the college-goers are those whose talent is above some threshold and who are not financially constrained. I discuss conditions for the existence of a separating equilibrium in the Appendix. The equilibrium evolution of the enrollment rates, the cut-off values of talent or the college premium is not necessarily monotonic. Furthermore, the actual enrollment rates and college completion rates imply that under mild conditions, the college premium is increasing in the cut-off value of talent. This means for a given enrollment rate, the lower the cut-off the smaller the wage gap. In other words, if we interpret the rising college premium as attracting less talented high school graduates to go to college, the decreasing minimum talent level tends to dampen the college premium. Intuitively, as we move to the extreme case of a pooling equilibrium, the effect of changes in the budget constraint on the college premium is smallest. Since talent is unobservable, the data is silent on the equilibirum selection. I restrict my attention to the pooling equilibrium for the following reasons: (1) the solution is closed-form and has nice properties; (2) the signaling effect brought by relaxing budget constraints in a separating equilibrium is likely to be even greater than that in a pooling equilibrium; (3) if we think empirically the talent cut-off in the separating equilibrium is decreasing, then the pooling equilibrium can be seen as a limiting case; (4) since our starting point is high school graduates, it is reasonable to assume that someone who can successfully complete the high school curriculum is prepared for college.

1.3.2.4 A Theoretical Bound of the Effect of the Signals

The next question is *how much* this story can account for the growth in the skill premium. This is of course an empirical question, but here I will derive a theoretical bound of the force of the signals. A widely held opinion is that compositional change in the labor force has little effect on the distribution of wage. This exercise addresses this concern theoretically and hopefully sheds some light on the kind of environment in which the force of signals tends to be strong.

Following Krusell et al. (2000), the growth rate in skill premium can be decomposed into two effects for the model with $P1$, the relative quantity effect and the relative efficiency effect,

$$g_{\ln \frac{\overline{W}}{\underline{W}}} \simeq (1-\rho)(g_{h_u} - g_{h_s}) + \rho(g_{\psi_s} - g_{\psi_u}),$$

where $g_x = \frac{dx/dt}{x}$; $h_s = x \int p dG$; $h_u = 1 - x \int p dG$; $\psi_s = E[\theta | CG]$ and $\psi_u = E[\theta | HSG]$.

The change in the distribution of signals leads to a change in the average talent given a signal, which amounts to a change in the efficiency of skilled labor relative to that of unskilled labor. To maximize the effect of the signals, we must choose the underlying parameters to maximize the relative efficiency effect $g_{\psi_s} - g_{\psi_u}$:

$$\sup_{\substack{G_t(\cdot) \\ p_t(\cdot)}} \frac{\int_0^{\overline{\theta}} \theta p(\theta) dG - \int_0^{\overline{\theta}} p(\theta) dG \int_0^{\overline{\theta}} \theta dG}{(1 - x(t) \int_0^{\overline{\theta}} p(\theta) dG)(\int_0^{\overline{\theta}} \theta dG - x(t) \int_0^{\overline{\theta}} \theta p(\theta) dG)} \dot{x}(t).$$

Remark 2 (1) $x(t)$ and $\dot{x}(t)$ are conveniently taken as given at each t. Though they are endogenous variables, I calibrate the enrollment rates to replicate those in data. So we may well take it as exogenous here.
(2) We allow $G_t(\cdot)$ and $p_t(\cdot)$ to be time-dependent. This maximizes the possible explanatory power of the signals and makes per period problem exactly the same. From now on, we will suppress the time subscript t.

Proposition 2 *The effect of signals is bounded by the negative growth rate of the fraction of people that don't attend college (if finite)*:

$$g_{\psi_s} - g_{\psi_u} \leq \frac{\dot{x}}{1 - x} = -g_{1-x}.$$

This result suggests that the signals work most effectively when the education can perfectly sort out the highest talents. Consider the following example in which there are only two talents, 1 or 0.

Example 1 *There is a fraction of ε (close to 0) of people with talent of 1 and the remaining are of talent 0. As a result, $E(\theta) = \varepsilon$. Suppose people with high talent can pass the exam almost surely, while people with low talent have the probability of success decreasing overtime in the following fashion:*

$$p_t(0) = \frac{1}{1 + x(t)}.$$

Note that at each instant of time the probability of success is still weakly increasing in the talents. The exam costs nothing. Then, one can verify that

$$\frac{E(\theta|\textit{with degree})}{E(\theta|\textit{without degree})} \to \frac{1}{1 - x}, as\ \varepsilon \to 0.$$

$$g_{\psi_s} - g_{\psi_u} = \frac{d}{dt} \ln \frac{E(\theta|\textit{with degree})}{E(\theta|\textit{without degree})} \to -g_{1-x}, as\ \varepsilon \to 0.$$

Note that in this example, the sorting mechanism becomes more and more efficient overtime, which also contributes to the growth of skill premium. This example shows that the suggested bound can be achieved in the limit. However, in the setting where the probabilities of success are constant overtime, we would expect in general slower growth in skill premium. The bottom line is that in an economy in which the distribution of degrees is highly upward skewed, the education signal has a bigger force.

Now we do a simple counterfactual calculation. Take the college enrollment rates from 1969 to 2005 and compute g_{1-x}[3]. Then, I take the skill premium in 1969, and let it grow at the maximum theoretical bound $-g_{1-x}$, whereby I get the fictitious wage gap in the dashed line contrasted with the real data, as is illustrated in Figure 1.5.

[Figure 1.5 about here.]

The signals, theoretically, have the potential to generate all of the growth in skill premium. But as will be clear in Section 1.4, our hands are tied significantly by the specification and parameterization of the model.

1.3.3 Optimality

In the current environment, there are two potential sources of inefficiency: the information problem represented by the private information of talents and the problem of missing credit market. We will investigate the consequences of these two problems one by one. In both cases, the objective of the social planner is to maximize period total output.

1.3.3.1 Benchmark One: Complete Information

Assume a social planner observes the individual talents. For $P1$, the social planner

[3] Since in the proof of the above proposition, $\dot{x}$ is assumed to be positive. I simply replace any negative growth in the data with zero.

simply chooses θ^* so that all agents with talent above θ^* are educated at a cost Q:

$$\begin{aligned}\Gamma(\theta^*) &= \max_{\theta^*}\{\lambda(1-\int_{\theta^*}^{\overline{\theta}} pdG)^{\rho-1}(\int_0^{\overline{\theta}} \theta dG-\int_{\theta^*}^{\overline{\theta}} \theta pdG)+\upsilon[K-(1-G(\theta^*))Q]^\rho \\ &\quad +(1-\lambda-\upsilon)(\int_{\theta^*}^{\overline{\theta}} pdG)^{\rho-1}\int_{\theta^*}^{\overline{\theta}} \theta pdG\}^{1/\rho} \\ &\quad s.t.\ \ 0\leq\theta^*\leq\overline{\theta}.\end{aligned}$$

This is not a concave problem and the solution is messy. Let $\rho = 1$ for tractability.

Proposition 3 *Consider* $\rho = 1$ with $P1$. *In cases in which* $2\lambda \geq 1-\upsilon$ *holds or both* $2\lambda < 1-\upsilon$ *and* $(1-2\lambda-\upsilon)\overline{\theta}p(\overline{\theta}) < \upsilon Q$ *hold*, *it is optimal not to provide education at all. If* $2\lambda < 1-\upsilon$ and $(1-2\lambda-\upsilon)\overline{\theta}p(\overline{\theta}) \geq \upsilon Q$, *the optimal cut-off in talent* θ^* *is given by* $(1-2\lambda-\upsilon)\theta^* p(\theta^*) = \upsilon Q$.

In cases where production relies more on unskilled labor than on skilled labor, or in cases where the opportunity cost of investing in education is high, it may be optimal not to provide education at all. But with incomplete information, there may still exist pooling equilibria defined in Section 1.3.2.3. The individual incentive to self-signal the talent causes both misallocation of factors and a waste of resources. More generally, in all those pooling equilibria, after some finite length of time, the economy will always over-invest in education, even though it may never reach the optimal amount of skilled labor even in the limit.

With $P2$, the degrees are irrelevant since talents are perfectly substitutable and the social planner simply uses all available resources.

Proposition 3' *For* $P2$, *the social planner employs all labor and capital and the period output is* $A[\alpha K^\rho + \beta(E(\theta))^\rho]^{1/\rho}$.

In the case with $P2$, there is no need to invest in education if education serves purely as a signal.

1.3.3.2 Benchmark Two: Relaxing Borrowing Constraints

In this section, agents of the same generation are allowed to borrow from each other. Let $b(t)$ be the amount of debt (or credit) that the agent acquires before he receives his income, which has to be paid back at the end of that period.

Definition 2 Equilibrium with within-generation credit markets

An equilibrium of this economy is a list $(c(t), k(t), sh(t), b(t), R(t), \overline{W}(t), \underline{W}(t))$ for each agent (θ, k_0), given initial capital distribution $F(\cdot)$ and distribution of talent $G(\cdot)$ and the exogenous positive saving rate σ and the production technology, so that

(i) Agents optimally choose $sh(K(t))$ and $b(t)$, given $R(t)$, $\overline{W}(t)$, $\underline{W}(t)$;

(ii) Firm maximizes period profit;

(iii) Factor markets clear;

(iv) Credit markets clear: $\int_{\underline{\theta}}^{\overline{\theta}} \int_0^{\overline{k_0}} b(t; \theta, k_0) dG(\theta) dF(k_0) = 0$.

Notice that *Lemma 1* still holds. It is easy to construct an equilibrium in which all agents go to college from day 1.

Proposition 4 *Under Assumptions 1- 3 and $P1$, for Q sufficiently small, there exists an equilibrium in which all agents go to college from day 1.*

In this equilibrium, the college attendance rate is always 1 and the wage gap remains constant

$$\frac{\overline{W}}{\underline{W}} = \frac{1-\lambda-\upsilon}{\lambda} \left(\frac{\int pdG}{1-\int pdG}\right)^{\rho-1} \frac{\int \theta pdG}{\int pdG} \frac{1-\int pdG}{\int \theta dG - \int \theta pdG}.$$

Furthermore, for all economies that have an equilibrium with borrowing constraints as is defined in *Definition 1,* there is also an equilibrium with with-in generation credit markets as is defined in *Definition 2,* in which there is full attendance. The equilibrium with within generation credit markets is easier to support: it exists for even higher cost of education. Now the evolution of the aggregate capital is described by

$$\dot{K}(t) = \sigma[\upsilon(K(t) - Q)^{\rho} + \Pi]^{\frac{1}{\rho}}[4].$$

For the same set of parameters, the equilibrium with within generation credit markets has more skilled labor, less unskilled labor and less capital. Hence, only in an economy where skilled labor is very productive, the relaxation of borrowing constraint may bring about more output. More generally, from a social planner's point of view, relaxing borrowing constraint does not necessarily lead to a Pareto improvement with transfers, since this allows for more competition through unproductive signals. The equilibrium without credit markets converges to the benchmark equilibrium in the limit.

1.4 Calibration

1.4.1 Data

The relevant data series are the log wage gap between college graduates and high school graduates, the college enrollment rate and the college completion rate.

Skill premium. To be consistent with the theoretic prediction that cohorts born more recently when the signaling effect of a degree is stronger face higher premium than what earlier cohorts face, the calculation of college premium should be cohort-based. I computed the wage series using the CPS March data from 1969 to 2005 by age groups and focus on the age group 23-6. The construction process is essentially the same as Autor, Katz and Kearney (2008).

[4] $\Pi = \lambda(1 - \int pdG)^{\rho-1}(\int \theta dG - \int \theta pdG) + (1 - \lambda - \upsilon)(\int pdG)^{\rho-1}\int \theta pdG.$

College enrollment rate. The college enrollment rate is available from 1960 to 2006 from the American College Testing Program on NCES website. The enrollment rate is obtained by dividing the total number of college enrollment in a given year by the total number of high school completers, who graduated from high school and completed GED within the preceding 12 months.

College completion rate. Take the number of bachelor's degrees conferred by degree-granting institutions each year and divide it by the total college enrollment four years before. The degree data are available by year from 1970 to 2006 from NCES. The model counterpart is $\int_0^{\overline{\theta}} p(\theta)dG(\theta)$, the average passing rate of college-goers. I plot the series of college completion rates in Figure 1.6.

[Figure 1.6 about here.]

Initial income distribution in 1972. I take the wage/salary income distribution of the fulltime-fullyear-employed 40-50 years old in 1972 from CPS March. These people were likely to have children around 20-year-old in the same year. CPS sampling weights are used.

Cost of college. The cost of college in the model is the tuition, fees, room and board (TFBR) net grants and aids. The TFBR is available from 1976 to 2005 from College Board and the Grants and Aids are available from 1986 to 2006 on selected years. After interpolating the missing observations linearly, the real net cost is almost constant from 1986 to 2006, averaged at 5467 in 2006 dollars.

1.4.2 Calibration Strategy

The data is structured as follows. The model year refers to the year for which the skill premium is calculated. Within the same period in the model, the enrollment rate six years and college completion rate two years before the model year are used. This is to accommodate the fact that the skill premium is calculated for the age group 23-26. Since the annual degree data starts in 1970 and the skill premium series ends in 2005, the first period in the model is 1972, while the last is 2005.

In order to introduce more variability to the model, I allow the average talent given a college degree to grow and transform the formulae of wage gap to make use of the data of college completion rate. More specifically, let the average talent given a college degree follow a linear trend

$$h_t \equiv E[\theta|CG] = \frac{\int_0^{\overline{\theta}} \theta p_t(\theta) dG}{\int_0^{\overline{\theta}} p_t(\theta) dG} = h_0 + \gamma t.$$

The model is silent about the change in h_t, since the signaling effect from increasing enrollment works through a deteriorating wage offer to unskilled labor. In reality, there are reasons to believe that the average talent of a college graduate grows over time: better screening mechanism in college admission, or better college financing to the talented, or improving human capital accumulation through college, among others. Permitting h_t to grow over time in this reduced form of course increases the overall fit of our model to data, but we will see that the magnitude of the signaling effect modeled in this paper does *not* hinge much on the growth rate of h_t.

The data counterpart of college completion rate π_t is $\int p(\theta) dG$, whereby the models of wage differential are transformed into

$$\frac{\overline{W}}{\underline{W}_t} = \frac{1-\lambda-\upsilon}{\lambda}(\frac{x_t\pi_t}{1-x_t\pi_t})^{\rho-1}\frac{(h_0+\gamma t)(1-x_t\pi_t)}{\int\theta dG - x_t\pi_t(h_0+\gamma t)}; \tag{P1}$$

$$\frac{\overline{W}}{\underline{W}_t} = \frac{(h_0+\gamma t)(1-x_t\pi_t)}{\int\theta dG - x_t\pi_t(h_0+\gamma t)}; \tag{P2}$$

where x_t is the college enrollment rate. From the last section, the college completion rates rose sharply during the period 1985 to 1995. What does this imply? Assume $\gamma = 0$. It is easy to show that with $P2$, the wage gap is increasing in college completion rate as long as $h_0 \geq \int\theta dG$. With $P1$, the wage gap is increasing only if $h_0 > \int\theta dG$ and ρ is sufficiently close to 1. In the calibrated models, it is true for both productions that the growth in completion rates helps generating some portion of the college premium. This may be interpreted as a change of the talent distribution over time, or changes in the college screening technology.

Now we are ready to discuss the measurement of the signaling effect. To facilitate discussion, I restrict my attention to $P2$. Recall from Section 1.3.2.4 that in the model, the growth rate of skill premium has two components, the relative quantity effect and the relative efficiency effect. $P2$ only has the relative efficiency effect: there is no general equilibrium effect of changes in skilled/unskilled labor composition on college premium. In other words, when I vary the enrollment rate, the variation in the skill premium reflects solely the relative efficiency effect, which is exactly the signaling effect that I'm interested in. Hence, the signaling effect can be measured by a counterfactual simulation, in which I fix the enrollment rate constant at the initial level and simulate the wage gap. In the absence of college completion rate data, the wage gap is constant if $\gamma = 0$. However, in the transformed model with the completion rate data, there is some growth in the wage gap even if $\gamma = 0$. The signaling effect is then the residual contribution to the growth in college premium on top of the prediction of the counterfactual model. I calculate the compound annual growth rate (CAGR) of the wage gap predicted by a model holding enrollment fixed and compare it with the CAGR of the wage gap predicted by a calibrated model with endogenous enrollment rates. The measure of the signaling effect is the percentage of growth rate that is contributed by varying enrollment rates:

$$1 - \frac{CAGR(\ln \frac{\overline{W}}{\underline{W}}|\text{holding enrollment fixed, }\gamma)}{CAGR(\ln \frac{\overline{W}}{\underline{W}}|\gamma)}. \tag{1.2}$$

Essentially, for any γ, I can compute the measure of the signaling effect within a model parametrized by γ (call it model γ) in the above way. As I vary γ, the overall fit of the model varies and can be measured likewise by

$$\frac{CAGR(\ln \frac{\overline{W}}{\underline{W}}|\gamma)}{CAGR(\ln \frac{\overline{W}}{\underline{W}}|\text{data})}.$$

This is the percentage of growth explained by model γ with respect to data. Multiplying the above two measures, I come to a measure of the overall signaling effect of model γ. We will see that this measure is remarkably stable across different values of γ.

To tackle the difficult problem brought by the unobservables, I ask the following two questions: one, what is the contribution of signals given that the unobservable behave in the most favorable way to me; two, what is the effect of signals as I limit the contribution of the unobservables. To answer them, I follow three steps.

In the first step, I jointly estimate some key parameters in a non-linear-least-square model of wage gap. More specifically, for $P1$, I normalize $h_0 = 1$, take $\upsilon = \lambda = \frac{1}{3}$, and jointly estimate γ, $E\theta$ and ρ; for $P2$, I take $\alpha = \frac{1}{3}, \beta = \frac{2}{3}$, normalize $h_0 = 1$, and estimate γ and $E\theta$. But, all I take from this stage is the value of γ. I interpret this value as representing the most favorable term I can get from the unobservables. Details of the estimation are available upon request.

In the second step, I calibrate the model in the standard fashion, taking γ from the first step as given. In particular, the saving rate σ is pinned down by minimizing the distance between the model enrollment rates and the data.

In the third step, I calibrate models which correspond to different values of γ, ranging from 0 to the first step estimate. I look at the measurement of the signaling effect and find it to be quite constant across different γ.

1.4.3 Calibration Results

1.4.3.1 P1

The first stage estimation for P1 yields $\gamma = 0.712\%$. To gain a sense of the magnitude of γ, the avarage talent of college graduates grows to 1.23 times the original level within 33 periods. It turns out that with the first stage estimate of γ, the model over-predicts the growth in college premium. Hence, in the calibration, I pick the γ that matches the model prediction of college premium in the last period with that in the data.

In the second stage, I calibrate the model as follows.

$Model$	$Value$	$Source$
M[5]	45000	Decision rule
γ	0.5%	Match last period model college premium with data
h_0	1	Normalization
x_0	0.5006	College enrollment rate in 1966[6]
Q	5467	Real TFRB net aids averaged over 1986 and 2006
$F(\cdot)$	$--$	Income distribution in 1972 times $F^{-1}(1-x_0)$
K_0	20816	Mean of $F(\cdot)$
υ	0.3	Average capital share of national income in NIPA
ρ	0.98	Monotonicity of skill premium in enrollment rate
λ	0.3283	To match the initial college premium in 1972
σ	$2.82e-7$	To match model enrollment rate with the data

M requires some explanation. M scales the productivity of talent to a scale comparable to that of capital, so that in each period the decision rule $p(0)(\overline{W}-\underline{W})-RQ>0$ holds. The value of ρ implies strong substitutability among the three inputs. Krusell et al.(2000) estimate the elasticity of substitution between unskilled labor and equipment to be 1.67 and that between skilled labor and equipment to be 0.67, which suggests some substitutability between unskilled labor and the combo of skilled labor and capital. In this model, ρ must be high enough to guarantee the monotonicity of the wage gap in enrollment rate. Both the growth rate γ and the trend in college completion rates contribute to the growth of the college premium. When holding the college enrollment rate fixed at the initial condition, the model still predicts around 85% of the growth. To be more specific, the CAGR of college premium in the model is 3.46%, while in the counterfactual with constant enrollment, it is 2.98%. This suggests that the signaling effect contributes around 14% in the growth of college premium (Panel 1.1).

[5] $h_t = M(h_0+\gamma t)$.To guarantee the existence of the pooling equilibrium, I need $p(0)(\overline{W}-\underline{W})-RQ>0$. A sufficient condition is that $\pi(\overline{W}-\underline{W})-RQ>0$. The scale of h_t guarantees that.

[6] The enrollment rate in 1966 is 0.5011. The difference results from a kernel density estimation of the income distribution.

[Panel 1.1 about here.]

1.4.3.2 P2

In the model with $P2$, the same parameter values apply unless noted below.

$Model$	$Value$	$Source$
M	100	Decision rule
γ	0.3435%	1st stage estimation
$E\theta$	0.9058	To match the initial college premium in 1972
Q	5467	Real TFRB net aids averaged over 1986 and 2006
α	$1/3$	NIPA
β	$2/3$	NIPA
ρ	-1	Empirical estimate, see Antras (2004)
σ	$4.5239e-005$	To match the model enrollment rate with the data

Now the CAGR of the model college premium is 3.36%, while in the counterfactual model it is 2.76%. Therefore, the signals contribute about 18% in the growth of college premium (Panel 1.2). Note that the model, by itself, is not an elaborate model about the evolution of the college enrollment, therefore it fails to catch the swing in the college enrollment rates. However, even if I feed the actual enrollment rate into the model, the prediction of college premium doesn't change much (Panel 1.3). The counterfactual prediction accentuates the trough and peak for obvious reasons. But the model is able to replicate the long-run trends subject to the limited source of variability.

[Panels 1.2 and 1.3 about here.]

1.4.3.3 Measuring the Signaling Effect

Now I restrict my attention to $P2$. I recalibrate the model for 30 equally spaced values of γ ranging from 0 to 0.3435%.

As is expected, the explanatory power of the model increases as I increase γ (see Figure 1.7). However, Figure 1.8 shows that the signaling effect of model γ is actually decreasing in γ. Hence, in terms of the overall effect of signaling, the estimate stays fairly constant within the range of 16-18.5% (see Figure 1.9).

The merit of this exercise is that we can be reasonably confident in saying that around 17% of the growth rate in college premium comes from the signaling mechanism modeled here. This estimate allows rooms for many other potential explanations to be at play at the same time, be it demographic change or skill-biased technological change or capital-skill complementarity, since it is conceptually equivalent to a γ less than the first step estimate. In general, with productions that allow decreasing return to scale in the skills, the increasing trend of enrollment rates changes the relative supply of skilled labor, which will tend to dampen the signaling effect. Hence the measure as defined in (1.2) tends to underestimate the effect of signaling, since it is a product not only of the signaling effect but also the general equilibrium effect of increased supply of skilled labor.

[Figures 1.7, 1.8 and 1.9 about here.]

1.5 Conclusion

Though the idea of education as a job market signal is well known, its application to the evolution of wage distribution hasn't been well articulated in theory. This paper is such an attempt. I have developed a model with agents heterogeneous in initial wealth and talent, who make schooling decisions. The growth in the college enrollment rate due to increased accessibility to college makes a high school diploma a clearer signal of low talent. If talent is useful in production, the college degree will be rewarded a higher premium relative to the high school diploma. This brings about a growing wage gap between college graduates and high school graduates. The model is calibrated, with two specifications of production technologies. The effect of signals on the college premium is estimated to be around 17% for models that can potentially allow for other explanations of rising college premium. Simplistic as it seems, the theory has a big potential to explain a wider range of phenomena. I close the paper with directions for future research.

One immediate extension is to extend the two dimensional choice variable to the multidimensional choice of getting bachelor's, master's or doctor's degree. Eckstein and Nagypal (2004) argues that the most important group contributing to the increase in college wage premium is workers with a postgraduate degree. This is consistent with my theory. The increase in the number of Bachelor's degrees issued will demand even higher degrees to effectively signal one's talent, which leads to the growing graduate school premium. It is conceivable that with a continuum of choice of levels of education, that varies from community colleges to the Ph.D. programs in top universities, the distribution of the education premium to each will fan out over time as the signals work their way through the distribution.

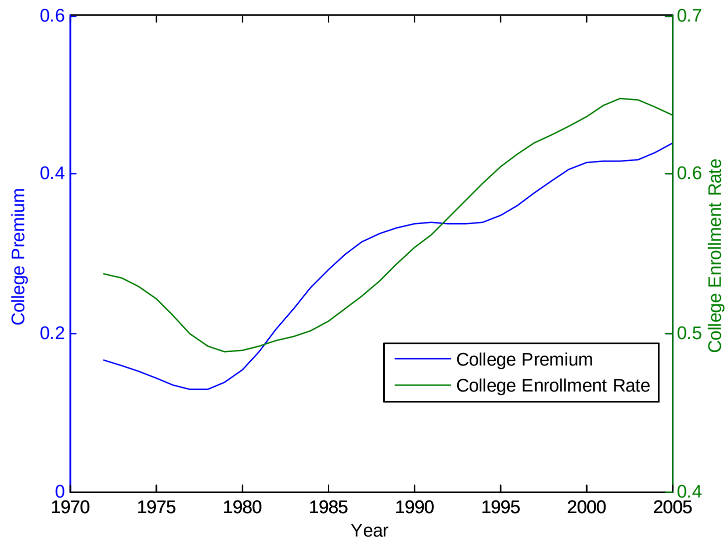
The framework can also be easily adapted to explaining the increasingly high premium of attending elite colleges. By casual observation, the best schools are becoming more and more accessible to the high talented students, thanks to more effective admission processes and more generous financial aid. As a result, the degree of elite schools must have become more correlated with talent than before. To estimate the fancy college premium and observe its evolution over time would be an interesting empirical question.

Another direction of research is to model the supply side of the college education. The key to the growing enrollment rate is the relaxation of household budget constraint over time through capital accumulation. But in reality there may be other ways that achieve the same effect. One example is the relaxation of the borrowing constraints, as is studied in Hendel, Shapiro and Willen (2001). Incorporating a sector of college will be a first step toward a general equilibrium approach. Colleges maximize some objective function by choosing costly admission processes. They can either admit students without much screening or undertake costly selection procedure. Colleges can be endowed with reputation such that in equilibrium some reputedly good colleges choose to be more selective, but will be compensated by higher prices they charge the students. Students in turn will be compensated by the top college premium. The story is more relevant if we can document the growing tuitions of top-notch schools and the growing returns to elite education.

Finally, one can conceive a full dynamic model, in which agents optimize over consumption and saving. Intuitively, this will help us more. Since the skill premium is growing over time, for subjective discount rate that is not too high, later cohorts will optimally choose to save more, which will allow their children to go to even fancier colleges or allow them to pursue postgraduate degrees, that will further enlarge the associated higher education premium. Combining a full dynamic model with a multiple or even continuum choice of levels of education would certainly make an elaborate model, though possibly analytically intractable. One would want to pay the extra cost of computation for more precise quantitative and policy-oriented analysis. After all, the parsimonious model we have here lays out the essential economic intuition just as well.

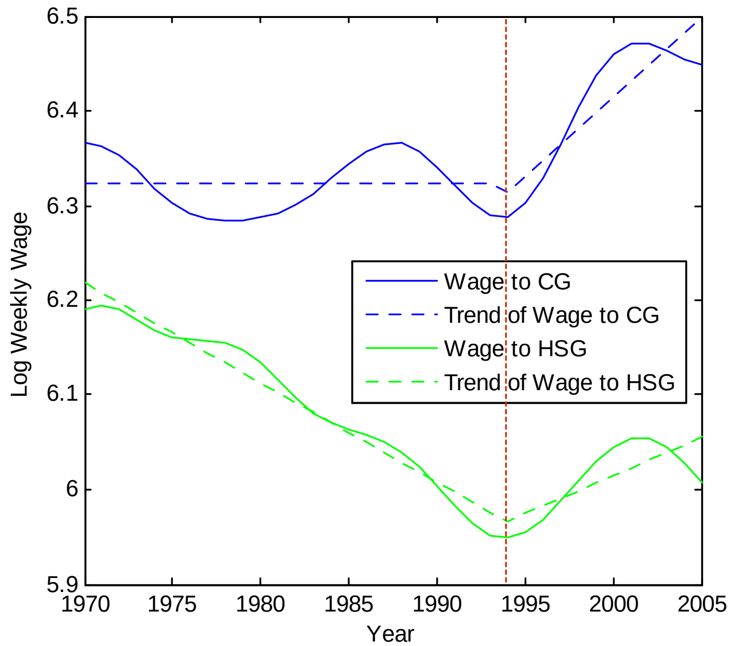
Figures and Tables

Figure 1.1: College Enrollment Rate and College Premium: HP-filtered, U.S. 1972-2005[7]



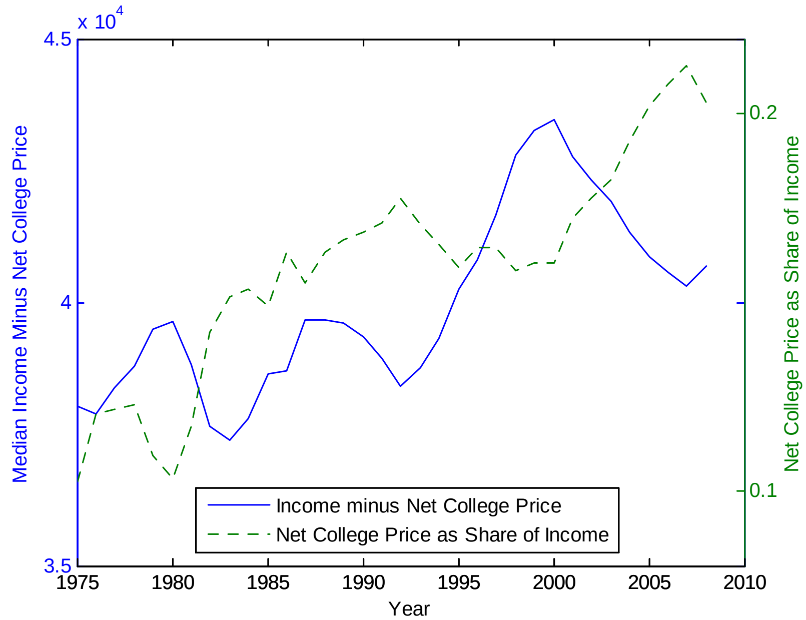
7 The college premium is the log weekly wage difference of a college graduate and a high school graduate for the age group 23-6, constructed from March CPS. Data are filtered by the Hodrick-Prescott Filter to remove the cycle.

Figure 1.2: HP-Filtered Log Weekly Wage to College Graduates and High School Graduates[8]



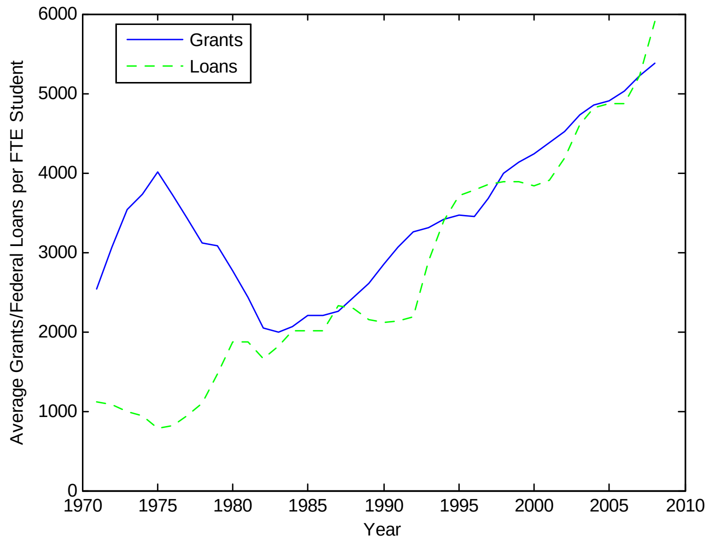
[8] Fitting a linear trend to the HP-filtered log weekly wage series yields: no trend with an average of 6.32 in $\log W^{CG}$ until 1993 while $\log W^{HSG} = 26.96 - 0.011 * Year$; from 1994 to 2005, $\log W^{CG} = -27.2323 + 0.0168 * Year$ and $\log W^{HSG} = -10.10 + 0.0081 * Year$; . All coefficients significant at 1%. The smoothness paramater in the HP-filter is 6.25.

Figure 1.3: The Difference of HP-filtered Median Household Income and Net College Price versus Net College Price as Share of Median Household Income (in 2008 Dollars) 1975/76-2007/08[9]



9 Data source: Trends in Student Aid 2009, Table 3; Trends in College Pricing 2009, Figure 5; U.S. Census Bureau, CPS, Annual Social and Economic Supplements, Tables H-6, H-8.

Figure 1.4: Average Grants and Federal Loans Per Full-Time-Equivalent Student (in 2008 Dollar) 1970/71-2008/09[10]



[10] Source: *Trends in Student Aid 2009*, Table 3.

Figure 1.5 Real and Fictitious Wage Gap

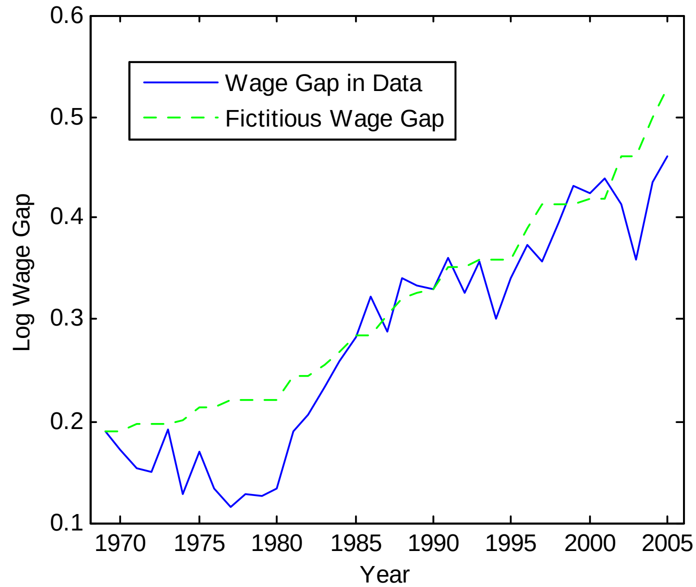
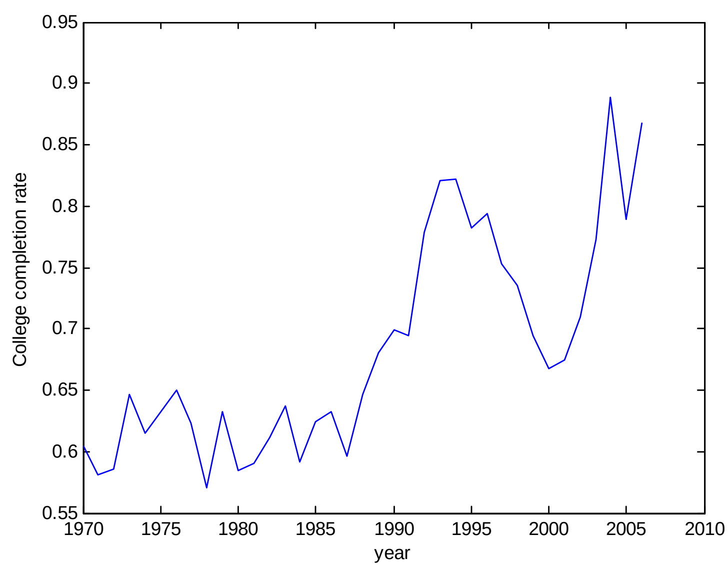
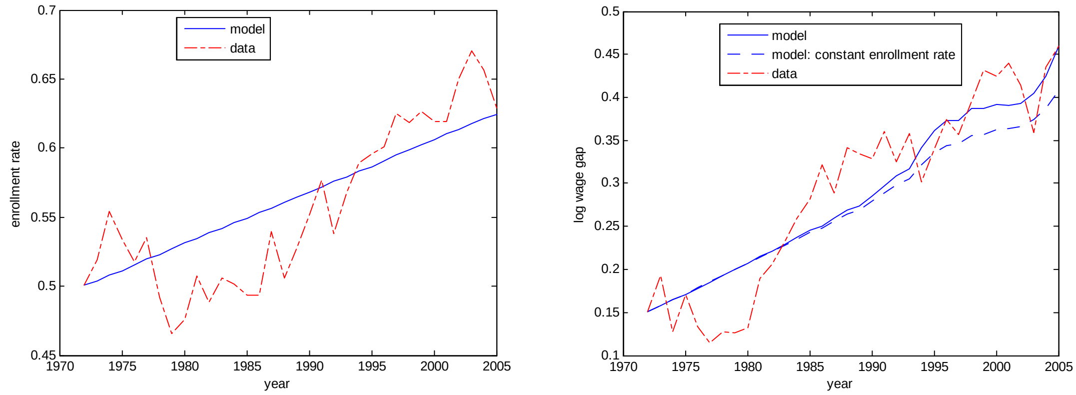


Figure 1.6 College Completion Rates[11]

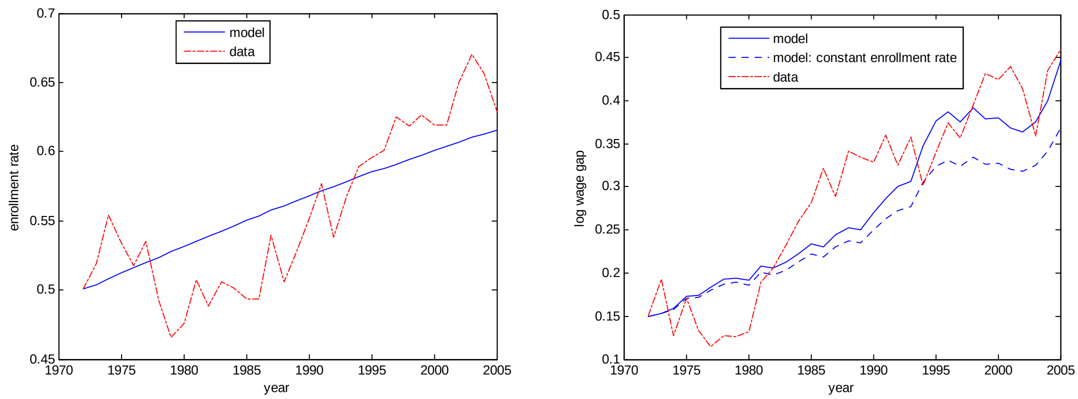


11 Source: NCES

Panel 1.1: Model prediction of college premium for $P1 : h_0 = 1, \rho = 0.98, \gamma = 0.5\%$



Panel 1.2: Model prediction of college premium for P2: $\gamma = 0.3435\%$



Panel 1.3: Prediction of college premium using endogenous enrollment rates vs. data

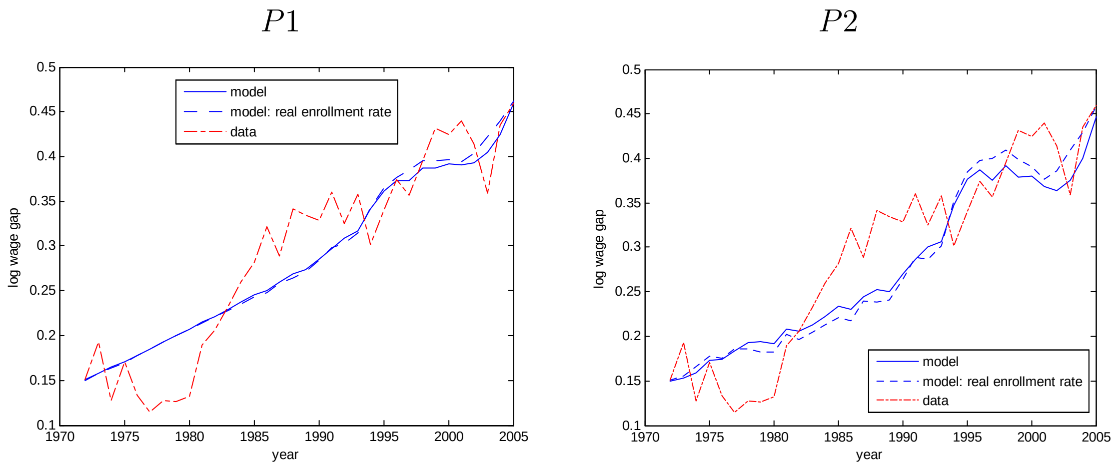


Figure 1.7: % of CAGR in College Premium Explained by Model γ

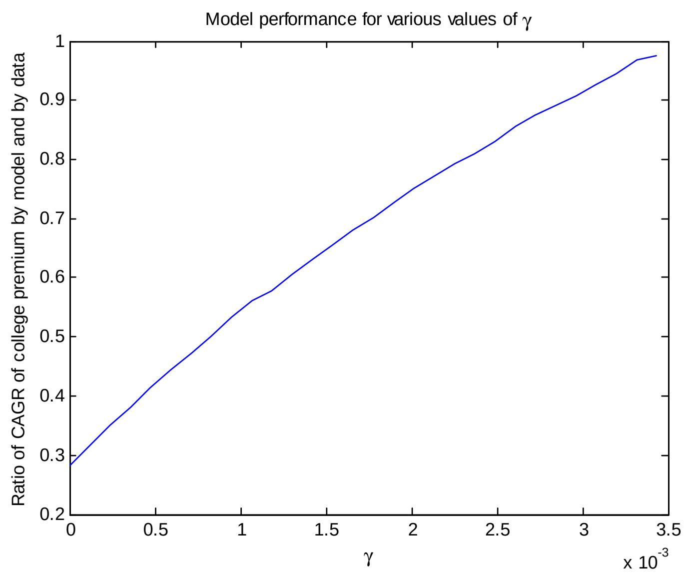


Figure 1.8: % of CAGR in College Premium in Model γ Explained by Signaling

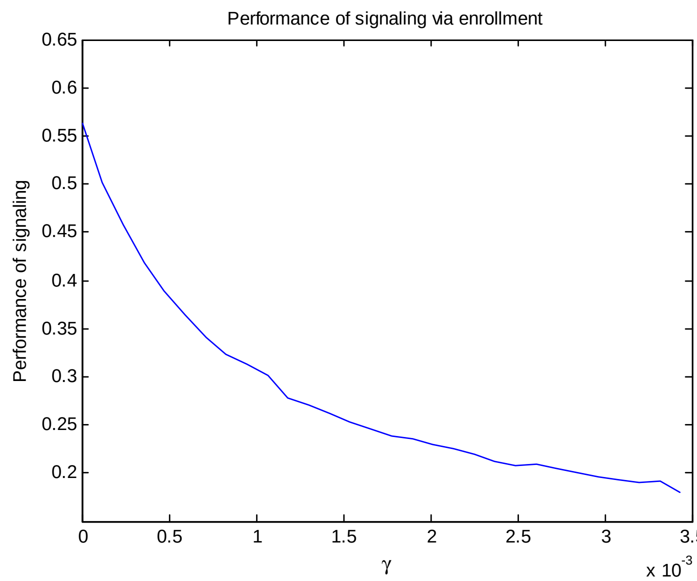


Figure 1.9: % of CAGR in College Premium Explained by Signaling for Model γ

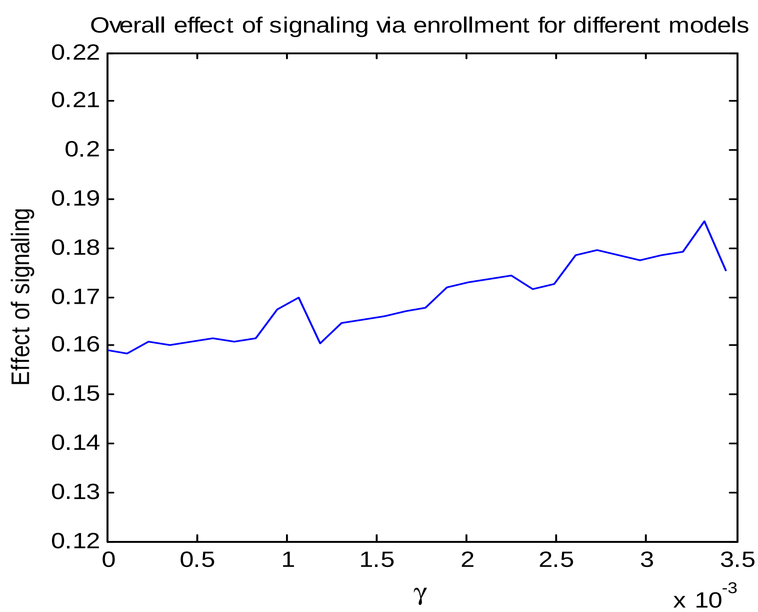


Table 1.1

Difference between Mean Parents' Income and Tuition and Fees net Grants and Federal Loans, by Income Groups and Types of Intitution, selected years (in 2008 dollars)

Year	Lowest 5th	Second 5th	Third 5th	Fourth 5th	Highest 5th
Public 4-year Institutions					
1985	15,110.83	36,668.34	55,474.54	79,295.33	145,075.60
1988	13,207.92	37,307.48	58,180.59	82,020.43	151,674.68
1991	16,125.12	36,804.96	55,521.14	80,999.30	156,513.43
1994	16,562.50	36,218.12	56,319.46	81,763.71	155,444.26
1998	18,708.25	39,304.02	60,786.51	88,908.78	160,174.07
2002	18,574.82	39,332.83	60,830.58	89,701.64	161,131.78
2006	18,505.38	39,554.77	61,375.99	92,128.54	168,814.00
Private Not-for-profit 4-year Institutions					
1985	13,440.63	35,098.99	53,755.32	75,810.07	168,132.03
1988	10,037.89	33,389.80	53,464.12	76,961.45	167,005.24
1991	13,456.60	33,829.92	50,992.03	76,435.25	153,830.67
1994	13,436.82	32,813.51	52,238.00	76,540.80	157,520.44
1998	15,863.34	35,309.56	56,792.18	82,824.42	156,307.34
2002	13,544.82	34,069.10	54,987.60	84,082.46	156,285.44
2006	13,430.70	34,850.36	54,063.27	83,832.11	167,342.44

Antres, P. 2004. "Is the U.S. Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution", *Contributions to Macroeconomics* 4, 1161-1162.

Archibald, R. and Feldman D., 2008. "How to Think About Changes in Higher Education Affordability", College of William and Mary Working Paper No. 76.

Autor, Katz and Kearney, 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists", *Review of Economics and Statistics*, 90(2), May, pp. 300-23.

Baumol, J. and Blackman, S. A. 1995. "How to Think About Rising College Costs", *Planning for Higher Education,* 23 (Summer) pp. 1-7.

Balart, P., (2010). "The Increase in College Premium and the Decline of College Wages: a Signaling Story"". Universitat Autònoma de Barcelona. Mimeo.

Bedard, K. 2001. "Human Capital versus Signaling Models: University Access and High School Dropouts", *Journal of Political Economy*, Vol. 109, no. 4, pp.749-75.

Cameron, S., and Heckman, J. 2001. "The Dynamics of Educational Attainment for Black, Hispanic, and White Males", *Journal of Political Economy*, Vol. 109, no. 3, pp. 455-499.

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

Carneiro, P. and Heckman, J. 2003. "Human Capital Policy", IZA Discussion Papers 821, Institute for the Study of Labor (IZA).

Card, D. and Lemieux, T. 2001. " Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-based Analysis", *The Quarterly Journal of Economics,* Vol. 116, No. 2, pp. 705-46.

Card, D. and Lemieux, T. 2001. "Going to College to Avoid the Draft: The Unintended Legacy of the Vietnam War", *American Economic Review,* May, Vol. 91(2), pp. 97-102.

Card, D. and DiNardo, J. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles", *Journal of Labor Economics*, Vol. 20, No. 4, pp.733-83.

Dynarski, S. 2003. "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion", *American Economic Review*, American Economic Association, vol. 93(1), pp. 279-288.

Eckstein, Z. and Nagypal E., 2004. "US Earnings and Employment Dynamics 1961 - 2002: Facts and Interpretation", 2004 Meeting Papers 182, Society for Economic Dynamics.

Fang, H. 2006. "Disentangling the College Wage Premium: Estimating a Model with Endogenous Education Choices," *International Economic Review*, Vol. 47, No. 4, November

2006, pp. 1151-85.

Guvenen, F. and Kuruscu B. 2009. "A Quantitative Analysis of the Evolution of the U.S. Wage Distribution: 1970-2000", Federal Reserve Bank of Minneapolis, Staff Report 427.

Hendel, I. Shapiro, J. and Willen, P. 2005. "Educational Opportunity and Income Inequality", *Journal of Public Economics*, Vol. 89, Issues 5-6, June 2005, pp. 841-70.

Hill, C., Winston, G. and Boyd, S. 2005. "Family Incomes and Net Prices at Highly Selective Private Colleges and Universities", *Journal of Human Resources*, Vol. 40, no. 4: pp. 769-790.

Kane, T. 1995. "Rising Public College Tuition and College Entry: How Well Do Public Subsidies Promote Access to College" National Bureau of Economic Research Working Paper 5164.

Kane, T. 2006. "Public Intervention in Postsecondary Education", in Hanushek, E.A. and Welch, F. editors, *Handbook of the Economics of Education*. Amsterdam: Elsevier.

Krusell, P., Ohanian, L., Rios-Rull, J. and Violante, G., 2000. "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis", *Econometrica*, Vol. 68, No. 5. (Sep., 2000), pp. 1029-1053.

Leslie, L. and Brinkman, P. 1988. *The Economic Value of Higher Education.* New York: Macmillan.

Long, B. 2007. "Do Loans Increase College Access and Choice? Examining the Introduction of Universal Student Loans", Federal Reserve Bank of Boston, NEPPC Working Paper 07-1

Riley, J. 2001. "Silver Signals: Twenty-Five Years of Screening and Signaling", *Journal of Economic Literature*, Vol. 39, No. 2, pp. 432-78.

Spence, M., 1973. "Job Market Signaling", *The Quarterly Journal of Economics*, Vol. 87, No. 3. (Aug., 1973), pp. 355-374.

Stiglitz, J. 1975. "The Theory of ’Screening’, Education, and the Distribution of Income", *The American Economic Review*, Vol. 65, No. 3 (Jun., 1975), pp. 283-300.

Taber, C. 2001. "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?", *The Review of Economic Studies*, Vol. 68, No. 3, pp. 665-91.

Winter C. 2009. "Accounting for the Changing Role of Family Income in Determining College Entry", IEW Working Paper iewwp402, Institute for Empirical Research in Economics.

1.6 Appendix

1.6.1 Theoretical Derivation

Lemma 1 Proof The value function is $v^i(k(t)) = \max\{v^{ci}(k(t)), v^{nci}(k(t))\}$, where

$$rv^c(k(t)) = p(\theta)\{(1-\sigma)[R(t)(k(t)-Q)+\overline{W}(t)] + \frac{dv^i}{dk}\sigma[R(t)(k(t)-Q)+\overline{W}(t)]\}$$
$$+[1-p(\theta)]\{(1-\sigma)[R(t)(k(t)-Q)+\underline{W}(t)] + \frac{dv^i}{dk}\sigma[R(t)(k(t)-Q)+\underline{W}(t)]\}.$$

$s.t. k(t) \geq Q$

$$rv^{nc}(k(t)) = (1-\sigma)[R(t)k(t)+\underline{W}(t)] + \frac{dv^i}{dk}\sigma[R(t)k(t)+\underline{W}(t)].$$

Given $\overline{W}, \underline{W}, R$, since it is optimal for(k_0, θ) to go to college,

$$\Delta(k,\theta) \equiv v^c(k) - v^{nc}(k) = (1-\sigma+\sigma\frac{dv}{dk})[p(\theta)(\overline{W}-\underline{W}) - RQ] > 0$$
$$\Rightarrow \Delta(k,\theta') = (1-\sigma+\sigma\frac{dv(k;\theta',k_0)}{dk})[p(\theta')(\overline{W}-\underline{W}) - RQ]$$
$$> 0. \forall \theta' > \theta$$

Hence, independent of the state variable k, (k_0, θ') would always prefers college *as long as* going to college is feasible, i.e. $k \geq Q$. Q.E.D.

.

Assumption 1 $\rho > 1 - \frac{(\int_0^{\overline{\theta}} \theta p dG - \int_0^{\overline{\theta}} p dG \int_0^{\overline{\theta}} \theta dG)(1-F(Q))}{\int_0^{\overline{\theta}} \theta dG - (1-F(Q)) \int_0^{\overline{\theta}} \theta p dG}$.

Lemma 2 Proof Under the specified strategy profile, the output and factor prices are

$$R(t) = \upsilon \Upsilon (K(t) - x(t)Q)^{\rho-1}. \quad \text{(A1)}$$

$$\overline{W}(t) = (1-\lambda-\upsilon)\Upsilon(x(t)\int_0^{\overline{\theta}} pdG)^{\rho-1} \frac{\int_0^{\overline{\theta}} \theta p dG}{\int_0^{\overline{\theta}} pdG}. \quad \text{(A2)}$$
$$\underline{W}(t) = \lambda \Upsilon (1 - x(t)\int_0^{\overline{\theta}} pdG)^{\rho-1} \frac{\int_0^{\overline{\theta}} \theta dG - x(t)\int_0^{\overline{\theta}} \theta p dG}{1 - x(t)\int_0^{\overline{\theta}} pdG},$$

where $\Upsilon = \{\lambda(1 - x\int_0^{\overline{\theta}} pdG)^\rho \frac{\int_0^{\overline{\theta}} \theta dG - x\int_0^{\overline{\theta}} \theta p dG}{1 - x\int_0^{\overline{\theta}} pdG} + \upsilon(K(t) - xQ)^\rho + (1 - \lambda - \upsilon)(x\int_0^{\overline{\theta}} pdG)^\rho \frac{\int_0^{\overline{\theta}} \theta p dG}{\int_0^{\overline{\theta}} pdG}\}^{\frac{1}{\rho}-1}$.

$$\begin{aligned}
\frac{d}{dx}(\frac{\overline{W}}{\underline{W}}) &= \frac{1}{1-x\int pdG}(\frac{\rho-1}{x}+\frac{\int\theta pdG-\int pdG\int\theta dG}{\int\theta dG-x\int\theta pdG}) \\
&\geq \frac{1}{1-x\int pdG}(\frac{\rho-1}{x_0}+\frac{\int\theta pdG-\int pdG\int\theta dG}{\int\theta dG-x_0\int\theta pdG}) \\
&\geq 0,\ \text{by } \textit{Assumption 1} \text{ and } x_0=1-F(Q).
\end{aligned}$$

This implies that $\ln(\frac{\overline{W}}{\underline{W}})$ is increasing in x. Note that $\forall Q$, *Assumption 1* is not empty.

Q.E.D.

Assumption 2 $\frac{1-\lambda-\upsilon}{\lambda} \geq (\frac{1-x_0\int pdG}{x_0\int pdG})^{\rho-2}\frac{\int\theta dG-x_0\int\theta pdG}{x_0\int\theta pdG}$.

Assumption 2 guarantees $\overline{W}(t) > \underline{W}(t)$.

Assumption 3 $Q < K(0)$.

Proposition 1 Proof The key is to verify that in the suggested equilibrium, all agents optimally make the schooling decision.

By *Lemma 1*, it is sufficient to look at the agent with the lowest talent and make sure he prefers to go to college. Suppose the college attendance is growing over time.

$$\begin{aligned}
& p(0)[\overline{W}(t)-\underline{W}(t)]-R(t)Q \\
= \ & \Upsilon\{p(0)[(1-\lambda-\upsilon)(x\int pdG)^{\rho-1}\frac{\int\theta pdG}{\int pdG} \\
& -\lambda(1-x\int pdG)^{\rho-1}\frac{\int\theta dG-x\int\theta pdG}{1-x\int pdG}]-\upsilon(K(t)-xQ)^{\rho-1}Q\},
\end{aligned}$$

where Υ, as is defined in *Lemma 2*, is positive. *Assumption 2* and *Lemma 2* implies

$(1-\lambda-\upsilon)(x\int pdG)^{\rho-1}\frac{\int \theta pdG}{\int pdG}-\lambda(1-x\int pdG)^{\rho-1}\frac{\int \theta dG - x\int \theta pdG}{1-x\int pdG}$ is increasing in x. Now

$$\begin{aligned}
& p(0)[\overline{W}(t)-\underline{W}(t)]-R(t)Q \geq \Upsilon\{p(0)[(1-\lambda-\upsilon)(x_0\int pdG)^{\rho-1}\frac{\int \theta pdG}{\int pdG} \\
& -\lambda(1-x_0\int pdG)^{\rho-1}\frac{\int \theta dG - x_0\int \theta pdG}{1-x_0\int pdG}]-\upsilon(K(0)-Q)^{\rho-1}Q\}\geq 0. \\
\Rightarrow\ & p(0)[(1-\lambda-\upsilon)(x_0\int pdG)^{\rho-1}\frac{\int \theta pdG}{\int pdG}-\lambda(1-x_0\int pdG)^{\rho-1}\frac{\int \theta dG - x_0\int \theta pdG}{1-x_0\int pdG}] \\
\geq\ & \upsilon(K(0)-Q)^{\rho-1}Q\equiv\Psi(Q).
\end{aligned}$$

By *Assumption 3,* $\frac{d\Psi(Q)}{dQ}=[K(0)-Q]^{\rho-2}[K(0)-\rho Q]>0$,with $\Psi(0)=0$; $\lim_{Q\to K(0)}\Psi(Q)=+\infty$. By *Assumption 2,*

$$p(0)[(1-\lambda-\upsilon)(x_0\int pdG)^{\rho-1}\frac{\int \theta pdG}{\int pdG}-\lambda(1-x_0\int pdG)^{\rho-1}\frac{\int \theta dG - x_0\int \theta pdG}{1-x_0\int pdG}]>0,$$

then there exists a $\widehat{Q}$ such that

$$\Psi(\widehat{Q})=p(0)[(1-\lambda-\upsilon)(x_0\int pdG)^{\rho-1}\frac{\int \theta pdG}{\int pdG}-\lambda(1-x_0\int pdG)^{\rho-1}\frac{\int \theta dG - x_0\int \theta pdG}{1-x_0\int pdG}].$$

For all $Q\leq\widehat{Q}$, $p(0)[\overline{W}(t)-\underline{W}(t)]-R(t)Q>0,\forall t$. So by *Lemma 1,* for Q sufficiently small, all agents want to go to college as soon as they can afford it. Lastly, for all those who are constrained, $\dot{k^i}=\sigma[R(t)k^i+\underline{W}(t)]>0$. This implies that indeed in the equilibrium there will be an increasing fraction of people who can afford education. Q.E.D.

Corollary 1 Proof An agent starts to go to college at time t that satisfies $k_0^i(t)+\int_0^t\dot{k^i}(s)ds=Q$, where the evolution of k^i follows $\dot{k^i}=\sigma[R(t)k^i(t)+\underline{W}(t)]$. At time t the faction of agents that goes to college is $1-F(k_0^i(t))$, which is increasing in t, since $k_0^i(t)$ is decreasing in t. Q.E.D.

Proposition 2 Proof I proceed in three steps.

Step 1: Transformation. Let $\widehat{p}(\theta)=\overline{\theta}p(\theta)g(\theta)$, which necessarily satisfies $\widehat{p}(\theta)\geq 0, 0\leq\int_0^{\overline{\theta}}\widehat{p}(\theta)d\theta\leq\overline{\theta}$. Let $\int_0^{\overline{\theta}}\theta dG\equiv a$. This problem is equivalent to a two-step maximization. Given a,

$$\sup_{\widehat{p}(\theta)} \dot{x}\overline{\theta}\frac{\int_0^{\overline{\theta}} \theta\widehat{p}d\theta - a\int_0^{\overline{\theta}} \widehat{p}d\theta}{(\overline{\theta} - x\int_0^{\overline{\theta}} \widehat{p}d\theta)(\overline{\theta}a - x\int_0^{\overline{\theta}} \theta\widehat{p}d\theta)}$$

$$s.t.\widehat{p}(\theta) \geq 0, 0 \leq \int_0^{\overline{\theta}} \widehat{p}(\theta)d\theta \leq \overline{\theta}, 0 \leq \int_0^{\overline{\theta}} \theta\widehat{p}d\theta \leq a\overline{\theta}$$

Then, maximize over all possible a.

Step 2: Change of variables. Let $y(\theta) = \int_0^{\theta} \widehat{p}(v)dv$. Integration by part gives $\int_0^{\overline{\theta}} \theta\widehat{p}d\theta = \int_0^{\overline{\theta}} \theta y'(\theta)d\theta = \overline{\theta}y(\overline{\theta}) - \int_0^{\overline{\theta}} y(\theta)d\theta$. The problem can be rewritten as

$$\sup_{\substack{y(\overline{\theta}),\\ \int_0^{\overline{\theta}} y(\theta)d\theta}} \dot{x}\overline{\theta}\frac{(\overline{\theta} - a)y(\overline{\theta}) - \int_0^{\overline{\theta}} y(\theta)d\theta}{(\overline{\theta} - xy(\overline{\theta}))[x\int_0^{\overline{\theta}} y(\theta)d\theta + \overline{\theta}(a - xy(\overline{\theta}))]}$$

$$s.t.\left\{\begin{array}{c} 0 \leq y(\overline{\theta}) \leq \overline{\theta}; y\,'(\theta) \geq 0; \\ \max\{0, \overline{\theta}(y(\overline{\theta}) - a)\} \leq \int_0^{\overline{\theta}} y(\theta)d\theta \leq (\overline{\theta} - a)y(\overline{\theta}). \end{array}\right\}$$

Step 3: Maximization. Firstly, $y(\overline{\theta})$ and $\int_0^{\overline{\theta}} y(\theta)d\theta$ can take values independently. Secondly, the objective is increasing in $y(\overline{\theta})$, but decreasing in $\int_0^{\overline{\theta}} y(\theta)d\theta$. But bigger $y(\overline{\theta})$ will increase the lowest level that $\int_0^{\overline{\theta}} y(\theta)d\theta$ can take.

If $y(\overline{\theta}) \leq a$, then the optimal values are $y(\overline{\theta}) = a$ and $\int_0^{\overline{\theta}} y(\theta)d\theta = 0$.

If $y(\overline{\theta}) \geq a$. Then at the optimum, no matter what value $y(\overline{\theta})$ takes, $\int_0^{\overline{\theta}} y(\theta)d\theta = \overline{\theta}(y(\overline{\theta}) - a)$. Substituting this relation into the objective function $\sup_{y(\overline{\theta})} \dot{x}\frac{\overline{\theta}-y(\overline{\theta})}{(\overline{\theta}-xy(\overline{\theta}))(1-x)}$. It is decreasing in $y(\overline{\theta})$. Hence, at the optimum, $y(\overline{\theta}) = a$ and $\int_0^{\overline{\theta}} y(\theta)d\theta = 0$.

In both cases, the maximum of the objective function is $\sup(g_{\psi_s} - g_{\psi_u}) = \dot{x}\frac{\overline{\theta}-a}{(\overline{\theta}-xa)(1-x)}$.

Now maximize with respect to a, $\sup(g_{\psi_s} - g_{\psi_u}) = \frac{\dot{x}}{1-x} = -g_{1-x}$,as $a \to 0$.Q.E.D.

Proposition 3 Proof Differentiate the objective function with respect to θ^* gives $g(\theta^*)[(2\lambda + \upsilon - 1)\theta^* p(\theta^*) + \upsilon Q]$. If $2\lambda \geq 1 - \upsilon$, maximum is obtained at $\theta^* = \overline{\theta}$. Suppose $2\lambda < 1 - \upsilon$. If $(1 - 2\lambda - \upsilon)\overline{\theta}p(\overline{\theta}) < \upsilon Q$, maximum is obtained at $\theta^* = \overline{\theta}$. Otherwise, first order necessary condition requires for $\theta^* \in [0, \overline{\theta}], (1 - 2\lambda - \upsilon)\theta^* p(\theta^*) = \upsilon Q$. SOC at θ^* gives

$[\lambda-(1-\lambda-\upsilon)][p(\theta^*)+\theta^* p'(\theta^*)]<0$. Hence $\Gamma(\theta^*)$ is a local maximum. It is easily shown that $\Gamma(\theta^*)>\Gamma(\underline{\theta})$ and $\Gamma(\theta^*)>\Gamma(\overline{\theta})$. Hence, θ^* achieves the global maximum. Q.E.D.

Proposition 4 Proof The agents' problem: the value function is $v^i(k(t))=\max\{v^{ci}(k(t)),v^{nci}(k(t))\}$, where,

$$rv^c(k(t))=p(\theta)\{(1-\sigma)[R(t)(k(t)-Q)+\overline{W}(t)]+\frac{dv^i}{dk}\sigma[R(t)(k(t)-Q)+\overline{W}(t)]\}$$

$$+[1-p(\theta)]\{(1-\sigma)[R(t)(k(t)-Q)+\underline{W}(t)]+\frac{dv^i}{dk}\sigma[R(t)(k(t)-Q)+\underline{W}(t)]\}$$

$$s.t. k(t)+b(t)\geq Q.$$

$$rv^{nc}(k(t))=(1-\sigma)[R(t)k(t)+\underline{W}(t)]+\frac{dv^i}{dk}\sigma[R(t)k(t)+\underline{W}(t)].$$

By the same logic as in *Proposition 1*,$\forall t, v^{ci}(k(t))-v^{nci}(k(t))=p(0)[\overline{W}(t)-\underline{W}(t)]-R(t)Q>0$. Now the factor prices in the proposed equilibrium are

$$\begin{aligned}\overline{W}(t) &= (1-\lambda-\upsilon)\widehat{\Upsilon}(\int pdG)^{\rho-2}\int\theta pdG.\\ \underline{W}(t) &= \lambda\widehat{\Upsilon}(1-\int pdG)^{\rho-2}(\int\theta dG-\int\theta pdG).\\ R(t) &= \upsilon\widehat{\Upsilon}(K(t)-Q)^{\rho-1},\end{aligned}$$

where $\Upsilon=\{\lambda(1-\int pdG)^{\rho-1}(\int\theta dG-\int\theta pdG)+\upsilon(K-Q)^{\rho}+(1-\lambda-\upsilon)(\int pdG)^{\rho-1}\int\theta pdG\}^{\frac{1}{\rho}-1}$.

$$\begin{aligned}& p(0)[\overline{W}(t)-\underline{W}(t)]-R(t)Q\\ =\ & \widehat{\Upsilon}\{p(0)[(1-\lambda-\upsilon)(\int pdG)^{\rho-2}\int\theta pdG-\lambda(1-\int pdG)^{\rho-2}(\int\theta dG-\int\theta pdG)]\\ & -\upsilon(K(t)-Q)^{\rho-1}Q\},\end{aligned}$$

By *Assumptions 1-3*,

$$
\begin{aligned}
& (1-\lambda-\upsilon)(\int pdG)^{\rho-2}\int \theta pdG-\lambda(1-\int pdG)^{\rho-2}(\int \theta dG-\int \theta pdG) \\
> \quad & (1-\lambda-\upsilon)(x_0\int pdG)^{\rho-1}\frac{\int \theta pdG}{\int pdG}-\lambda(1-x_0\int pdG)^{\rho-1}\frac{\int \theta dG-x_0\int \theta pdG}{1-x_0\int pdG}>0.
\end{aligned}
$$

$\exists\, \widehat{Q}^*$ such that

$$
\begin{aligned}
& p(0)[(1-\lambda-\upsilon)(\int pdG)^{\rho-2}\int \theta pdG-\lambda(1-\int pdG)^{\rho-2}(\int \theta dG-\int \theta pdG) \\
= \quad & \upsilon(K(0)-\widehat{Q}^*)^{\rho-1}.
\end{aligned}
$$

It is readily seen that $\widehat{Q}^*>\widehat{Q}$. $\forall Q<\widehat{Q}^*$, $p(0)[\overline{W}(t)-\underline{W}(t)]-R(t)Q\geq 0$. Hence, everyone attends college at all times, while the wage gap remains constant. Q.E.D.

1.6.2 Separating Equilibrium

I sketch here the proof of the existence of a separating equilibrium for $P2$. This exercise can be repeated for $P1$.

Assumption 4 $p(0)=0$.

Assumption 5 $g(\overline{\theta})=0$.

Assumption 6 $\beta p(E(\theta))[\frac{\int_{E(\theta)}^{\overline{\theta}}\theta pdG}{\int_{E(\theta)}^{\overline{\theta}}pdG}-\frac{\int \theta dG-(1-F(Q))\int_{E(\theta)}^{\overline{\theta}}\theta pdG}{1-(1-F(Q))\int_{E(\theta)}^{\overline{\theta}}pdG}]>\alpha[K_0-(1-G(E(\theta))Q]^{\rho-1}Q$.

Assumption 7 $1>2\int_{E(\theta)}^{\overline{\theta}}pdG$.

First, at time t, fix $x_t=1-F(k_{0t})$ and K_t. By Lemma 1, the cut-off level of talent $\widehat{\theta}_t$ satisfies $p(\widehat{\theta}_t)(\overline{W}_t-\underline{W}_t)=R_tQ$, or

$$
\beta p(\widehat{\theta}_t)[E_t(\theta|CG)-E_t(\theta|HSG)]=\alpha[K_t-(1-G(\widehat{\theta}_t))x_tQ]^{\rho-1}Q. \tag{1.3}
$$

The LHS is further equal to $\beta p(\widehat{\theta}_t)[\frac{\int_{\widehat{\theta}_t}^{\overline{\theta}}\theta pdG}{\int_{\widehat{\theta}_t}^{\overline{\theta}}pdG}-\frac{\int \theta dG-x_t\int_{\widehat{\theta}_t}^{\overline{\theta}}\theta pdG}{1-x_t\int_{\widehat{\theta}_t}^{\overline{\theta}}pdG}]$. One can show that RHS is decreasing in $\widehat{\theta}_t$ while LHS is increasing in $\widehat{\theta}_t$ if

(a) $\widehat{\theta}_t<\int \theta dG$; and (b) $1>2x_t\int_{\widehat{\theta}_t}^{\overline{\theta}}pdG$. We will restrict the solution $\widehat{\theta}_t$ to $[0,E(\theta)]$

to guarantee (a). (b) is ensured by (a) and Assumption 7. Note that we can rewrite (2) as $2[x_t(1-G(\widehat{\theta}_t))]\frac{\int_{\widehat{\theta}_t}^{\overline{\theta}} pdG}{1-G(\widehat{\theta}_t)} = 2\cdot$enrollment rate$_t\cdot$college completion rate$_t$. One can verify using the U.S. data from 1972 to 2005 that the above inequality is always satisfied. Under Assumptions 4 and 5, in order for (1.3) to have a solution in $[0, E(\theta)]$, one requires Assumption 6. Hence, for all values of x_t and K_t, there exists a cut-off point $\widehat{\theta}_t \in [0, E(\theta)]$, such that all agents with $\theta \geq \widehat{\theta}_t$ choose to go to college as long as they can afford it.

Second, the dynamic system that characterizes the equilbrium path is

$$\begin{aligned}
\dot{K}_t &= \sigma A[\alpha(K_t-(1-F(k_{0t}))(1-G(\widehat{\theta}_t))Q)^\rho + \beta E(\theta)^\rho]^{1/\rho}; \\
\dot{k}_{0t} &= -\sigma[R_t Q + \underline{W}_t];
\end{aligned}$$

where $R_t = \Upsilon\alpha[K_t-(1-F(k_{0t}))(1-G(\widehat{\theta}_t))Q]^{\rho-1}$, $\underline{W}_t = \Upsilon\beta[E(\theta)]^{\rho-1}\frac{\int \theta dG-(1-F(k_{0t}))\int_{\widehat{\theta}_t}^{\overline{\theta}} \theta pdG}{1-(1-F(k_{0t}))\int_{\widehat{\theta}_t}^{\overline{\theta}} pdG}$ and $\Upsilon = A[\alpha(K_t-(1-F(k_{0t}))(1-G(\widehat{\theta}_t))Q)^\rho+\beta E(\theta)^\rho]^{1/\rho-1}$. The cut-off of talent satisfies

$$\beta p(\widehat{\theta}_t)[\frac{\int_{\widehat{\theta}_t}^{\overline{\theta}} \theta pdG}{\int_{\widehat{\theta}_t}^{\overline{\theta}} pdG} - \frac{\int \theta dG-(1-F(k_{0t}))\int_{\widehat{\theta}_t}^{\overline{\theta}} \theta pdG}{1-(1-F(k_{0t}))\int_{\widehat{\theta}_t}^{\overline{\theta}} pdG}] = \alpha[K_t-(1-G(\widehat{\theta}_t))(1-F(k_{0t}))Q]^{\rho-1}Q.$$

The initial conditions are $K_0 = \int_0^{\overline{k}_0} k_0 dF, k_{00} = Q$.

Under Assumptions 4-7, the solution to the above dynamic system exists. However, the equilibrium paths of the cut-off point of talent, the enrollment rates and the college premium are not necessarily monotone.

1.6.3 Calibration

1.6.3.1 Data

Skill premium. The raw data are taken from the CPS March from 1969 to 2005. Only fullyear fulltime workers that have positive wage and schooling are considered. They are grouped by ages. The relevant age group here is those age 23 to 26. The log deflated weekly wage, which is the income from wage and salary divided by weeks worked, is then regressed on dummies of education, geographic region and race, by sexes. The education is the highest education attainment reported, high school dropouts, high school graduates, some college, college graduates or above. The geographical region is grouped in four, Northeast region, Midwest region, South region and West region. For more definitions on the data precession, please refer to Autor, Katz and Kearney (2008). For each sex, the log wage gap is the difference between the prediction for a white college graduate (but with no graduate degree) who lives in the average geographic region and that for a high school graduate counterpart. The log wage gap is the mean of the log wage gaps of the two sexes, weighted by their hours worked. CPS weights are used. I have explored variations of this basic set-up, including the log 10-year-income gap, the log wage gap between a 23 year old college graduate and a 19 year old high school graduate, among others. The results don't differ much.

Initial income distribution in 1972. Annual income from wage and salary are converted into 2006 dollars by CPI index. CPS weights are used. To match the initial enrollment rate, which is 0.5006, I find the 50th percentile in the empirical income distribution and normalize it to be equal to Q. That is,

$$F(Q/\xi) = 1 - 0.5006.$$

Further multiply all income in the sample by ξ and this gives the $F(\cdot)$ in the model. ξ can be thought of as the share of income that goes to educational expenses.

Cost of college. The real cost of college, computed using the data published in *Trends in College Pricing 2006* and *Trends in Student Aid 2007,* do not show an obvious trend from 1986 to 2006. I take Q to be the average over all these years, which is 5467.

1.6.4 Proofs for Model with $P2$

Lemma 2' Proof Let $a = E(\theta)$. The gross output is

$$Y = A\{\alpha(K - xQ)^\rho + \beta[L_H E(\theta|HSG) + L_C E(\theta|CG)]^\rho\}^{1/\rho}.$$

$$\begin{aligned} \overline{W} &= \Lambda\beta a^{\rho-1} E(\theta|CG); \\ \underline{W} &= \Lambda\beta a^{\rho-1} E(\theta|HSG); \\ R &= \Lambda\alpha(K - xQ)^{\rho-1}, \end{aligned}$$

where $\Lambda = A\{\alpha(K - xQ)^\rho + \beta[L_H E(\theta|HSG) + L_C E(\theta|CG)]^\rho\}^{1/\rho-1}$.

$$\begin{aligned} \ln\frac{\overline{W}}{\underline{W}} &= \frac{E(\theta|CG)}{E(\theta|HSG)} \\ &= \frac{\int \theta p dG}{\int p dG} \frac{1 - x\int p dG}{\int \theta dG - x\int \theta p dG} \end{aligned}$$

increasing in x. Q.E.D.

Proposition 1' Proof

$$\begin{aligned} & p(0)(\overline{W} - \underline{W}) - RQ \\ = \ & \Lambda\{p(0)\beta a^{\rho-1}[E(\theta|CG) - E(\theta|HSG)] - \alpha(K - xQ)^{\rho-1}Q\} \\ \geq \ & \Lambda\{p(0)\beta a^{\rho-1}[E(\theta|CG) - E(\theta|HSG)] - \alpha(K(0) - Q)^{\rho-1}Q\}. \end{aligned}$$

By the same token, there exists $\widetilde{Q}$, s.t.

$$p(0)\beta a^{\rho-1}[E(\theta|CG) - E(\theta|HSG)] = \alpha(K(0) - \widetilde{Q})^{\rho-1}\widetilde{Q}.$$

For all $Q \leq \widetilde{Q}$,

$$p(0)(\overline{W} - \underline{W}) - RQ \geq 0, \forall t.$$

Moreover, when this is the case, there will be indeed an increasing number of agents going to college. Q.E.D.

1.6.5 1st Stage Estimation Results

P1 The non-linear model for wage gap is

$$\ln(\frac{\overline{W}}{\underline{W}})_t = \ln\frac{1-\lambda-\upsilon}{\lambda} + (\rho - 1)\ln(\frac{x_t\pi_t}{1 - x_t\pi_t}) + \ln\frac{(h_0 + \gamma t)(1 - x_t\pi_t)}{\int \theta dG - x_t\pi_t(h_0 + \gamma t)},$$

which is transform into a statistical model with series of $\ln(\frac{\overline{W}}{\underline{W}})_t$, x_t and π_t:

$$\begin{aligned} y_t &= \ln(\frac{\overline{W}}{\underline{W}})_t + \ln\frac{x_t\pi_t}{1-x_t\pi_t} = \ln\frac{1-\lambda-\upsilon}{\lambda} + \rho\ln\frac{x_t\pi_t}{1-x_t\pi_t} + \ln\frac{(1-x_t\pi_t)(h_0+\gamma t)}{\int\theta dG - x_t\pi_t(h_0+\gamma t)} + \varepsilon_t \\ &= b_0\ln\frac{x_t\pi_t}{1-x_t\pi_t} + \ln\frac{(1-x_t\pi_t)(1+b_2 t)}{b_1 - x_t\pi_t(1+b_2 t)} + \varepsilon_t. \end{aligned}$$

Normalize $h_0 = 1$. Take $\upsilon = \lambda = 1/3$, and jointly estimate ρ, $\int\theta dG$ and γ. Note that $\int\theta dG$ and γ are relative to h_0 as a result of normalization.

Source	SS	df	MS	Number of obs	34
Model	7.4987	3	2.499	R-squared	0.9925
Residual	.05678	31	.0018	AdjR-squared	0.9918
				RootMSE	.04280
Total	7.5555	34	.0018	Res.dev.	-120.94
yt	Coef.	Std.Err.	t	P>t	[95% Conf.Interval]
$/b0$	.8594	.0353	24.35	0.000	.7874 .9313
$/b1$	.9985	.0269	37.06	0.000	.9436 1.053
$/b2$	.0071	.0011	6.74	0.000	.0049 .0093

Parameter b0 taken as constant term in model & ANOVA table

This implies

$$\begin{aligned} \rho &= 0.869356; \\ \int\theta dG &= 0.9986415; \\ \gamma &= 0.71175\%. \end{aligned}$$

P2 The non-linear model for the wage gap is

$$\ln\frac{\overline{W}}{\underline{W}_t} = \ln\frac{(h_0+\gamma t)(1-x_t\pi_t)}{\int\theta dG - x_t\pi_t(h_0+\gamma t)};$$

which is transformed into

$$
\begin{aligned}
y_t &= \ln(\frac{\overline{W}}{\underline{W}})_t - \ln(1 - x_t\pi_t) = \ln \frac{h_0 + \gamma t}{\int \theta dG - x_t\pi_t(h_0 + \gamma t)} \\
&= \ln \frac{1 + b_0 t}{b_1 - x_t\pi_t(1 + b_0 t)}.
\end{aligned}
$$

I normalize $h_0 = 1$, and jointly estimate γ and $\int \theta dG$.

Source	SS	df	MS	Number of obs	33	
Model	22.56	2	11.28	R-squared	0.9966	
Residual	.07802	31	.0025	AdjR-squared	0.9963	
				RootMSE	.0501668	
Total	22.64	33	.6861	Res.dev.	-105.9117	
yt	Coef.	Std.Err.	t	$P > t$	[95%Conf.Interval]	
$/b0$	.0034	.00056	6.17	0.000	.0023	.0046
$/b1$	.9032	.01031	87.65	0.000	.8822	.92425

Parameter b0 taken as constant term in model & ANOVA table

This implies

$$
\begin{aligned}
\gamma &= 0.3435\%, \\
\int \theta dG &= 0.9032.
\end{aligned}
$$

Chapter 2 Information Production in the Process of Securitization

"The three credit rating agencies were key enablers of the financial meltdown. The mortgage-related securities at the heart of the crisis could not have been marketed and sold without their seal of approval. Investors relied on them, often blindly. In some cases, they were obligated to use them, or regulatory capital standards were hinged on them. This crisis could not have happened without the rating agencies. Their ratings helped the market soar and their downgrades through 2007 and 2008 wreaked havoc across markets and firms."

——— *The Financial Crisis Inquiry Commission Report 2011*

2.1 Introduction

The financial crisis of 2007-2010 is considered by many as the worst financial crisis since the Great Depression. Among the many causes of the crisis, the Financial Crisis Inquiry Commission identifies the failure of the rating agencies as the key ingredient of the financial melt-down. The role of the rating agencies is to evaluate the return of structured financial products under new macroeconomic conditions and to communicate it to the investors. More specifically, in order to rate a mortgage-backed security (MBS), the rating agencies use quantitative models to estimate the loss distributions of the relevant classes of mortgages and simulate the cash flow of the structure to determine the level of credit enhancement needed for a given grade. All this is done under their forecast of macroeconomic conditions. In abstract terms, I can interpret the rating process as a *post*-origination information production process, i.e. producing information about the return of existing loans in banks' portfolio. The quote from the Commission Report points to the importance of understanding the post-origination information production in the process of securitization[12]. In particular, we knew that the loans were getting riskier, but more importantly why were we contented with knowing little about the riskiness of the loans?

This paper develops a simple framework to look at banks' incentive to produce information and its effect on the loan origination decision. In doing so, I reduce rating agencies' problem of information production to banks', assuming that rating agencies de facto produce and use information that banks consider optimal. This assumption is consistent with the issuer-pay model, which aligns rating agencies' and banks' objectives. In practice, the rating agencies rely on the banks to provide the statistics of the loans that comprise the structured financial product, hence there is practically a boundary of knowledge about the loans imposed by the banks. In my model, a bank makes two decisions: loan origination and information production. When making origination decision, the bank must decide loans from which risk class(es) to include in its portfolio. After the loans are made, the bank then chooses how much information about the returns of those existing loans to produce. Perfect disclosure is assumed: whatever information produced, it is public information.

Consider first the problem of information production. A loan is represented by a draw from a distribution of returns. The uncertainty in the riskiness of a mortgage is modeled

[12] For some anecdotal evidence of the inadequicies of rating agencies' practices, such as using outdated 30-year-fixed-interest models to evaluate the subprime mortgages, see *Ohio AG vs. Credit Rating Agencies*, p. 30.

as the uncertainty in the variance parameter of the distribution. A risk-averse bank, as the originator of the loans, can costly produce a signal about the variance parameter. More information makes the financial product to be marketed less risky to investors, increasing the price of the financial product; on the other hand, it makes the price correlate more with the information than with investors' common prior, which leads to increased price volatility. The net benefit of information is the difference between the two aforementioned effects. The optimal amount of information balances the marginal net benefit with the marginal cost of information.

Now if we consider the process of securitization as a means to diversify geographical risks of different mortgages, we would expect that the information production be reduced as compared to a case where diversification is not available. Diversification makes the return of securities less sensitive to the information about a particular loan, decreasing the marginal net benefit of information. When the bank is faced with the opportunity of originating loans of different risk classes, the securitization implies that the bank will originate more and riskier loans, since securitization improves the profitability of issuing a single loan. These statements will be made precise in the models developed below. Throughout the paper, two institutions are compared: one in which loans are sold individually and the one in which pass-through securities backed up by the loans are sold. The first scenario is a benchmark case which disables diversification.

Another way of understanding the issue is to see the process of securitization as a way to redistribute risks. In the context of the model, in any non-trivial cases where information is beneficial at all, the bank is actually better off holding all assets[13]. It is kept from doing so by assumption in the model (i.e. it has to sell all assets, either individually or through securitization) and by required reserves and required capital in reality as argued in Pennacchi (1988)[14]. By producing information, the bank effectively retains risks through the volatility of asset prices. For example, if the bank produces an infinitely precise signal about the return, the investors will pay to the bank exactly what the security is worth. Hence, before the realization of the signal, the bank bears all the risks. This arrangement is closer to the ideal distribution of risks but too costly from the bank's perspective. So it balances the improved distribution of risks through information provision with the cost of doing so. Securitization transforms the problem of risk-distribution such that the welfare is less sensitive to the information production. I will identify three sources of welfare improvements from securitization.

So far in a static world where there exist only idiosyncratic risks, the process of securitization is welfare improving. This may account for the success of this institution before the crisis, when the perception of the environment is essentially static and the major risks of concern are geographical risks and prepayment risks. The inability of making good judgment about the changes in the environment sowed the seeds of the market crash

---

13 See Appendix.

14 Pennacchi (1988) argues that loan sales reduce the cost of capital for the bank when the required reserves and required capital constraints are binding. However loan sales introduce a moral hazard problem: since the return on the loans hinges on bank's monitoring activities, selling loans reduces its commitment to monitoring ex ante. In this paper, the information is not productive by itself. Hence it would be optimal to finance the loans by loan sales if it were feasible. Securitization has long raised concerns from the regulators as a way of regulatory arbitrage (Calomiris and Mason, 2004). By transferring mortgage loans to the Real Estate Mortgage Investment Conduits (REMIC), banks can effectively structure a mortgage-backed securities offering as a sale of assets rather than debt financing, removing the loans from the balance sheet and bypassing the minimum capital requirement.

in 2007. However, even in a static environment with idiosyncratic risks, the securitization can make the bank better off at the cost of investors if the investors are optimistic about the returns on loans. A slightly modified version of the model is used to illustrate the point.

The main message of the paper is that securitization reduces the production of payoff-relevant information and induces the bank to originate more and riskier loans. More specifically, the quality of the loans seems to deteriorate along two dimensions: the estimated or perceived risk, such as those measured by loan-to-value ratios, and the precision of the estimates, such as those measured by the percentage of low/no documentation loans, as is shown in Table 2.1.

Table 2.1: Combined Loan-to-Value Ratio and Share of Low/No Doc Loans

$Year$	2001	2002	2003	2004	2005	2006	2007
$CLTV$ %	79.4	80.1	82.0	83.6	84.9	85.9	82.8
% $Low/No\,Doc$	23.5	29.6	32.2	33.6	36.6	37.7	33.3

Source: Table 1 in Demyanyk and van Hemert (2008).

Some empirical studies argue that securitization is likely to cause weak underwriting standards. Dell'Ariccia, Igan and Laeven (2008) find that larger decline in lending standards as measured by loan denial rates and loan-to-income ratio occurs in areas with higher securitization rates. Mian and Sufi (2008) use zip code level borrowers' data to show a striking correlation between the increase in securitization and the expansion of credit and its dissociation from income growth. In fact, 2002 to 2005 is the only period in the last eighteen years when income and mortgage growth are negatively correlated in their sample. Keys, Mukherjee, Seru and Vig (2010) use a regression discontinuity argument to show that increased securitization had adverse effect on banks' screening incentives.

By exploiting a rule of thumb that loans with FICO score above a threshold of 620 can be more easily securitized, they show that the loans with score just above 620 are more likely to default than those just below the threshold.

On the theoretical ground, there is abundant work on the institution of securitization, which focuses on the comparison between the originate-to-distribute model and the originate-to-hold model (Pennacchi, 1988; Gorton and Pennacchi, 1995; Petersen and Rajan, 2002; Parlour and Plantin, 2008). Although both those models and my model imply that under securitization the bank originates loans of worse qualities, i.e. with lower expected returns or riskier returns or both, the reasons behind it are different. In those models, securitization, or essentially loan sales, introduces a moral hazard problem: selling loans reduces banks' incentive to monitor the loans, which improves the return on those loans. Related, Diamond (1984) identifies banks pooling the loans and offering debt contracts to lenders as an optimal arrangement to mitigate this moral hazard problem. In contrast, this paper looks at the information production *after* the loans are made and explore how that affects the origination decision. Securitization, by diversifying and transforming the sensitivity of bank's payoff to information, alters its solution to the information production problem. As a result, improved risk-sharing and reduced information production makes the origination of riskier loans more affordable to the bank. As suggested by the quote at the beginning of the paper, the main problem is not that the loans are perceived as riskier, but that we are contented with imprecise perceptions. My model predicts that under securitization not only the bank lends to riskier borrowers, i.e.

the perceived risk is increased, it also spends less resource improving its perception of the risk. This "carefree" mindset of the bank and the rating agencies opens the possibility of vastly erroneous predictions from under-invested models in an dynamic environment whose key parameters keep evolving.

In terms of the welfare implications, the paper differs from the aforementioned papers too. In those papers, since securitization or loan sales reduces productive monitoring, it reduces welfare. Here, in a stationary world with idiosyncratic risks and homogeneous priors, securitization is shown to be welfare-improving. This feature of securitization justifies its existence and popularity in the 1980s and 1990s, but also sows the seeds of slow adjustment and learning in a dynamic environment, which leads to the current crisis. This is an example where information, which has a purely re-distributive role, is Pareto-improving when the initial allocation is sub-optimal. For more discussion on the social and private values of information, please refer to Hirshleiffer (1971) and Hakansson, Kunkel and Ohlson (1982) among others.

Last but by no means the least, there is a long line of research on the information production by heterogeneous investors in a competitive market with noisy rational expectations: Grossman (1976), Grossman and Stiglitz (1980), Hellwig (1980), Verrecchia (1982) and Diamond and Verrecchia (1991) are some important contributions to the literature. However, I chose a set-up where the bank determines the information production. It is motivated by the fact that in the market of mortgage-backed securities there is practically little scope for the investors to acquire information. It is common practice among the government sponsored agencies such as Fannie Mae and Freddie Mac to limit the amount of information available to the market. Information at the pool level is disclosed, but specific information that helps identify the particular loans in the pool is withheld. Diamond and Verrecchia (1991), Fishman and Hagerty, (1990) and Glaeser and Kallal (1997) provide theoretical insights into this phenomenon in models with asymmetric information. To the extent that banks *honestly* disclose coarser information, I abstract away the informational asymmetry between the bank and the investors.

The paper is organized as follows. The second section contains a basic model of information production. The third section solves the origination and information production decisions jointly in a model with heterogeneous lending opportunities. Section 4 presents a model with heterogenous priors and discusses some welfare implications. The conclusion follows.

2.2 The Basic Model

This is a static model of information production. There are a single bank and a large number of homogeneous investors. The bank has the constant absolute risk aversion (CARA) utility with risk tolerance $\rho : U(w) = -\exp(-\frac{w}{\rho})$. Investors' preferences are also of the CARA type with risk tolerance $r : u(w) = -\exp(-\frac{w}{r})$. The bank is endowed with n loans. The returns on the loans are drawn independently from a common normal distribution parameterized by (μ, σ^2). Ex ante, the variance σ^2 is unknown and is (correctly) believed to follow a normal distribution: $\widetilde{\sigma}^2 \sim \mathrm{N}(\mu_\sigma, \sigma_\sigma^2)$. Suppose $\mu_\sigma \gg \sigma_\sigma$.[15]

[15] For $\frac{\mu_\sigma}{\sigma_\sigma} \geq 4$, $\Pr(\widetilde{\sigma}^2 \leq 0) \leq 3.2e-005$. In the numerical example in the next section, I will show that under

The normality is assumed for ease of interpretation. I show in the Appendix that the qualitative results also hold in an environment where the bank is risk neutral and the variance of the return is distributed non-central $\chi^2(1)$. The bank has a technology to produce a public signal $\widetilde{s}^2$ about the unknown variance $\widetilde{\sigma}^2$:

$$\widetilde{s}^2 = \widetilde{\sigma}^2 + \widetilde{\epsilon},$$

where the noise $\widetilde{\epsilon}$ is distributed as $\mathrm{N}(0, \frac{1}{\pi})$ independently from $\widetilde{\sigma}^2$. The cost of signals is an increasing and convex function of the precision π: $c'(\pi) \geq 0$, $c"(\pi) \geq 0.$

Consider two problems. In Problem 1, the bank sells the loans individually to the investors. In Problem 2, it pools the loans and issues N shares of securities to N investors, each share being a claim of $\frac{1}{N}$ of the total return on the pool. Assume $N \geq n.$

The sequence of the play is as follows. Given the loans, the bank chooses the precision of the signal π. After the signal is realized and observed by all, the bank meets with an investor and offers a loan if in Problem 1 and offers a share of security if in Problem 2. The bank has full bargaining power and the investor is charged his willingness to pay, which is a function of the realization and the precision of the signal. This type of meetings continue until the bank sells off all of the loans.

I proceed by backward induction. Suppose the bank chooses a signal with precision π. I can write the joint distribution of the variance $\widetilde{\sigma}^2$ and signal $\widetilde{s}^2$ as

$$\begin{bmatrix} \widetilde{\sigma}^2 \\ \widetilde{s}^2 \end{bmatrix} \sim \mathrm{N}(\begin{bmatrix} \sigma_0^2 \\ \sigma_0^2 \end{bmatrix}, \begin{bmatrix} \sigma_{\sigma_0}^2 & \sigma_{\sigma_0}^2 \\ \sigma_{\sigma_0}^2 & \sigma_{\sigma_0}^2 + \frac{1}{\pi} \end{bmatrix}.$$

The conditional distribution of $\widetilde{\sigma}^2$ given s^2 is still normal, with the following posterior

the chosen parametrization, the normal distribution is a good approximation of the truncated normal.

mean $\mu_{\sigma 1}$ and variance $\sigma^2_{\sigma 1}$

$$\begin{aligned} \mu_{\sigma 1}(s^2) &= \frac{1}{\pi\sigma^2_\sigma + 1}\mu_\sigma + \frac{\pi\sigma^2_\sigma}{\pi\sigma^2_\sigma + 1}s^2; \\ \sigma^2_{\sigma 1} &= \frac{\sigma^2_\sigma}{\pi\sigma^2_\sigma + 1}. \end{aligned}$$

Problem 1

Consider the case where the bank sells the loans individually. Investors, conditioning on s^2, infer that the variance $\widetilde{\sigma}^2$ has the distribution

$$f_{\widetilde{\sigma}^2|s^2}(\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma_{\sigma_1}} \exp\{-\frac{[\sigma^2 - \mu_{\sigma 1}(s^2)]^2}{2\sigma^2_{\sigma_1}}\}.$$

Furthermore, conditional on σ^2, the return on a loan, $\widetilde{u}_l$, is distributed according to

$$f_{\widetilde{u}_l|\sigma^2}(u) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(u-\mu)^2}{2\sigma^2}\}.$$

Hence, the conditional distribution of $\widetilde{u}_l$ given s^2 is

$$\begin{aligned} f_{\widetilde{u}_l|s^2}(u) &= \int f_{\widetilde{u}_l,\widetilde{\sigma}^2|s^2}(u,\sigma^2)d\sigma^2 \\ &= \int f_{\widetilde{u}_l|\sigma^2}(u) f_{\widetilde{\sigma}^2|s^2}(\sigma^2)d\sigma^2, \end{aligned}$$

which is not normal. However investors' willingness to pay $p_l(s^2)$ is again normal. With some standard algebra[16],

$$\exp(-\frac{p_l(s^2)}{r}) = E[\exp(-\frac{\widetilde{u}_l}{r})|\widetilde{s}^2 = s^2]$$
$$= \exp\{-[\frac{\mu}{r} - \frac{\mu_{\sigma 1}(s^2)}{2r^2} - \frac{\sigma^2_{\sigma_1}}{8r^4}]\}.$$
$$\Rightarrow p_l(s^2) = \mu_0 - \underbrace{[\frac{\mu_{\sigma 1}(s^2)}{2r} + \frac{\sigma^2_{\sigma_1}}{8r^3}]}_{\text{an investor's risk premium}}$$
$$= \mu_0 - \frac{1}{2r}(\frac{1}{\pi\sigma^2_\sigma + 1}\mu_\sigma + \frac{\pi\sigma^2_\sigma}{\pi\sigma^2_\sigma + 1}s^2) - \frac{\sigma^2_{\sigma_1}}{8r^3}.$$

As is indicated above, the investor's risk premium reflects his perceived risk, $\mu_{\sigma 1}(s^2)$, and his uncertainty of the perception, $\sigma^2_{\sigma_1}$. The perceived risk, or the posterior mean of the variance parameter, is a weighted average of the prior and signal. When the signal is perfectly informative, i.e. $\pi \to \infty$, investors rely only on the signal, and vice versa.

Now I solve bank's problem of information provision. Note that to the bank the volatility of the profit comes solely from $\mu_{\sigma 1}(\widetilde{s}^2)$.

$$\begin{aligned} & \max_\pi EU(n\widetilde{p}_l - c(\pi)) \\ = \; & \max_\pi n\mu - \underbrace{[\frac{n}{2r}\mu_\sigma + \frac{n}{8r^3}\sigma^2_{\sigma_1}]}_{\text{investors' risk premia}} - \underbrace{\frac{n^2}{8\rho r^2}var[\mu_{\sigma 1}(s^2)]}_{\text{the bank's risk premium}} - c(\pi) \qquad (2.1) \\ = \; & \max_\pi n\mu - [\frac{n}{2r}\mu_\sigma + \underbrace{\frac{n}{8r^3}\frac{\sigma^2_\sigma}{\pi\sigma^2_\sigma + 1}}_{\text{decreasing in precision}}] - \underbrace{\frac{n^2}{8\rho r^2}\frac{\pi\sigma^4_\sigma}{\pi\sigma^2_\sigma + 1}}_{\text{increasing in precision}} - c(\pi). \end{aligned}$$

In changing π, the bank is trading of investors' risk premia with its own risk premium:

[16] The willingness to pay can be derived by applying the mean-variance argument twice. First, the willingness to pay given σ^2 is $p' = \mu - \frac{\sigma^2}{2r}$. Second, since $\widetilde{p}' = \mu - \frac{\widetilde{\sigma}^2}{2r}$ is itself normal, hence $p_l(s^2) = E[\widetilde{p}'|\widetilde{s}^2 = s^2] - \frac{var[\widetilde{p}'|\widetilde{s}^2 = s]}{2r} = \mu - \frac{\mu_{\sigma 1}(s)}{2r} - \frac{\sigma^2_{\sigma_1}}{8r^3}$.

a higher precision reduces the risks that the investors face, decreasing their risk premia, while it induces more volatile asset prices, increasing the bank's risk premium. When $\rho > nr,$ the marginal effect of π on investors' risk premia is bigger than that on the bank's. Hence, information production helps shift risks from the investors' side to the bank's.

Assumption 1 $\rho > nr.$

Under Assumption 2.1, I can define the notions of first- and second- best allocation of risks in this model. The first-best allocation would be for the bank to hold all assets, yielding $n\mu - \frac{n\mu_\sigma}{2\rho} - \frac{n^2\sigma_\sigma^2}{8\rho^3}$. This is infeasible if the bank faces binding required reserves and capital, as discussed in the introduction. The second-best would be for the bank to produce a perfectly precise signal to retain the risks, but it is too costly. Hence the optimal information balances the redistribution of risks and the cost of doing so.

$$\begin{aligned} & \max_{\pi} EU(n\widetilde{p}_l - c(\pi)) \\ = \ & \max_{\pi} \underbrace{n\mu - \frac{n}{2r}\mu_\sigma - \frac{n^2}{8\rho r^2}\sigma_\sigma^2}_{\text{payoff from the 2nd best}} - \underbrace{(\frac{n}{8r^3} - \frac{n^2}{8\rho r^2})\frac{\sigma_\sigma^2}{\pi\sigma_\sigma^2+1}}_{\text{shortfall from the 2nd best}} - c(\pi). \end{aligned}$$

The optimal information production $\overline{\pi}$ is determined by the first order condition (FOC), which is both sufficient and necessary,

$$\frac{n(\rho - nr)}{8\rho r^3}\sigma_\sigma^4 = (\overline{\pi}\sigma_\sigma^2 + 1)^2 c'(\overline{\pi}). \tag{2.2}$$

Problem 2

Now consider the case where the bank sells N shares of pass-through securities. For a given σ^2, the return on a share of the security is distributed normal, $\widetilde{u}_s \sim \mathrm{N}(\frac{n}{N}\mu, \frac{n\sigma^2}{N^2})$. The willingness to pay upon seeing s^2 can be derived similarly,

$$p_s(s^2) = \frac{n}{N}\mu - \frac{n\mu_{\sigma 1}(s^2)}{2rN^2} - \frac{n^2\sigma^2_{\sigma_1}}{8r^3N^4}.$$

Increasing N amounts to increasing the investor base. As N gets bigger, each investor is closer to being risk neutral. The bank's problem now becomes

$$\begin{aligned} & \max_{\pi} EU(N\widetilde{p}_s - c(\pi)) \\ = \quad & \max_{\pi} n\mu - \frac{n}{2r}\frac{\mu_\sigma}{N} - \frac{n}{8r^3}\frac{\sigma^2_{\sigma_1}}{N^3/n} - \frac{n^2}{8\rho r^2}\frac{var(\mu_{\sigma 1}(s^2))}{N^2} - c(\pi). \end{aligned} \tag{2.3}$$

Compare (2.1) and (2.3) and it is clear that the bank's utility is now less sensitive to the precision π. The following proposition states that the bank chooses less information under securitization.

Proposition 1 *Let $\overline{\pi}$ be the optimal precision the bank chooses when it sells loans individually and $\widehat{\pi}$ be the optimal precision when it securitizes. Under Assumption 2.1,*

(1) when $\rho > Nr, \overline{\pi} > \widehat{\pi} > 0$.

(2) when $nr < \rho \leq Nr, \overline{\pi} > \widehat{\pi} = 0$.

Proof. *When $\rho > Nr$,both problems have interior solutions. The FOC in Problem 2 is*

$$\frac{n^2(\rho - Nr)}{8\rho r^3 N^3}\sigma^4_\sigma = (\widehat{\pi}\sigma^2_\sigma + 1)^2 c'(\widehat{\pi}). \tag{2.4}$$

Note that the common RHS of (2.2) and (2.4),

$$\frac{d}{d\pi}(\pi\sigma^2_\sigma + 1)^2 c'(\pi) = 2\sigma^2_\sigma(\pi\sigma^2_\sigma + 1)c'(\pi) + (\pi\sigma^2_\sigma + 1)^2 c''(\pi) > 0.$$

The LHS in Problem 1 is unambiguously greater than that in Problem 2. Therefore, $\overline{\pi} > \widehat{\pi} > 0$. (2) is immediate. ■

It is obvious that increasing N, i.e. expanding the investor base, reduces bank's incentive to produce information. This is reminiscent of Peress (2010), who showed that the bigger a given stock's investor base, the less investors engage in information production.

There are three potential sources of welfare gains from the securitization. In the non-trivial case where $\rho > Nr$, write the two objective functions in Problems 1 and 2 as follows.

$$\begin{aligned}
\max_\pi EU(n\widetilde{p}_l - c(\pi)) &= n\mu - \frac{n}{2r}\mu_\sigma - \frac{n^2}{8\rho r^2}\sigma_\sigma^2 - (\overline{\pi} + \frac{1}{\sigma_\sigma^2})c'(\overline{\pi}) - c(\overline{\pi}). \\
\max_\pi EU(N\widetilde{p}_s - c(\pi)) &= \underbrace{n\mu - \frac{n}{2r}\frac{\mu_\sigma}{N} - \frac{n^2}{8\rho r^2}\frac{\sigma_\sigma^2}{N^2}}_{\text{risk-sharing effect}} \underbrace{-(\widehat{\pi} + \frac{1}{\sigma_\sigma^2})c'(\widehat{\pi})}_{\text{information sensitivity effect}} - \underbrace{c(\widehat{\pi})}_{\text{cost-saving effect}}.
\end{aligned}$$

Firstly, the second-best payoff in Problem 2 is bigger due to a risk-sharing effect. On one hand, by pooling and subdividing the risks from the loans, the securitization reduces the total amount of risks to be distributed (when $N = n$). On the other, a bigger investor base enables the risks to be distributed more widely. The second source of improvements comes from the fact that the securitization reduces the sensitivity of bank's payoff to information. Interestingly, although in the action space $\widehat{\pi}$ is farther away from the 2nd-best action, i.e. a precision of infinity, than $\overline{\pi}$ is, in the payoff space bank's utility in Problem 2 is closer to the 2nd-best payoff than that in Problem 1. The securitization achieves a utility level that is closer to the 2nd-best with a lower level of information production. The third source of gain is in the saving in the cost of information.

In the next section, I show how the intuition in this basic model is brought into play in a model with heterogeneous lending opportunities and endogenous origination.

2.3 A Model with Heterogeneous Lending Opportunities

In this section, I endogenize the bank's origination decision in light of its optimal information production. Without loss of generality, assume from now on $N = n$.

Imagine there are $\overline{N}$ types of projects in need of funding. The return on each type of projects are represented by a normal distribution with a common and known mean μ[17] and an unknown variance $\widetilde{\sigma}_i^2$, $i = 1, 2, ..., \overline{N}$. The variances are from normal distributions ordered in the following way:

$$\begin{aligned} \mu_{\sigma 1} &< \mu_{\sigma 2} < ... < \mu_{\sigma n} < ....; \\ \sigma^2_{\sigma_1} &< \sigma^2_{\sigma_2} < ... < \sigma^2_{\sigma_n} < ..... \end{aligned} \tag{2.5}$$

Again, assume $\mu_{\sigma i} \gg \sigma^2_{\sigma_i}, \forall i = 1, ...\overline{N}$. The lower-indexed types of projects not only have lower perceived risks, but also the beliefs are more precise. Think of smaller $\mu_{\sigma i}$ representing lower loan-to-value ratio and smaller $\sigma^2_{\sigma_i}$ representing full documentation loans. There is a fixed supply of projects whose riskiness follow each distribution. I normalized it to 1. The bank clearly prefers originating lower indexed loans to higher indexed ones.

The bank can produce signals about each of the random parameters:

$$\widetilde{s}_i^2 = \widetilde{\sigma}_i^2 + \widetilde{\varepsilon}_i, \forall i = 1, ..., \overline{N}.$$

where $\widetilde{\varepsilon}_i' s$ are noises distributed independently across i according to $\mathrm{N}(0, \frac{1}{\pi_i})$. The cost of information $c(\pi_i)$ has the usual properties.

[17] The following analysis goes through in an alternative environment where the mean of the returns increases sufficiently slowly in the index.

The bank's problem now is to decide at which index to stop, taking into consideration the optimal post-origination information production. Again I compare its solution in Problem 1 where it later sells the loans individually and that in Problem 2 where it later securitizes the loans.

Assumption 2 $1 < \overline{N} \leq \frac{3}{4}[\frac{\rho}{r}] - 1$.
Assumption 3 Linear costs: $c_i(\pi_i) = c \cdot \pi_i$, $\forall i = 1, ..., \overline{N}$.
Assumption 4 $\sigma_{\sigma n}^2 > \sqrt{\frac{8\rho r^3 n^3 c}{\rho - rn}}$, $\forall n = 1, 2, ..., \overline{N}$.

The linear cost structure allows me to solve the optimal information production explicitly. Assumption 2.5 rules out the trivial case where there is no information production under the securitization.

Problem 1

The bank's problem can be summarized as

$$\begin{aligned} & \max_n \sum_{i=1}^{n} \Psi_i(\overline{\pi}_i) \\ = \quad & \max_n \sum_{i=1}^{n} [\mu - \frac{1}{2r}\mu_{\sigma i} - \frac{1}{8r^3}\frac{\sigma_{\sigma i}^2}{\overline{\pi}_i \sigma_{\sigma i}^2 + 1} - \frac{1}{8\rho r^2}\frac{\overline{\pi}_i \sigma_{\sigma i}^4}{\overline{\pi}_i \sigma_{\sigma i}^2 + 1} - c\overline{\pi}_i], \end{aligned}$$

where

$$\overline{\pi}_i = \max\{0, \sqrt{\frac{\rho - r}{8\rho r^3 c}} - \frac{1}{\sigma_{\sigma i}^2}\}.$$

Evidently, the optimal information provision is independent of the size of the pool, n. One can prove the following properties.

Property 1.1 $\overline{\pi}_i$ is independent of n and increasing in i.
Property 1.2 $\Psi_i(\overline{\pi}_i)$ is decreasing in i.

Property 1.1 is immediate. To derive Property 1.2, note that $\frac{\sigma_{\sigma i}^2}{\overline{\pi}_i \sigma_{\sigma i}^2 + 1} = \sqrt{\frac{8\rho r^3 c}{\rho - r}}$ and all other terms in the objective function is decreasing in $\overline{\pi}_i$. More generally, these two properties hold for all cost functions that are increasing and concave. The benefit from an additional loan decreases, as the bank makes riskier loans. The optimal size of the pool is therefore

$$n^* = \min\{\overline{N}, n|\Psi_n(\overline{\pi}_n) \geq 0 \text{ and } \Psi_{n+1}(\overline{\pi}_{n+1}) < 0\}.$$

To better illustrate the intuition, I parametrize the model[18] and plot the $\overline{\pi}_i$ and $\Psi_i(\overline{\pi}_i)$ as functions of i in Figures 2.1 and 2.2. The optimal size of the pool is 11 in this numerical example.

Problem 2

The bank's problem can be represented as maximizing the sum of bank's utility contributed by each loan.

$$\max_n \sum_{i=1}^{n} \Phi_i^n(\widehat{\pi}_i^n) = \max_n \sum_{i=1}^{n} [\mu_0 - \frac{1}{2rn}\mu_{\sigma i} - \frac{1}{8r^3n^3}\frac{\sigma_{\sigma i}^2}{\widehat{\pi}_i^n\sigma_{\sigma i}^2+1} - \frac{1}{8\rho r^2n^2}\frac{\widehat{\pi}_i^n\sigma_{\sigma i}^4}{\widehat{\pi}_i^n\sigma_{\sigma i}^2+1} - c\widehat{\pi}_i^n].$$

where

$$\widehat{\pi}_i^n = \max\{0, \sqrt{\frac{\rho - rn}{8\rho r^3n^3c}} - \frac{1}{\sigma_{\sigma i}^2}\}.$$

Clearly, the optimal information production of the ith loan not only depends on $\sigma_{\sigma i}^2$, but also on the size of the pool, n. This is because now the return on each share of security depends on the total number of the loans in the pool.

Property 2.1 $\widehat{\pi}_i^n$ depends on n. For a given n, $\widehat{\pi}_i^n$ is increasing in i.

Property 2.2 For a given i, $\widehat{\pi}_i^n$ is decreasing in n.

Property 2.3 For any $n > 1$, $\widehat{\pi}_i^n < \overline{\pi}_i$, $\forall i \leq n$.

[18] $\rho = 28$, $r = 1$, $\mu = 45$, $\overline{N} = 20$, $c = 0.01$. $\sigma_{\sigma n}^2 = 1.1\sqrt{\frac{8\rho^3n^3c}{\rho - rn}}$ and $\mu_{\sigma n} = 4\sigma_{\sigma n}$, $\forall n = 1, ..., \overline{N}$. In the Appendix, I show the normal distribution is a good approximation of the truncated normal distribution of the variance parameter in this environment.

Property 2.2 reflects the fact that a larger number of loans decreases the sensitivity of bank's utility to information, reducing the information production of each existing loan. Under the same parametrization as in Problem 1, I plot the information production $\widehat{\pi}_i^n$ as a function of i for various numbers of n in Figure 2.3. As n increase, the curve is defined on a bigger domain with the entire curve shifting down.

Property 2.4 For a given n, $\Phi_i^n(\widehat{\pi}_i^n)$ is decreasing in i, $\forall i \leq n$.

Property 2.5 Under Assumptions 2.5-2.5, $\Phi_i^n(\widehat{\pi}_i^n) < \Phi_i^{n+1}(\widehat{\pi}_i^{n+1})$, $\forall i \leq n$ and $\forall n \leq \overline{N}$.

Property 2.4 holds for pretty much the same reason as Property 1.2. For a fixed size of the pool, the utility contribution from a riskier and more uncertain[19] loan is smaller, for two reasons. One, the price of the securities discounts higher perceived risks (bigger μ_σ). Two, the bank needs to produce more information about a more uncertain loan (bigger σ_σ^2), hence introducing more volatility in asset prices and a higher cost.

When the bank increase n, it of course loads additional risks on its book, but a bigger pool implies less information production for all the existing loans. If the saving in information production and reduction in the price volatility is large, it can make projects that the bank wouldn't find profitable in Problem 1 profitable here. Consider an increase of the number of the loans from n to $n+1$. A sufficient condition for $\Phi_i^n(\widehat{\pi}_i^n) < \Phi_i^{n+1}(\widehat{\pi}_i^{n+1})$ to hold is $\frac{1}{8r^3n^3}\frac{\sigma_{\sigma i}^2}{\widehat{\pi}_i^n\sigma_{\sigma i}^2+1} > \frac{1}{8r^3(n+1)^3}\frac{\sigma_{\sigma i}^2}{\widehat{\pi}_i^{n+1}\sigma_{\sigma i}^2+1}$, which is equivalent to $(\rho - rn)n^3 < [\rho - r(n+1)](n+1)^3$. Assumption 2.5 is the sufficient and necessary condition for the latter to hold for all $n \leq \overline{N}$. An implication of Property 2.5 is that the optimal choice in Problem 2, denoted by n^{**}, has to be greater than n whenever

[19] Let me abuse the usage of "uncertain" to mean a larger σ_σ^2.

$\Phi_n^n(\widehat{\pi}_n^n) > 0$. Now I am ready to establish the result that the bank issues credit to more and by construction riskier and more uncertain borrowers, under securitization.

Proposition 2 *Under Assumptions 2.5-2.5,* $n^{**} \geq n^*$.
Proof. *It is sufficient to show that* $\Phi_{n^*}^{n^*}(\widehat{\pi}_{n^*}^{n^*}) > 0$. *Rewrite*

$$\begin{aligned}
\Psi_{n^*}(\overline{\pi}_{n^*}) &= \mu_0 - \frac{1}{2r}\mu_{\sigma n^*} - \frac{1}{8\rho r^2}\sigma^2_{\sigma n^*} - c(\overline{\pi}_{n^*} + \frac{1}{\sigma^2_{\sigma n^*}}) - c\overline{\pi}_{n^*}. \\
\Phi_{n^*}^{n^*}(\widehat{\pi}_{n^*}^{n^*}) &= \mu_0 - \frac{1}{2rn^*}\mu_{\sigma n^*} - \frac{1}{8\rho r^2 n^{*2}}\sigma^2_{\sigma n^*} - c(\widehat{\pi}_{n^*}^{n^*} + \frac{1}{\sigma^2_{\sigma n^*}}) - c\widehat{\pi}_{n^*}^{n^*} \\
&> \Psi_{n^*}(\overline{\pi}_{n^*}) \geq 0.
\end{aligned}$$

■

In the numerical example, the optimal number of loans in Problem 2 is the largest possible, 20. Figure 2.4 illustrates how $\Phi_i^n(\widehat{\pi}_i^n)$ as a function of i shifts up as n increases.

Securitization is unambiguously welfare-improving in this context. Since $\Psi_i(\overline{\pi}_i) < \Phi_i^n(\widehat{\pi}_i^n)$, $\forall i \leq n$ and $\forall n$, the bank already enjoys a higher utility level under securitization than under piecemeal loan sales when the pool size is n^*. Enlarging the pool size to n^{**} improves bank's utility even further. On the other hand, all investors have zero surplus always.

Proposition 3 *Securitization is welfare-improving.*

In this section, I show that when the bank decides the number of loans and information production jointly, it includes more loans (hence riskier and more uncertain loans) into its portfolio under securitization than otherwise. The per-loan information production is lower under securitization than otherwise. In the numerical example, the total information production is also lower under securitization. The securitization, by changing the sensitivity of payoffs to information, dissuades the use of information to shift risks. It is shown to be welfare-improving.

2.4 A Model with Optimistic Investors

Adopt the set-up of the basic model and retain the assumption that $N = n$. The old notation is used to model bank's beliefs. The investors have beliefs

$$\begin{bmatrix} \widetilde{\sigma}^2 \\ \widetilde{s}^2 \end{bmatrix} \sim \mathrm{N}(\begin{bmatrix} \overline{\mu}_\sigma \\ \overline{\mu}_\sigma \end{bmatrix}, \begin{bmatrix} \sigma_\sigma^2 & \sigma_\sigma^2 \\ \sigma_\sigma^2 & \sigma_\sigma^2 + \frac{1}{\pi} \end{bmatrix},$$

and $\overline{\mu}_\sigma < \mu_\sigma$. Call these investors *optimists*. Rajan, Seru and Vig (2010) suggest that as the level of securitization increases, lenders tend to originate loans that rate high based on characteristics known to the investors and ignore other credit-relevant information. This may give rise to an optimistic opinion about the returns among the investors. Accordingly, I model the bank as having an objective belief μ_σ and the investors as being more optimistic. Obviously if $\overline{\mu}_\sigma$ were very small, the bank would never bother to produce information to correct the investors' beliefs. Here, I will restrict my attention to a case where the investors are mildly optimistic so that the bank produces a positive amount of information in the benchmark case while producing no information under securitization.

Assumption 5 $\mu_\sigma - \frac{\rho - nr}{4\rho r^2}\sigma_\sigma^2 < \overline{\mu}_\sigma \leq \mu_\sigma - \frac{\rho - nr}{4\rho r^2 n}\sigma_\sigma^2.$

The same logic as in the Basic Model applies and I solve for optimal information production in both problems and derive conditions under which the bank is better off at the cost of the investors under securitization.

Proposition 4 *Under Assumptions 2.1 and* 2.4, *the sufficient and necessary condition,* *under which investors are worse off when the bank securitizes loans than when it sells loans individually, is*

$$c'(\frac{n-1}{\sigma_\sigma^2}) < \frac{\sigma_\sigma^2}{2rn}(\frac{\rho - nr}{4\rho r^3}\sigma_\sigma^4 - \mu_\sigma + \overline{\mu}_\sigma).$$

Proof. *It's easy to verify that in Problem 1, the bank chooses $\overline{\pi}$ that solves*

$$n(\frac{\rho - nr}{4r^2\rho}\sigma_\sigma^2 - \mu_\sigma + \overline{\mu}_\sigma) = \frac{2r}{\sigma_\sigma^2}(\overline{\pi}\sigma_\sigma^2 + 1)^2 c'(\overline{\pi}), \tag{2.6}$$

while in Problem 2, $\widehat{\pi} = 0$. In Problem 1, the utility of an investor who is endowed with w_0 wealth is

$$\begin{aligned} Eu(w_0 - \widetilde{p} + \widetilde{u}) &= -\exp(-\frac{w_0}{r})E\exp(\frac{\widetilde{p} - \widetilde{u}}{r}) \\ &= -\exp(-\frac{w_0}{r})E_{\widetilde{s}^2}\{\exp(\frac{\widetilde{p}}{r}E[\exp(-\frac{\widetilde{u}}{r})|s^2]\} \\ &= -\exp[-\frac{1}{r}(w_0 - \frac{\mu_\sigma - \overline{\mu}_\sigma}{2r(\overline{\pi}\sigma_\sigma^2 + 1)})] < Eu(w_0). \end{aligned}$$

So investors incur loss from the trade (unknowingly at the time of the trade for them). Repeat the same exercise for Problem 2, $Eu(w_0 - \widetilde{p} + \widetilde{u}) = -\exp[-\frac{1}{r}(w_0 - \frac{\mu_\sigma - \overline{\mu}_\sigma}{2nr})] < Eu(w_0)$. Hence, the necessary and sufficient condition for investors to have a lower utility level in Problem 2 is

$$\begin{aligned} n &< \overline{\pi}\sigma_{\sigma_0}^2 + 1, \text{ or} \\ \overline{\pi} &> \frac{n-1}{\sigma_\sigma^2}. \end{aligned}$$

Since the RHS of (2.6) is increasing in π. The above condition is equivalent to the inequality in the statement of the proposition. ■

In general, there are many ways to satisfy the above condition. One trivial way is to make the marginal cost of the signal sufficiently low. But I will discuss two other economically more meaningful scenarios. In order to make comparisons, let me also fix the parameters throughout these scenarios. In particular,

$$\rho = 10;\ n = 20;\ \sigma_0^2 = 9;\ \overline{\sigma}_0^2 = 8.5;$$

$$\text{and } c(\pi) = \pi^2/20.$$

Numerical Examples

Scenario 1: Very risk averse investors: $r = 0.25$; $\sigma_\sigma^2 = 5$.

Very risk averse investors are sensitive to information. A marginal increase in the precision of the information increases their willingness to pay much more than it increases the price volatility. This induces the bank to produce very precise signals in Problem 1, $\overline{\pi} = 8.99$. Here the utility of a single investor in Problem 1 is -1.0909, while his utility under securitization is -1.2214.

Scenario 2: Very imprecise prior: $\sigma_\sigma^2 = 100$; $r = 0.48$.

When investors and the bank have very rough idea about the variance of the distribution of the return ex ante, for a given level of investors' risk tolerance, the marginal benefit of information is high. Here $\overline{\pi} = 1.9934$. A single investor's utility when the bank sells loans is -1.0054, while his utility when the bank securitizes is -1.0558.

The parameters are chosen such that if we swap the value of r in one scenario with that in the other, the result that investors are worse-off under securitization goes away. Basically when the prior belief about the variance is very imprecise, in order to dissuade the bank from producing information under securitization, the investors cannot be overly risk averse. Similar intuition goes through when investors are very risk averse.

Here I have outlined two cases in which securitization with zero information production does *not* lead to Pareto improvement. In either case, the bank is better at the cost of investors. In one scenario, investors are much more risk averse than the bank. In the other, agents have very rough idea about the riskiness of the returns ex ante.

2.5 Conclusion

In this paper, I propose a model of information production of banks, where information is a costly tool to redistribute risks. The securitization, on one hand reduces the total amount of risks to be distributed by providing diversification; on the other reduces the costly information production by decreasing the sensitivity of bank's payoff to information. As a result, given a pool of loans, in a stationary environment with idiosyncratic risks, the bank achieves a higher utility level under securitization with less information production. Furthermore, when the bank decides how many loans to make taking into consideration the optimal information production, it makes more and riskier loans, whose riskiness is also more uncertain. Essentially, securitization increases the profitability of loans of all risk classes, so that banks can afford extending credit to more riskier borrowers, which would have generated negative profit if the bank were to sell the loan individually.

The findings have bearing on policy issues hotly debated in the aftermath of the crisis. The low level of investment in statistical modeling and analysis is bred and rationalized by the relatively stationary environment in the 80s and 90s, when the mortgage market consisted of mostly prime mortgages and the risks are mostly at the geographical or individual level. However, this rational inattention is no longer adequate or justifiable in 2000s, when the mortgage market exhibits an increasing sign of betting on house price appreciation, which introduces an increasingly big component of systemic risks. The rating agencies fail to understand that the changes in the nature of the mortgage contracts and the macroeconomic environment render the conditions under which the securitization is socially desirable obsolete.

A future direction of work would be to incorporate a learning mechanism into the model. If I interpret the level of information production as an effort level of learning about the parameter of some trend in a dynamic environment, then the securitization slows down the learning process by economizing on information production. This essentially translates into a more volatile sequence of future returns and opens the possibility of erroneous model predictions of the environment. The desirability of the institution of securitization would then depend on weighing the benefits from diversification and reduced information costs against the slower learning process.

Figure 2.1: $\overline{\pi}_i$ as a function of i in Problem 1

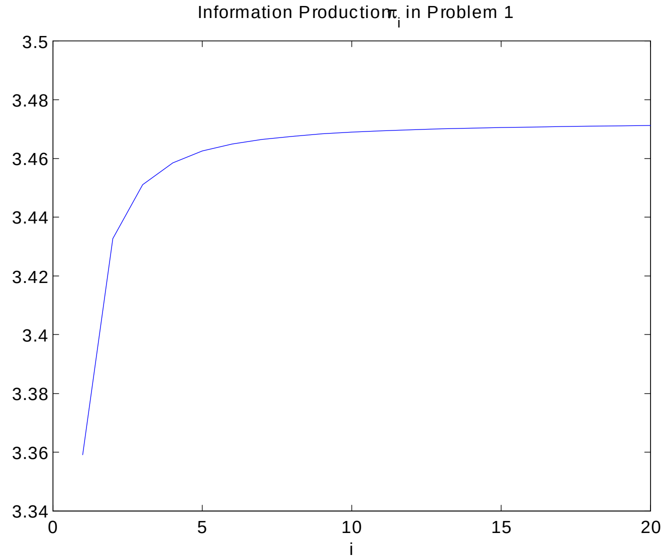


Figure 2.2: $\Psi_i(\overline{\pi}_i)$ as a function of i in Problem 1

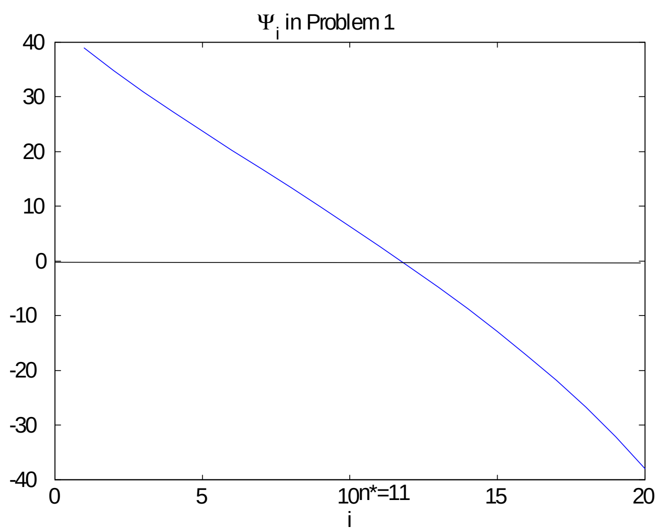


Figure 2.3: $\widehat{\pi}_i^n$ as a function of i for various n in Problem 2

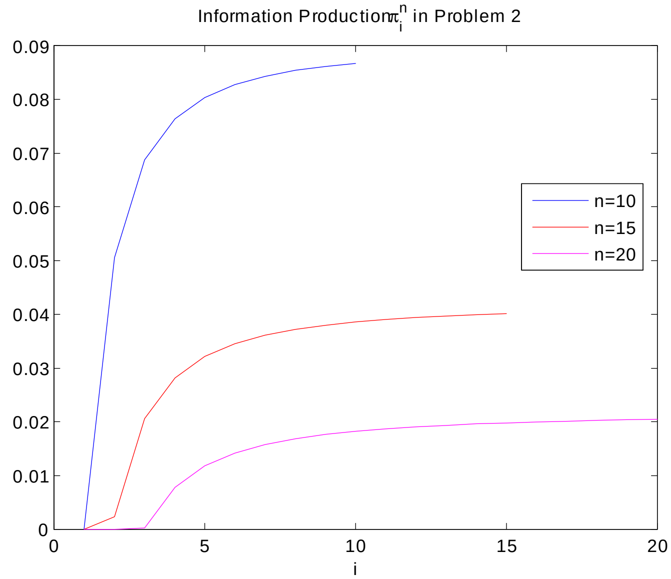
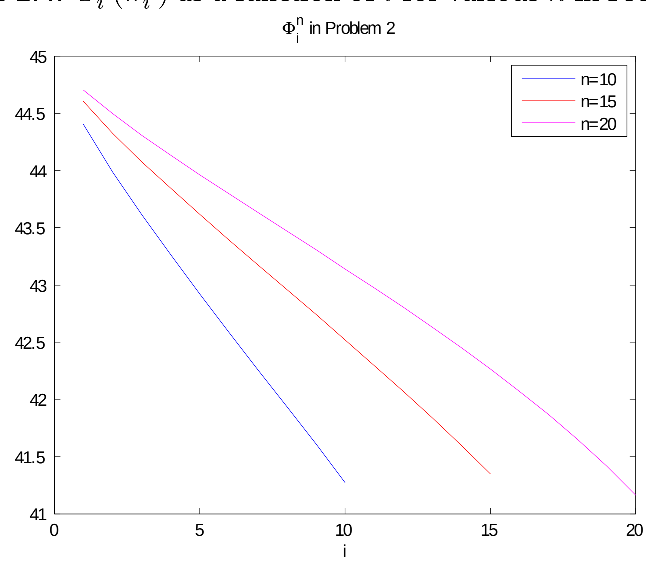


Figure 2.4: $\Phi_i^n(\widehat{\pi}_i^n)$ as a function of i for various n in Problem 2



Ashcraft, A. and T., Schuermann, 2008. "Understanding the securitization of subprime mortgage credit," Staff Reports 318, Federal Reserve Bank of New York.

Calomiris, C. and J. Mason, 2004. "Credit Card Securitization and Regulatory Arbitrage", *Journal of Financial Services Research*, Vol. 26, No. 1, August 2004.

Cordray, Ricahrd, 2009. *Ohio AG vs. Credit Rating Agencies*. Freely available at http://www.scribd.com/doc/22835949/Ohio-AG-complaint-vs-Credit-Rating-Agencies.

Dell'Ariccia, G., D. Igan and L. Laeven, 2008. "Credit Booms and Lending Standards: Evidence From The Subprime Mortgage Market", CEPR Discussion Papers 6683.

Demyanyk, Y. and O. Van Hemert, 2007. "Understanding the Subprime Mortgage Crisis", Federal Reserve Bank of St. Louis Supervisory Policy Analysis Working Papers 2007-05.

Diamond, D. 1985. "Optimal Release of Information By Firms", *The Journal of Finance* 40 No. 4, 1071-1094.

Diamond, D. and R. Verrecchia, 1991. "Disclosure, Liquidity and the Cost of Capital", *The Journal of Finance*, 46, 1325-1359.

Fishman, M. and K. Hagerty, 1990. "The Optimal Amount of Discretion to Allow in Disclosure", *The Quarterly Journal of Economics*, 105, 427-444.

Glaeser, E. and H. Kallal. 1997. "Thin Markets, Asymmetric Information, and Mortgage-Backed Securities", *Journal of Financial Intermediation* 6, 64-86.

Gorton, G. and G. Pennacchi, 1995. "Banks and Loan Sales: Marketing Nonmarketable Assets", *Journal of Monetary Economics* 35, 389-411.

Grossman, S., 1976. "On the Efficiency of Competitive Stock Markets Where Trades Have Diverse Information", *The Journal of Finance*, Vol. 31, No. 2, Papers and Proceedings of the Thirty-Fourth Annual Meeting of the American Finance Association Dallas, Texas December 28-30, 1975 (May, 1976), pp. 573-585.

Grossman, S. and J. Stiglitz, 1980. "On the Impossibility of Informationally Efficient Markets", *The American Economic Review*, Vol. 70, No. 3 (Jun., 1980), pp. 393-408.

Hellwig, M., 1980. "On the Aggregation of Information in Competitive Markets", *Journal of Economic Theory*, 22, 477-498, 1980.

Hakansson, N., G. Kunkel, and J. Ohlson. 1982. "Sufficient and Necessary Conditions for Information to Have Social Value in Pure Exchange", *The Journal of Finance* 37 December, 1169-81.

Hirshleiffer, J. 1971. "The Private and Social Value of Information and the Reward to Inventive Activity", *American Economic Review* 61, 561-74.

Keys, B., T. Mukherjee, A. Seru and V. Vig, 2010 "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans," *The Quarterly Journal of Economics,* 125(1), 307-362.

Mayer, C., K. Pence, and S. Sherlund. 2009. "The Rise in Mortgage Defaults." *Journal of Economic Perspectives*, 23(1): 27–50.

Mian A. and A. Sufi, 2008. "Summary of "the consequences of mortgage credit expansion", Proceedings, Federal Reserve Bank of Chicago, issue May, 129-132.

Milligan, Jack, 2008. "The Model Meltdown", *Mortgage Banking*, April 1, 2008. The article is freely available at http://www.thefreelibrary.com/The+model+meltdown%3a+the+credit-rating+agencies+were+caught+by...-a0178442313.

Parlour C. and G. Plantin, 2008. "Loan Sales and Relationship Banking", *The Journal of Finance* 63, 1291-1314.

Pennacchi, G. 1988. "Loan Sales and the Cost of Bank Capital", *The Journal of Finance*, 43, 375-95.

Peress, J., 2010. "The Trade-off between Risk Sharing and Information Production in Financial Markets", Journal of Economic Theory,145(1), 124-155

Petersen M. and R. Rajan, 2002. "Does Distance Still Matter? The Information Revolution in Small Business Lending," *The Journal of Finance*, 57(6), 2533-2570.

Rajan, U., A. Seru and V. Vig. 2010. "The Failure of Models That Predict Failure: Distance, Incentives and Defaults", Chicago GSB Research Paper No. 08-19; EFA 2009 Bergen Meetings Paper; Ross School of Business Paper No. 1122. Available at SSRN: http://ssrn.com/abstract=1296982.

Verrecchia, R., 1982. "Information Acquisition in a Noisy Rational Expectations Economy", *Econometrica*, Vol. 50, No. 6 (Nov., 1982), pp. 1415-1430.

Appendix

Optimal Decision of Holding/Selling Loans

Within the framework of the basic model. Consider the problem where the bank chooses the fraction of assets to sell. Given that the bank forms N shares of securities, it chooses to sell αN shares to investors. The bank's problem is

$$\max_{\alpha,\pi} EU(\alpha N\widetilde{p}^I + (1-\alpha)\widetilde{p}^B - c(\pi)),$$

where

$$\begin{aligned}\widetilde{p}^I &= \frac{n}{N}\mu - \frac{n\widetilde{\mu}_{\sigma 1}}{2rN^2} - \frac{n^2\sigma_{\sigma_1}^2}{8r^3N^4};\\ \widetilde{p}^B &= \frac{n}{N}\mu - \frac{n\widetilde{\mu}_{\sigma 1}}{2\rho N^2} - \frac{n^2\sigma_{\sigma_1}^2}{8\rho^3N^4}.\end{aligned}$$

Hence the FOC with respect to α is

$$(\frac{1}{r}-\frac{1}{\rho})[\frac{n\mu_\sigma}{2N} + \frac{n^2}{4\rho N^2}\frac{\pi\sigma_\sigma^4}{\pi\sigma_\sigma^2+1}(\frac{\alpha}{r}+\frac{1-\alpha}{\rho})] + \frac{n^2}{8N^3}\frac{\pi\sigma_\sigma^4}{\pi\sigma_\sigma^2+1}(\frac{1}{r^3}-\frac{1}{\rho^3}) = 0.$$

When $\rho < r$, it is optimal to set $\alpha = 1$. When $\rho > r$, it is optimal to set $\alpha = 0$. In all the interesting cases, when the bank chooses to produce a positive amount of information, $\rho > r$, that is it is also optimal for it to actually keep the loans if it can.

Non-central Chi-Square Distributed Variance and a Risk Neutral Bank

In this section, I sketch the intuition for an alternative set-up, in which the variance is distributed non-central Chi-square and the bank is risk neutral. The economic environment and problems are otherwise the same as the basic model.

Denote the return to an individual loan as $\widetilde{u} \sim \mathrm{N}(\mu, \sigma^2)$. Consider an auxiliary random variable $\widetilde{\sigma} \sim \mathrm{N}(\mu_\sigma, \sigma_\sigma^2)$ The variance of the return is the square of this auxiliary random variable. The signal structure is

$$\widetilde{s} = \widetilde{\sigma} + \widetilde{\epsilon}, \widetilde{\epsilon} \sim \mathrm{N}(0, \sigma_\epsilon^2) \text{ independent from } \widetilde{\sigma}.$$

The Bayesian posterior given a signal $\widetilde{s} = s$ is

$$\widetilde{\sigma}|\widetilde{s} = s \sim \mathrm{N}(\frac{\sigma_\epsilon^2}{\sigma_\sigma^2 + \sigma_\epsilon^2}\mu_\sigma + \frac{\sigma_\sigma^2}{\sigma_\sigma^2 + \sigma_\epsilon^2}s, \frac{\sigma_\sigma^2\sigma_\epsilon^2}{\sigma_\sigma^2 + \sigma_\epsilon^2}).$$

Denote $\mu_{\sigma 1}(s) = \frac{\sigma_\epsilon^2}{\sigma_\sigma^2+\sigma_\epsilon^2}\mu_\sigma + \frac{\sigma_\sigma^2}{\sigma_\sigma^2+\sigma_\epsilon^2}s$ and $\sigma_{\sigma 1}^2 = \frac{\sigma_\sigma^2\sigma_\epsilon^2}{\sigma_\sigma^2+\sigma_\epsilon^2}$. In this set-up, the normalized variance $(\frac{\widetilde{\sigma}-\mu_{\sigma 1}(s)}{\sigma_{\sigma 1}})^2$ is distributed $\chi^2(1)$.

Assume $\sigma_\sigma^2 < r^2$. This implies that $\sigma_{\sigma 1}^2 < r^2$. We will see from what follows that this assumption guarantees well-defined normal densities. Now consider the investors' willingness to pay in the first problem.

$$\begin{aligned}
\exp(-\frac{p(s)}{r}) &= E[\exp(-\frac{\widetilde{u}}{r})|\widetilde{s} = s] \\
&= \int\int \frac{1}{2\pi\sigma\sigma_{\sigma 1}} \exp\{-\frac{1}{2}[\frac{2u}{r} + \frac{(u-\mu)^2}{\sigma^2} + \frac{(\sigma - \mu_{\sigma 1}(s))^2}{\sigma_1^2}]\}d\sigma du \\
&= \int\int \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(u - \frac{\mu r - \sigma^2}{r})^2}{2\sigma^2}\}du \frac{1}{\sqrt{2\pi}\sigma_{\sigma 1}} \exp\{\frac{\sigma^2 - 2\mu r}{2r^2} - \frac{(\sigma - \mu_{\sigma 1}(s))^2}{2\sigma_{\sigma 1}^2}\}d\sigma \\
&= \int \frac{1}{\sqrt{2\pi}\sigma_{\sigma 1}} \exp\{-\frac{(\sigma - \frac{r^2\mu_{\sigma 1}(s)}{r^2-\sigma_{\sigma 1}^2})^2}{2\frac{r^2\sigma_{\sigma 1}^2}{r^2-\sigma_{\sigma 1}^2}}\}d\sigma \exp\{\frac{r^2\mu_{\sigma 1}(s)^2}{2\sigma_{\sigma 1}^2(r^2 - \sigma_{\sigma 1}^2)} - \frac{\mu}{r} - \frac{\mu_{\sigma 1}(s)^2}{2\sigma_{\sigma 1}^2}\} \\
&= \frac{r}{\sqrt{r^2 - \sigma_{\sigma 1}^2}} \exp\{\frac{\mu_{\sigma 1}(s)^2}{2(r^2 - \sigma_{\sigma 1}^2)} - \frac{\mu}{r}\} \\
\Rightarrow p(s) &= \mu - \frac{r\mu_{\sigma 1}(s)^2}{2(r^2 - \sigma_{\sigma 1}^2)} - r\ln\frac{r}{\sqrt{r^2 - \sigma_{\sigma 1}^2}}.
\end{aligned}$$

Now since the bank is risk neutral, it only cares about the expected profit,

$$\overline{\Pi}(\pi) = n\mu - \frac{nr}{2(r^2 - \sigma_{\sigma 1}^2)}E\mu_{\sigma 1}(s)^2 - nr\ln\frac{r}{\sqrt{r^2 - \sigma_{\sigma 1}^2}} - c(\pi),$$

where

$$
\begin{aligned}
E\mu_{\sigma 1}(\widetilde{s})^2 &= E(\frac{\sigma_\epsilon^2}{\sigma_\sigma^2+\sigma_\epsilon^2}\mu_\sigma + \frac{\sigma_\sigma^2}{\sigma_\sigma^2+\sigma_\epsilon^2}\widetilde{s})^2 \\
&= \frac{\sigma_\epsilon^4}{(\sigma_\sigma^2+\sigma_\epsilon^2)^2}\mu_\sigma^2 + \frac{2\sigma_\epsilon^2\sigma_\sigma^2}{(\sigma_\sigma^2+\sigma_\epsilon^2)^2}\mu_\sigma^2 + \frac{\sigma_\sigma^4}{(\sigma_\sigma^2+\sigma_\epsilon^2)^2}E\widetilde{s}^2 \\
&= \frac{\sigma_\epsilon^4}{(\sigma_\sigma^2+\sigma_\epsilon^2)^2}\mu_\sigma^2 + \frac{2\sigma_\epsilon^2\sigma_\sigma^2}{(\sigma_\sigma^2+\sigma_\epsilon^2)^2}\mu_\sigma^2 + \frac{\sigma_\sigma^4}{(\sigma_\sigma^2+\sigma_\epsilon^2)^2}(\mu_\sigma^2+\sigma_\sigma^2+\sigma_\epsilon^2) \\
&= \mu_\sigma^2 + \frac{\sigma_\sigma^4}{\sigma_\sigma^2+\sigma_\epsilon^2} \text{ decreasing in } \sigma_\epsilon^2,
\end{aligned}
$$

as is the variance of the price in the basic model. Hence, the second-order effect of signals manifests itself through the expected value of the posterior variance, in contrast to the variance of the posterior variance in the basic model. $var(\mu_{\sigma 1}(\widetilde{s}))$ affects $E\mu_{\sigma 1}(\widetilde{s})^2$ directly, and therefore affects the expected profit directly. The mechanism remains the same as that in the main text, but with somewhat less clear-cut interpretation. An increase in the precision of the signal decreases posterior variance, $\sigma_{\sigma 1}^2$, which tends to increase the expected profit; on the other hand, it also increases the expectation of the posterior variance $E\mu_{\sigma 1}(\widetilde{s})^2$, which tends to decrease the expected profit. The optimal level of information balances these two forces.

Now compare Problem 1 and Problem 2. Assume $N = n$.In Problem 2, $\widetilde{u} \sim \mathrm{N}(\mu, \frac{\sigma^2}{n})$, where ex ante $\widetilde{\sigma} \sim \mathrm{N}(\mu_\sigma, \sigma_\sigma^2)$. The willingness to pay is derived as follows,

$$
\begin{aligned}
\exp(-\frac{p(s)}{r}) = & E[\exp(-\frac{\widetilde{u}}{r})|\widetilde{s}=s] \\
= & \int\int \frac{1}{2\pi\frac{\sigma}{\sqrt{n}}\sigma_{\sigma 1}} \exp\{-\frac{1}{2}[\frac{2u}{r} + \frac{n(u-\mu)^2}{\sigma^2} + \frac{(\sigma-\mu_{\sigma 1}(s))^2}{\sigma_{\sigma 1}^2}]\} d\sigma du \\
= & \int\int \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n}}} \exp\{-\frac{(u-\frac{\mu rn-\sigma^2}{rn})^2}{2\frac{\sigma^2}{n}}\} du \frac{1}{\sqrt{2\pi}\sigma_{\sigma 1}} \exp\{\frac{\sigma^2-2\mu rn}{2r^2 n} - \frac{(\sigma-\mu(s))^2}{2\sigma_{\sigma 1}^2}\} d\sigma \\
= & \int \frac{1}{\sqrt{2\pi}\sigma_{\sigma 1}} \exp\{-\frac{(\sigma-\frac{r^2 n\mu_{\sigma 1}(s)}{r^2 n-\sigma_{\sigma 1}^2})^2}{2\frac{r^2 n\sigma_{\sigma 1}^2}{r^2 n-\sigma_{\sigma 1}^2}}\} d\sigma \exp\{\frac{r^2 n\mu_{\sigma 1}(s)^2}{2\sigma_{\sigma 1}^2(r^2 n-\sigma_{\sigma 1}^2)} - \frac{\mu}{r} - \frac{\mu_{\sigma 1}(s)^2}{2\sigma_{\sigma 1}^2}\} \\
= & \frac{r}{\sqrt{r^2-\frac{\sigma_{\sigma 1}^2}{n}}} \exp\{\frac{\mu_{\sigma 1}(s)^2}{2n(r^2-\frac{\sigma_{\sigma 1}^2}{n})} - \frac{\mu}{r}\} \\
\Rightarrow p = & \mu - \frac{r\mu_{\sigma 1}(s)^2}{2n(r^2-\frac{\sigma_{\sigma 1}^2}{n})} - r\ln\frac{r}{\sqrt{r^2-\frac{\sigma_{\sigma 1}^2}{n}}}.
\end{aligned}
$$

The expected profit in this case is

$$
\widehat{\Pi}(\pi) = n\mu - \frac{r}{2n(r^2-\frac{\sigma_{\sigma 1}^2}{n})} E\mu_{\sigma 1}(s)^2 - nr\ln\frac{r}{\sqrt{r^2-\frac{\sigma_{\sigma 1}^2}{n}}} - c(\pi).
$$

One can show that $\overline{\pi} \geq \widehat{\pi}$, for pretty much the same reason, that is the diversification implied by the securitization decreases the marginal benefit of information.

Numerical Example: Approximation of Truncated Normally Distributed Variances

In the model with heterogeneous lending opportunities, I gave a numerical example based the theoretical results obtained under normality assumptions. Here I verify that normally distributed variances under the current parametrization do not affect the results much. Under the assumption that $\mu_{\sigma i} > 4\sigma_{\sigma i}, \forall i$, the prior distribution of the variance, which is a truncated normal distribution is well approximated by an unrestricted normal distribution.

Problem 1

For all loans, the optimal signal is quite precise: $\overline{\pi}$ is in the neighborhood of 3.4694. This implies that the posteriors

$$\begin{aligned} \widetilde{\sigma}^2|\widetilde{s}^2 &= s^2 \dot{\sim} \mathrm{N}(s^2, 0.288); \\ \widetilde{s}^2 &\sim \mathrm{N}(\mu_{\sigma i}, \sigma_{\sigma i}^2 + 0.2882). \end{aligned}$$

Since by assumption $\mu_{\sigma i} > 4\sigma_{\sigma i}$, $\forall i$., the resulting $\widetilde{s}^2$ is well approximated by $\mathrm{N}(\mu_{\sigma i}, \sigma_{\sigma i}^2 + 0.2882)$.

Problem 2

For lower indexed loans, which has $\pi = 0$, the posterior distribution coincides with the prior distribution. However, at $n^{**} = 20$, $\widehat{\pi}_{20}^{20} = 0.0204$, which implies

$$\begin{aligned} \widetilde{\sigma}_{20}^2|\widetilde{s}_{20}^2 &= s^2 \sim \mathrm{N}(4.95 + 0.97s^2, 47); \\ \widetilde{s}^2 &\sim \mathrm{N}(152.72, 1506.6), \text{ with } \frac{152.72}{\sqrt{1506.6}} = 3.93. \end{aligned}$$

The unconditional probability of a signal being greater than 23, so that the conditional distribution has a negligible lower tail, is almost 1, i.e. $\Pr(\widetilde{s}^2 > 23) = 1 - 4.1598e - 4$.

Chapter 3 Using Subjective Expectations Data to Allow for Unobserved Heterogeneity in Hotz-Miller Estimation Strategies

3.1 Introduction

Progress on structural estimation within applied microeconomics has been limited, given the difficulty of implementation in "frontal" or "full solution" strategies, i.e. strategies that solve the complicated optimization and/or equilibrium problem at each trial of the structural parameter vector in the estimation routine.[20] The work of Hotz & Miller (1993) shows how to estimate the structural parameters of a discrete choice dynamic programming model without solving the optimization problem even once. The Hotz-Miller strategy has generated some methodological work on estimation of structural models that builds upon this initial insight[21]. However, an inherent problem in the Hotz-Miller type of strategy exploited by these papers is that, because of its very own

[20] Within the full solution paradigm, Rust (1987) and Keane & Wolpin (1994,1997) provided subtantial computational savings that stimulated most of the empirical research to date with these type of models. See Keane & Wolpin (2009), Todd & Wolpin (2009), and Keane, Todd & Wolpin (2010) for surveys of a substantial number of applications using full solution methods in development, labor, consumer behavior and other fields in applied microceconomics. More recently, Su & Judd (2007) proposed a novel, promising approach (MPEC) to further alleviate the computational burden associated with estimation via full solution methods by recasting the problem in a constrained optimization framework. See also Dube, Fox and Su (2009).

[21] See Hotz, Miller, Sanders & Smith (1994) to extend the original estimator to deal with the "Data Curse of Dimensionality" and for possible generalizations to allow for continuous choices and states. See Aguirregabiria & Mira (2002) for a recursive implementation of Hotz-Miller and for convergence to Full Information Maximum Likelihood. See Altug & Miller (1998) for a consistent account of aggregate shocks. See Jofre-Bonet & Pesendorfer (2003) for dynamic auctions. See Golan & Levy-Gayle (2008) for estimation of dynamic dignaling models. See Bajari, Hong, Krainer & Nekipelov (2009) for similar ideas applied to estimation of static games. See Aguirregabiria & Mira (2007), Bajari, Benkard & Levin (2007), Pesendorfer & Schmidt-Dengler (2008) and Pakes, Ostrovsky & Berry (2008) for dynamic discrete games and Choo & Siow (2005) for the use of Hotz-Miller approaches in facilitating estimation of a dynamic two-sided matching game.

nature, it cannot accommodate permanent sources of unobserved heterogeneity.[22] The first step recovers equilibrium behavior policies from the data, and as such, these can only be recovered based on observables. On the other hand, the more computationally intensive "frontal strategies" can handle permanent unobserved heterogeneity by integrating out the unobserved types in the likelihood function.[23]

Given its computational simplicity but its limitation regarding the handling of unobserved heterogeneity, in recent years there have been some efforts directed towards generalizing the Hotz-Miller approach to allow for unobserved heterogeneity.[24] In this paper we explore the potential use of expectation data such as, for example, subjective assessments of future choice probabilities to allow for estimable unobserved heterogeneity in these two-step estimation strategies for dynamic structural models.[25] We show that while requiring a particular type of data, our strategy can be an interesting alternative in the toolkit of structural microeconometricians if and when such data is available. In that sense, we think of our approach as complementary to the above literature. Our aim is to

---

22 This important limitation was noted early on by Eckstein & Wolpin (1989) among others.

23 This is the so-called Heckman & Singer (1984) approach taken by Wolpin (1984), van der Klaauw (1996), Keane and Wolpin (1997), Eckstein and Wolpin (1999) , Carro & Mira (2006), Mira (2007), Arcidiacono, Khwaja and Ouyang (2007), Blau & Gilleskie (2008), Liu, Mroz & van der Klaauw (2009), among many others. Alternative approaches to handle unobserved heterogeneity, which still require DP solutions have been advanced by Ackerberg (1999,2009) and Bajari, Fox, Kim & Ryan (2009). Whether discrete or continuous, parametric or non-parametric, all of the above are "random effects" approaches in the sense that only the probability of an observation being of a given type is contemplated.

24 Buchinsky, Hahn & Hotz (2005) propose a clustering approach that is similar to ours in the sense of being essentially a fixed effects approach. Houde & Imai (2006) and Arcidiacono & Miller (2008) suggest alternative estimation strategies in a random effects context. Arcidiacono & Miller (2008) allow for the unobserved heterogeneity to transition in systematic ways over time. Kasahara & Shimotsu (2008, 2009a) and Hu & Shum (2009) focus on estimation and identification of related dynamic discrete choice models with time-invariant unobserved types.

25 We focus on expectations about future choice probabilitites because they are more widely available. Other questions may elicit expectations about the future value of some state variables and could also be used to identify types with our method.

expand the toolkit that empirical researchers have when it comes to estimating dynamic structural models in computationally feasible ways. Our explicit use of elicited subjective expectations distinguishes our contribution from these other approaches taken in the literature. We will be focusing on single agent models, as the availability of expectations data seems more widespread in areas more amenable to single agent applications. However our idea can be applied to multiple agent contexts, in particular to dynamic discrete games. Indeed, much of the literature that built upon the Hotz-Miller strategy to estimate dynamic games is now being generalized to allow for game and/or player level unobserved heterogeneity.[26],[27]

We first characterize the power of expectation data to identify and estimate these models with the computational simplicity of a Hotz-Miller type of approach while, at the same time, allowing for unobserved heterogeneity, assuming that expectations are precisely elicited. For example, we first assume we have the ideal scenario in which there is no "heaping" or "focal measurement error" in Self-Reported Choice Probabilities (SR-CPs from now on).[28] Second, we show that when the use of more realistic, *focal*, subjective expectation data is contemplated in real applications, most of our results from the "ideal" case hold. Finally, we characterize how a modified version of our "linking technology" can alleviate some of the problems created by focal, reference point-based SR-CPs, if we

---

26 Aguirregabiria & Mira (2007) , Aguirregabiria, Mira & Roman (2007), Arcidiacono & Miller (2008), Siebert and Zulehner (2008), Hu & Shum (2008). Blevins (2009)

27 Aguirregabiria & Mira (2009) provide a comprehensive overview of structural estimation in the context of dynamic discrete choice models using full solution and non-full solution methods. Their review covers single agent and mulltiple agent models

28 By "focal measurement error" we mean the systematic tendency of respondents to report round numbers (focal points) when assesing their future choice probabilities.

have more than one SR-CP available.[29]

In addition to the theoretical insight, several datasets already include this kind of questions so our estimation strategy can be readily applied in a variety of settings. In the U.S. alone, all the major longitudinal surveys such NLSY or HRS include these type of questions. Looking ahead, however, the insights from our proposed estimation strategy are also informative about questionnaire design. In particular, about how these SR-CPs should be elicited to add the most value in a computationally feasible structural estimation strategy.

Finally, it is worth mentioning that there exist two strands of literature on the use of expectations data that are somewhat, but not directly related to our work: a) Relaxing Rational Expectations. This is a strand of literature that uses expectation data in more direct but still very important manner. The basic idea is to leverage data on expectations to be more flexible about the modelling of expectations. Key contributions here are Manski (2004) and Attanasio (2009). b) Using expectations data in estimation strategies for structural models that do not exploit the Hotz-Miller inversion. In this approach, like in ours, the expectation data are directly linked to the expectations used in the optimization problem. See Wolpin & Gonul (1985), van der Klaauw (2000), Wolpin (1999) and van der Klaauw & Wolpin (2008) for important contributions. In these cases, it is shown that these data are similar to revealed choice data and their use can provide more efficient estimators.

[29] Throughout this paper we allow for a specific form of (lack of) precision in the elicited sujective expectations. Allowing for more flexible forms of self-reporting error in is certainly important. In principle our framework could be generalized to allow for more flexible forms of measurement error in self reports but such generalization is beyond the scope of this paper.

These are important gains in estimator efficiency, but the contribution of such expectation data in those contexts is somewhat different than the one explored here.

The rest of the paper is organized as follows: The next section presents an extremely simple machine replacement example. We will use this example throughout the paper to fix ideas. Section 3.3 adds unobserved heterogeneity to the set up and discusses alternative conditions under which the use of expectation data succeeds in identifying such heterogeneity. Section 3.4 provides Montecarlo experiments that describe the performance of our estimation strategy. Section 3.5 discusses some extensions for our framework. Conclusions follow.

3.2 Example: Estimating a Simple Dynamic Structural Model of Machine Replacement Decisions

Consider a simplified capital replacement problem similar to that in Rust (1987). Firms each use one machine to produce output in each period. These machines age, becoming more likely to breakdown, and in each time period the firms have the option of replacing the machines. Let x_t be the age of the machine at time t and let the expected current period profits from using a machine of age x_t be given by:

$$\Pi\left(x_t, d_t, \varepsilon_{0t}, \varepsilon_{1t}\right) = \begin{cases} \theta_1 x_t + \varepsilon_{0t} & \text{if } d_t = 0 \\ R + \varepsilon_{1t} & \text{if } d_t = 1 \end{cases}$$

where $d_t = 1$ if the firm decides to replace the machine at t, R is the net cost of a new machine, and the ε_ts are time specific shocks to the utilities/profits from replacing and not replacing. Let's assume that these ε_ts are i.i.d. across firms and time periods, and while not required for the implementation of our methods below, let's further assume that they follow a type I extreme value distribution. We consider a model with stochastic aging in which

$$x_{t+1} = \begin{cases} \min\{5, x_t + 1\} & \text{with probability } \pi_f & \text{if } d_t = 0 \\ x_t & \text{with probability } 1 - \pi_f & \text{if } d_t = 0 \\ 1 & \text{with probability } 1 & \text{if } d_t = 1 \end{cases}$$

Note that in this very simple model the state space only has 5 points and therefore full-solution methods can easily be used to estimate the model. We do this for illustrative purposes, but it should be kept in mind that the method we propose below can deal with more realistic state spaces in which standard full solution methods cannot be used or can only be used at subtantial computational cost. Estimation is standard, and can proceed using either Rust (1987) nested fixed point algorithm or Hotz-Miller (1993) two-step estimator, among other alternatives. The Hotz-Miller strategy avoids the solution of the complicated dynamic structural model. The associated optimization problem is not solved even once. However, one is able to recover the structural parameters and can, after estimation, solve the model at those parameters if needed for, say, baseline simulation of artificial data and/or counterfactual policy experiments.

3.3 Adding Unobserved Heterogeneity

We now modify the machine replacement example to allow for heterogeneity in the structural parameters capturing age related maintenance costs θ_{1k} and machine replacement costs R_k. We first consider the case of finite discrete types . We then analyze the continuous case.

In the discrete case we index types by $k = 1, ..., K$. An alternative set up considers the existence of unobserved state variables $x^u = \{x_1^u, x_2^u, ...\}$ or, alternatively, a single unobserved discrete state variable $k_t \in \{1, ..., K\}$ that captures every possible combination of unobserved states in x^u. We allow for the possibility that unobserved states may transition over time and allow for this transition to potentially depend on the choice d_t. Then, in general, we consider a state transition given by

$$f_{xk}\left(x', k' | d, x, k\right)$$

We can entertain several assumptions that restrict the generality of $f_{xk}\left(x', k' | d, x, k\right)$

- **Assumption F1($(\mathbf{x}, \mathbf{k})$ are conditionally independent):** conditional on (d, x, k), x' and k' are independently distributed:
$$f_{xk}\left(x', k' | d, x, k\right) = f_k\left(k' | d, x, k\right) f_x\left(x' | d, x, k\right)$$

Similar to Arcidiacono & Miller (2008), we can also assume that

- **Assumption F2 (Exogenous Transitions for Unobserved States):** the transition of the unobserved state variables does not depend on the current choice nor the current observed state, but follows an exogenous and flexible markov stochastic process:
$$f_k\left(k' | d, x, k\right) = f_k\left(k' | k\right) = \pi_{kk'}$$

As in much of the literature using full-solution methods, in some situations we can further assume

- **Assumption F3 (Time Invariant Unobserved Heterogeneity):** the unobserved states are time invariant.
$$\pi_{kk} = 1 \text{ for all } k_t \in \{1, ..., K\}$$

In some cases we can further assume that

- **Assumption F4 (Homogeneous Transitions for Observed State Variables):** the evolution of the observed states, x, does not depend on the unobserved heterogeneity, k.
$$f_x\left(x' | d, x, k\right) = f_x\left(x' | d, x\right) \text{ for all } k \in \{1, ..., K\}$$

In this setup, a standard estimation strategy would proceed by integrating out unobserved heterogeneity in the likelihood function, treating types as discrete random effects in the population. Alternatively, a modification of the Hotz-Miller strategy, exploiting subjective probabilities of future choices, can be used to estimate the structural parameters allowing for unobserved heterogeneity and without solving the dynamic program. In the remaining of this section we consider this possibility in detail.

3.3.1 Estimation using Hotz-Miller with Precise Subjective Choice Probability Data

Suppose we have available self-reported probabilities of next period machine replacement for each firm after the current period replacement decisions have been made.

Let

$$p_i^{SR}\left(d_{t'+1} = 1 | x_{t'}, d_{t'}, k\right)$$

denote the 1-period ahead self-reported probability of choosing $d = 1$ (replacement choice) at time $t' + 1,$ elicited at t' from the technician in charge of machine maintenance at firm i, of unobserved type $k,$ who, in addition is at the observed state $x_{t'}$ and who has recently made choice $d_{t'}.$ Throughout this section we assume that these probabilities are elicited with great precision. For future reference we establish this feature of the data in the following assumption

- **Assumption SR-Precise:** The subjective probabilities are elicited with precision. In particular, self-reports are not rounded off to the nearest "focal" probability.

A key question is then: under what conditions can we use these expectation data to reveal the underlying unobserved heterogeneity? The basic intuition can be grasped in the context of our machine replacement example. Presumably if we have two firms A and B with machines in the same state in the current period $x_{At} = x_{Bt} = x_t$*, and these two firms make the same choice,* $d_{At} = d_{Bt} = d_t$, but report different probabilities of replacement tomorrow,

$$
\begin{aligned}
p^{SR}\left(d_{A,t+1}=1|x_{At},d_{At},k_A\right) &\neq p^{SR}\left(d_{B,t+1}=1|x_{Bt},d_{Bt},k_B\right)\\
p^{SR}\left(d_{A,t+1}=1|x_t,d_t,k_A\right) &\neq p^{SR}\left(d_{B,t+1}=1|x_t,d_t,k_B\right)
\end{aligned}
$$

it must be the case that there is something unobserved by the econometrician but observed by the technician in charge of machine maintenance in each firm that induces the difference in the self-reports. In other words, the unobserved state k is different for the two firms, $k_A \neq k_B$. Therefore, differences in self-reports are informative about underlying unobserved heterogeneity. In particular, note that if there are only two types, there can only be two and only two different $p^{SR}\left(d_{t+1}=1|x_t,d_t,k\right)$ reported by observations that have the same state-choice combination (x_t,d_t).

It follows that the number of types, K can be readily identified by counting the number of different $p^{SR}\left(d_{t+1}=1|x,d,k\right)$ elicited at each state-choice cell, (x,d).[30] Can we use the self-reported probabilities to estimate the machine replacement problem a la Hotz-Miller but allowing for unobserved heterogeneity? The answer is yes. Below we provide details on how to do so.

3.3.1.1 Linking Technology and Type Revelation with Precise Self Reports

We now introduce our "linking technology". The basic idea is pretty simple and illustrates the power of eliciting self-reported choice probabilities to recover the underlying

[30] Note that it is important to consider future choice probability elicitation at particular state-choice combinations, not just particular states. The reason is that, among observations with the same state at t, x_t, those who make different choices will induce different probability distributions for the state variables next period, and then, even if they are of the same type, they will end up reporting different future choice probabilities. By focusing on those who are at the same state *and* made the same current choice, we avoid this problem.

unobserved heterogeneity.

Let's assume there are only two types $k = 1, 2$. Then at any time t, the set of observations i with common observable state x_t *and who made the same current choice* d_t must be either type $k = 1$ or type $k = 2$. If they are of the same type, they face the same prospects regarding their state variables next period.[31] Moreover, they also face a common distribution of idiosyncratic error terms next period $f(\varepsilon_{t+1})$. Hence, they will provide the same report about the probability of making the choice next period. However, observations that are of different unobserved types will report a different probability.[32] We should then see two, and only two, different values of SR-CPs for each observed state-choice combination.[33]. Essentially, self-reported probabilities allow us to "reveal" type membership. Then, after uncovering the unobserved type, estimation methods such as those proposed by Hotz & Miller (1993) or Hotz, Miller, Sanders & Smith (1994) apply directly, treating type as an additional observed discrete state. Moreover, for the purposes of identification, the model can be reduced to one without unobserved heterogeneity. Then identification results such as those in the work of Magnac & Thesmar (2002) apply directly.[34]

The linking technology, which we introduce more formally below, is a technique to

31 Note that under Assumption F4, they would face the same prospects even if they were of different types.

32 Different types having the same 1-period ahead choice probability is a measure-zero event if the choice is feasible next period and the utility of the choice depends on the type.

33 Note that this holds regardless of whether Assumption F4 is true or, instead, the transitions for the observed states depend on the unobserved type.

34 Magnac & Thesmar (2002) do consider identification of a model with correlated fixed effects without relying on expectations data. However, the structure of unobserved heterogeneity they focus on is somewhat different than the one considered here.

"link" observations in the data. The linking is done via self-reports, which act as the chain's interconnecting links for each unobserved type. When we have two self-reports available for each observation, the only, rather weak, requirement for the linking technology to work is the absence of isolated islands in the space of feasible state-choice combinations. These "isolated islands" are sets of state-choice combinations (x, d) in which the pairs of self-reports of individuals are all contained and have no bridges to other regions of the state-choice space. Below we set up some notation and more formally define the linking technology along with the No-Islands assumption in which we rule out the existence of such islands.[35]

Definition 1 (*Revelation of types*) *A revelation of types is defined by an equivalence relation $\sim$ on the set of observations $I = \{1, 2, ...N\}$. Call the cardinality of the quotient set $I/\sim$ the revealed number of types and denote it by M. By the Fundamental Theorem of Equivalence Relations, an equivalence relation $\sim$ on a set, partitions that set. The underlying parameter K is unknown. M does not necessarily recover K.*

Let the pair of self-reports be elicited at t' and t'' for all observations.

Definition 2 (*Linking Technology*) *Define a binary relation, R, in the following way: $\forall i, j \in \{1, 2, ...N\}$,*

$$i \; R \; j$$

iff

$$\{\widetilde{p}_i^{SR}(x_{it'}, d_{it'}), \widetilde{p}_i^{SR}(x_{it''}, d_{it''})\} \cap \{\widetilde{p}_j^{SR}(x_{jt'}, d_{jt'}), \widetilde{p}_j^{SR}(x_{jt''}, d_{jt''})\} \neq \emptyset.$$

The linking technology is a relation $\sim$ on $\{1, 2, ..., N\}$: $\forall i, j \in \{1, 2, ...N\} = I$,

$$i \sim j$$

iff $\exists$ a subset of observations $\{i_1, i_2, ...i_n\} \subseteq I$, such that

$$i \; R \; i_1 \; R \; i_2 \; R ... R \; i_n \; R \; j.$$

The linking technology defines an equivalence relation. It is easily checked that $\sim$ satisfies reflexivity, symmetry and transitivity.

Assumption SR-No Islands is defined after specifying a particular linking technology.

[35] Alternatively, observations in the isolated islands can be discarded provided that suitable assumptions about their representativeness hold.

- **Assumption SR-No Islands:** Define Σ^k to be the set of all state-choice cells at which a type k observation makes a self-report in the data. Then, $\forall(x,d),(x',d') \in \Sigma^k, \exists$ observations m and n of type k, with m reporting at (x,d), and n reporting at (x',d'), and $m \sim n$.

Lemma 5 *Under Assumption SR-Precise and SR-No Islands, the linking technology recovers the true number of types and type membership for each observation.*
Proof. *See Appendix.* ■

We focus on the case in which we have permanent unobserved heterogeneity or "types" and where we have 1-period ahead SR-CPs. In the extensions section we discuss some variations. In section 3.5.1 we consider the elicitation of S-periods ahead SR-Cumulative CP. Later in sections 3.5.2 and 3.5.3 we address the case in which the unobserved heterogeneity is continuous as well as the case in which unobserved state variables evolve as a Markov process. In the remainder of this section we maintain Assumptions F1 and F3.

3.3.1.2 1-Period Ahead SR-CPs

For now, let's assume the available self-reports are about 1-period ahead CPs. In general, these self-reports can occur before or after the choice has been made this period. In what follows, and unless noted otherwise, we assume that the 1-period ahead SR-CPs are elicited after the current choice, d_t has been made.

If the model in question were deterministic, it would be clear which state point next period the SR-CP is giving choice information about. In models with stochastic transitions we need a more detailed "theory of self-report" that specifies what goes through the respondent's mind between the time she listens to the question and the time she provides the answer. Our theory of self report is the following: We assume the question is asked at time t after x_t has been realized and d_t has been chosen. Upon listening to the question "what's the probability that you will set $d_{t+1} = 1$?" respondents use the solution to the dynamic programming problem to calculate the implied CCPs, $p\left(d_{t+1} = 1|x_{t+1}, k\right)$ at each feasible state next period, x_{t+1}. Note that there will be many probabilities, especially when the state space is large. After computing these, however, they need to provide a single answer. One reasonable way forward is to assume that respondents then report the average of these CCPs using the one-period ahead transition probability for the state variables, $f_x\left(x'|d, x\right)$ as weights. In other words, the question elicits the "expected CCP". Formally,

$$\begin{aligned}
\text{SR-CP} &= E\left[CCP\right] \\
p^{SR}\left(d_{t+1}|x_t, d_t, k\right) &= E_{x_{t+1}|x_t,d_t}\left[\Pr\left(d_{t+1}|x_{t+1}, k\right)\right] \\
&= \sum_{x_{t+1}} \Pr\left(d_{t+1}|x_{t+1}, k\right) f_x\left(x_{t+1}|d_t, x_t\right)
\end{aligned}$$

In some case, it is also possible that the question actually elicits the one-period ahead **CCP at the Modal State.** In this case the respondent reports the CCP at the mode of the distribution of her own state variables next period. Given homogenous transitions, we can infer what that state is and we are then back to the simpler deterministic case. We can then "link" the self-reports at all those states. We call this the "Solve and Link" strategy: We solve out for the implied (modal) observed state at which the self-reported probability is being elicited. Then we link all the CCPs to trace out the unobserved types in the observed state space. It should be emphasized that rather than being something like "what's the probability that you'd choose $d_{t+1} = 1$?" here we are assuming the question eliciting the SR-CP is something more like the following: "Look one period ahead and consider what's your most likely situation at that time. In that situation, what would be the probability that you'd choose $d = 1$?". Note that the introduction to the second question more explicitly instructs the respondent to situate himself in the most likely (modal) state one period ahead and then only report the choice probability assuming she will in fact be in that state. Formally, $p^{SR}\left(d_{t+1}|x_t, d_t, k\right) = \Pr\left(d_{t+1}|x^m_{t+1}, k\right)$ where $x^m_{t+1} = Mode\left(x_{t+1}|x_t, d_t\right)$ is the modal state at time $t+1$ given x_t and d_t. That is $x^m_{t+1} = \arg\max_{x_{t+1}}\left\{f\left(x_{t+1}|x_t, d_t\right)\right\}$

In what follows, and unless noted otherwise, we assume that subjective expectation questions elicit the expected CCP.

- **Assumption SR-E[CCP]**: The subjective probability questions elicit the expected CCP.

We focus on the case in which we have two self-reports available for each individual.[36] In this case we can work within a very general class of models. We can exploit the self-reports to group observations into types, without trying to recover the implied CCPs. By having at least two self-reports we can connect observations at different points in the state space who belong to the same type. In particular, any two observations who share one common self-report at a given state-choice combination are of the same type and their other self-reports add to our signals to identify that type. The "linking technology" is extremely powerful. By having two self-reports we can trace out types in unrestricted models in which the profile of choice probabilities for different types may be allowed

[36] Again, well known surveys such as NLYS and HRS do include at least two self reports about subjective probability of future choices for the same individual. If only one Self-Report is available we need to restrict ourselves to cases where the CCPs are monotonic on type across the state space. For example, we could restrict ourselves to a class of models where one type always has higher choice probability. This is an important restriction. When profiles of SR-CPs for different types "cross" at some point in the state space, identification problems arise.

to cross in the state space. The linking technology allows us to overcome the ambiguity created by these crossings.

Recovering type-specific CCPs using 2 Self-Reports of Expected CCPs. While the identification of number of types (and type membership for each observation) doesn't actually require it, we can also explore the conditions under which we can recover the actual type-specific conditional choice probabilities. We will later make use of these results in more complex settings, but it is useful to introduce the issue now. When expected CCPs are reported, the respondent reports an average of CCPs, with the average taken using the transition probability. To recover the underlying CCPs we use the alternative "Link and Solve" strategy:

(1) We first **link** SR-CPs from the same type and form a system of equations.
(2) We then **solve** the system of equations and recover the type-specific CCPs.

To be specific, the first 1-period ahead SR-CP reported at t' gives us one equation for respondent i of type k_i.

$$p^{SR}\left(d_{i,t'+1}=1|x_{it'},d_{it'},k_i\right)=\sum_{x_{i,t'+1}}\Pr\left(d_{i,t'+1}=1|x_{i,t'+1},k_i\right)f\left(x_{i,t'+1}|x_{it'},d_{it'},k_i\right)$$

where $p^{SR}()$ and $f()$ are known and $\Pr\left(d_{i,t'+1}=1|x_{i,t'+1},k_i\right)$ for all $x_{i,t'+1}$ are the unknowns. In general, we have $|X|$ unknowns so we need more equations. We then link this equation with a similar equation based on the respondent's second self-report and with the self-reports of other respondents j of the same type who have been linked to i to form a linear system of equations that has as many equations as unknowns.

$$\begin{aligned}
p^{SR}\left(d_{i,t'+1}=1|x_{it'},d_{it'},k_i\right) &= \sum_{x_{i,t'+1}} \Pr\left(d_{i,t'+1}=1|x_{i,t'+1},k_i\right) f\left(x_{i,t'+1}|x_{it'},d_{it'},k_i\right)\\
p^{SR}\left(d_{i,t''+1}=1|x_{it''},d_{it''},k_i\right) &= \sum_{x_{i,t''+1}} \Pr\left(d_{i,t''+1}=1|x_{i,t''+1},k_i\right) f\left(x_{i,t''+1}|x_{it''},d_{it''},k_i\right)\\
&\vdots\\
p^{SR}\left(d_{j,t'+1}=1|x_{jt'},d_{jt'},k_j\right) &= \sum_{x_{j,t'+1}} \Pr\left(d_{j,t'+1}=1|x_{j,t'+1},k_j\right) f\left(x_{j,t'+1}|x_{jt'},d_{jt'},k_j\right)
\end{aligned}$$

where $k_i = k_j, \forall i, j$ is guaranteed by the "linking technology". We can then solve for the CCPs, $\{\Pr(d=1|x,k)\}_{x\in X}$ by using standard techniques to solve systems of linear equations. There are $|X|$ unknowns and at most $|X| \times |D|$ different self-reports.

Note that once these type-specific CCPs have been recovered, they could be plugged-in directly instead of the non-parametric first stage probabilities in the typical Hotz-Miller two-step approach.

We have focused on discrete types, 1-period ahead self-reports and time-invariant unobserved heterogeneity. Our framework can be extend to relax each of these. We briefly discuss these extensions below in Section 3.5.

3.3.2 Estimation using Hotz-Miller with "Focal" Subjective Choice Probability Data

Unfortunately, in many contexts the SR-CPs are not as clean as we assumed them to be in the previous section. While people may take more care in thinking about these probabilities when making actual choices, it is likely that they exercise less care when quickly computing these probabilities in a few seconds when answering to the interviewer.[37] In particular, there is likely to be substantial "heaping" or "bunching" at common reference points like 0, 0.10, 0.50, 0.90 and 1. See Walker (2003), Hill, Perry & Willis (2004) and Blass, Lach & Manski (2010). Surprisingly, there is no bunching at 0.33 and 0.66 which a priori appear to be good focal points when the probability reflects 1 out 3 or 2 out of 3 odds. Interestingly, respondents seem to be more precise when reporting probabilities close to the boundaries. For example, it is not uncommon to observe self-reports of 0.01, 0.02, 0.98 and 0.99. It is understandable that respondents care more about distinguishing 0 from 0.01 or 0.99 from 1 than 0.50 from 0.51 or 0.49. We accommodate these empirical regularities of probability self-reporting behavior in our discussion below.

Therefore, in order to make our method more empirically relevant, in this section we address the issue of "less than ideal" subjective choice probability assessments and characterize to what extent the results derived in the previous sections hold in the more realistic case in which Assumption SR-Precise does not hold. We will work with a set of $B = 25$ "focal" or "reference" values, $b,$ that have been consistently found in practice to account for most of the self-reported probabilities.[38] With a little abuse of notation, let B also denote the cardinality of the set $B.$

[37] See Karni (2009) for a formalization of truthful elicitation of probabilities.

[38] Indeed, casual inspection of some of the responses to these type of questions in the National Longitudinal Survey of Youth 1997 NLSY97 reveals a pattern of clustering around the values in the particular set B we defined. However, our methods can be used with any set B. That is, the set of focal values can be modified on a case by case basis if the pattern of bunching in a specific survey is coarser or more detailed than this one.

Then

$$b \in B = \{0;1;2;5;10;15;20;25;30;35;40;45;50;$$
$$55;60;65;70;75;80;85;90;95;98;99;100\}.$$

Focal SR-CPs may lead to "bunching" which may create uncertainty in the identification of the types. Say, for example, we have two observations of different types at the same (x, d). For simplicity, consider 1-period ahead E[CCP] self-reports. Say under assumption SR-Precise type 1 reports 68% while type 2 reports 72%. Now, in a more realistic scenario in which SR-Precise no longer holds, we will have both types reporting 70%.

In this section we will show that a variation of our linking technology, coupled with mild assumptions on the pattern of bunching across types and about the sampling of self-reports, can succeed in overcoming this problem. We maintain assumptions $F1$, $F3$, and $F4$ on the transition probability for state variables

A precise self-report of i at time t is defined to be a function of x_{it}, d_{it}, and k_i , which can be 1-period- or s-period-ahead expected CCP, modal CCP, etc. Following the notation in the previous sections, let $p_i^{SR}(d_{t+1} = 1|x_{it}, d_{it}, k_i)$ be a the self-reported choice probability that satisfies SR-Precise. Now, consider two SRs at t' and t''. In this section, we want to focus on the case in which the SR-CPs are probabilities that are rounded-off to the nearest focal point. We add an F to the self-report probability notation to emphasize it is now a focal self-report: $p_i^{SRF}(x_{it}, d_{it}, k_i)$. Formally,

$$p_i^{SRF}(x_{it}, d_{it}, k_i) = \arg\min_{b\in B} |p_i^{SR}(d_{t+1} = 1|x_{it}, d_{it}, k_i) - b|$$

where $p_i^{SR}(d_{t+1} = 1|x_{it}, d_{it}, k_i)$ may be a modal CCP or an Expected CCP. Actually, if $p_i^{SR}(d_{t+1}=1|x_{it},d_{it},k_i)=E\left[CCP\right]$ we need to account for an additional layer of round-off in the underlying CCPs, which we then denote FCCPs:[39]

$$\begin{aligned} E\left[FCCP\right] &= \sum_{x_{i,t'+1}} \text{FCCP}\left(x_{i,t'+1},k_i\right) f\left(x_{i,t'+1}|x_{it'},d_{it'},k_i\right) \\ &= \sum_{x_{i,t'+1}} \left[\arg\min_{b\in B}\left|\text{CCP}\left(x_{i,t'+1},k_i\right)-b\right|\right] f\left(x_{i,t'+1}|x_{it'},d_{it'},k_i\right) \\ &= \sum_{x_{i,t'+1}} \left[\arg\min_{b\in B}\left|\Pr\left(d_{i,t'+1}=1|x_{i,t'+1},k_i\right)-b\right|\right] f\left(x_{i,t'+1}|x_{it'},d_{it'},k_i\right) \end{aligned}$$

We assume all observations follow this "rounding" procedure. Since k_i is unobserved, from the econometrician's point of view, the SRs can be associated with states and actions only: $\widetilde{p}_i^{SRF}(x_{it},d_{it})=p_i^{SRF}(x_{it},d_{it},k_i)$.

Definition 3 (Bunching) *Two SRs are said to be bunched at (x,d) for **observations** i and j of different types, if $p_i^{SRF}(x,d,k_i)=p_j^{SRF}(x,d,k_j)$ and $k_i\neq k_j$. Two SRs are said to be bunched at (x,d) for **types** k and k', if $p^{SRF}(x,d,k)=p^{SRF}(x,d,k')$.*

Note that "Bunching" is defined both, for observations and for types. When focal self-reports generate bunching in the data, some variation of our basic linking technology works under some additional assumptions.

Assumption B1 (Immediate Detection of Bunching Observations) If a pair of SRs by two observations i and j who belong to different types, bunch at the state-choice (x,d), then their other SRs must be elicited at another common state-choice (x',d'), at which the two types' focal SRs differ:

$$\widetilde{p}_i^{SRF}(x',d')\neq\widetilde{p}_j^{SRF}(x',d').$$

Assumption B1 essentially makes sure that all bunchings of a pair of observations can

[39] This additional layer of rounding off corresponds to the idea that an additional source of discrepancy between the theoretical $E\left[CCP\right]$ and the self-report resides in the respondent's inability to exactly compute the value function "off the top of her head". This inability induces computation of FCCPs, rather than CCPs at each feasible state point next period. Then, a second layer of rounding is introduced when the average of these rounded CCPs is itself rounded off when the answer is provided to the interviewer. Note that this assumption only introduce some limited rationality at the self-report stage. Behavior continues to be fully rational.

be detected immediately. It will be relaxed later in the sense that we will not require immediate detection of bunching observations, but will require detection of bunching types.

Definition 4 (Bunching state-choice for $\{i, j\}$) *The bunching state-choice for $\{i, j\}$ is the state-choice (x, d) at which their SRs bunch. Denote it by (x^B_{ij}, d^B_{ij}).*

Assumption B1 guarantees that whenever there are two observations i and j of different types reporting at a bunching state-choice for them, the bunching of different types is immediately detected. Hence, the to-be-defined "linking technology under bunching" can use this notion of "bunching state-choice".

In particular, under Assumption B1, two observations i and j can bunch at most at one state-choice cell.

In Figure 3.1, the squares mark the precise SRs, which are rounded-off to the nearest focal points, marked by circles. Whenever there is bunching of two different precise SRs, we include the square-marked precise SRs for illustrative purposes. As is evident from the figure, observations i and j have the same focal self-reports at the state-choice (x^B_{ij}, d^B_{ij}).

However, Assumption B1 is not enough to identify the types. Consider the following example in Figure 3.2. There is no way of telling whether the observations are grouped as $\{j, i\}$ and $\{g, h\}$ or $\{i, g\}$ and $\{h, j\}$. In light of this, we make Assumption B2, which bridges the two SRs by the same type.

Assumption B2 (Bridging Bunchings) For all observations i and j who belong to the same type, but the singleton intersection of whose SRs is at (x^B_{ih}, d^B_{ih}) for some h, there exists another observation l of the same type as i and j, who has SRs in the two non-bunching state-choice cells.

Figure 3.1: Illustration of Immediate Detection of Bunching Observations

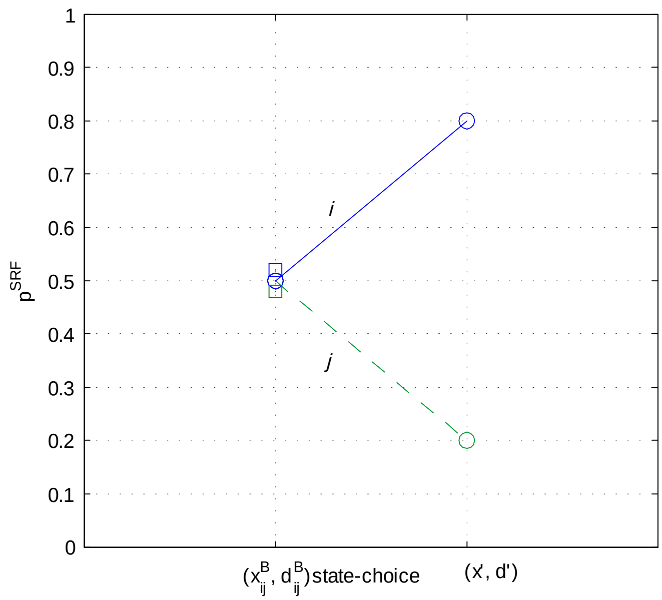


Figure 3.2: Problem Without Assumption B2

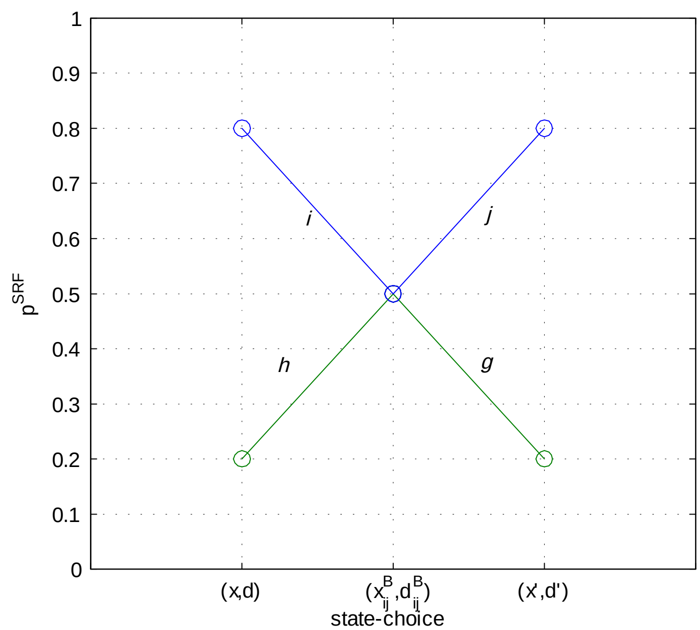


Figure 3.3: Illustration of Bridging the Bunching

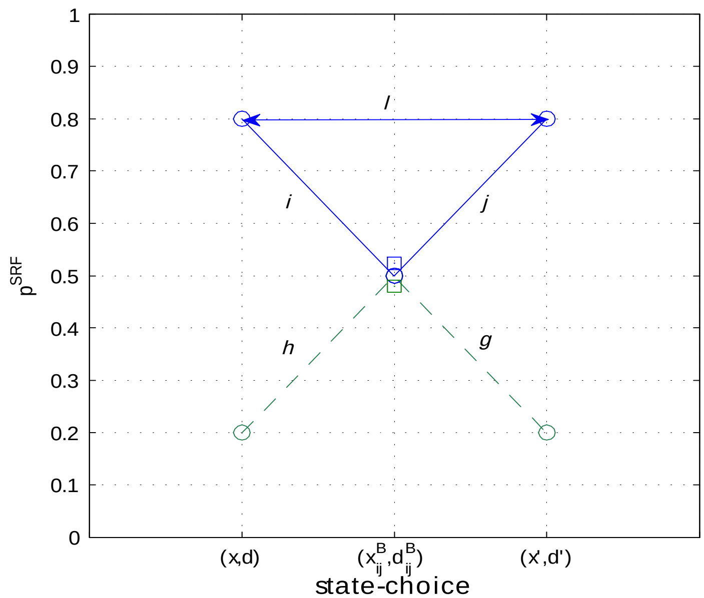


Figure 3.3 illustrates how observation l bridges a bunching.

Definition 5 (Linking Technology under Bunching) *Define a binary relation, R^B, in the following way:* $\forall i, j \in \{1, 2, ...N\}$,

$$i \; R^B \; j$$

iff the following conditions are met:

(1) The pairs of self reports for i and j are such that

$$\{\widetilde{p}_i^{SRF}(x_{it'}, d_{it'}), \widetilde{p}_i^{SRF}(x_{it''}, d_{it''})\} \cap \{\widetilde{p}_j^{SRF}(x_{jt'}, d_{jt'}), \widetilde{p}_j^{SRF}(x_{jt''}, d_{jt''})\} \neq \emptyset,$$

(2) if $\exists$ observation h,

$$\{\widetilde{p}_i^{SRF}(x_{it'}, d_{it'}), \widetilde{p}_i^{SRF}(x_{it''}, d_{it''})\} \cap \{\widetilde{p}_j^{SRF}(x_{jt'}, d_{jt'}), \widetilde{p}_j^{SRF}(x_{jt''}, d_{jt''})\} = \{\widetilde{p}_i^{SRF}(x^B_{ih}, d^B_{ih})\}$$

then

$$\begin{aligned} & \exists l, \{\widetilde{p}_l^{SRF}(x_{lt'}, d_{lt'}), \widetilde{p}_l^{SRF}(x_{lt''}, d_{lt''})\} \\ = \; & \{\widetilde{p}_i^{SRF}(x_{it'}, d_{it'}), \widetilde{p}_i^{SRF}(x_{it''}, d_{it''})\} \triangle \{\widetilde{p}_j^{SRF}(x_{jt'}, d_{jt'}), \widetilde{p}_j^{SRF}(x_{jt''}, d_{jt''})\}, \end{aligned}$$

where $\triangle$ denotes the set difference.

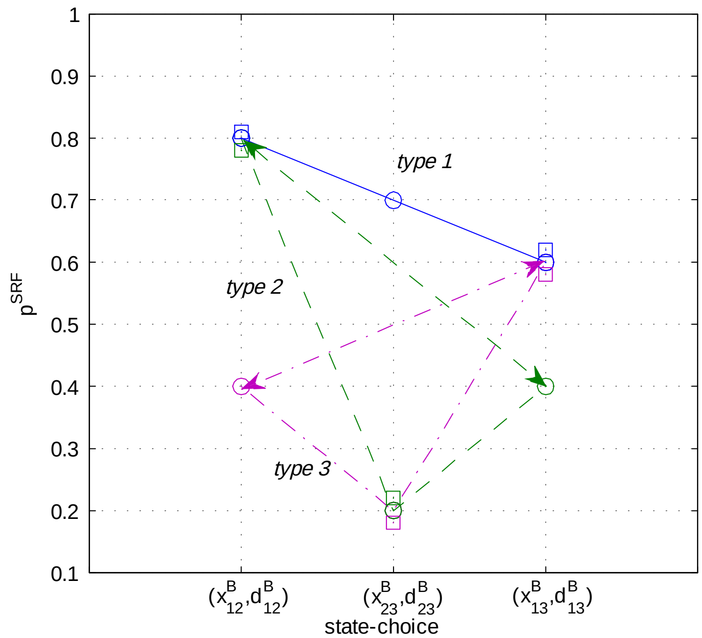
The *linking technology under bunching* is a relation $\sim^B$ on $\{1, 2, ..., N\}$: $\forall i, j \in \{1, 2, ...N\} = I$,

$$i \sim^B j$$

– iff $\exists$ a subset of observations $\{i_1, i_2, ...i_n\} \subseteq I$, such that

$$i \; R^B \; i_1 \; R^B \; i_2 \; R^B ... R^B \; i_n \; R^B \; j.$$

Figure 3.4: Identification of the Number of the Types



It can be easily proved that the linking technology under bunching also defines an equivalence relation.

Lemma 6 *Under Assumptions B1, B2 and SR-No Islands, the linking technology under bunching recovers the true types exactly.*

Proof. See Appendix. ■

Note that the number of types is identified after the partition. In particular, it is not identified by counting the number of different SRs in each state-choice cell. Consider Figure 3.4. The partition identifies 3 types, though at each state-choice cell, there are only 2 different SRs. With a slight abuse of notation, $(x^B_{kk'}, d^B_{kk'})$ here denotes the bunching state-choice cell for type k and type k'. The arrows indicate "bridges".

Assumption B3 **(Detection of Bunching Types)** If two types, k and k', bunch at the state-choice (x,d), then $\exists$ two observations i of type k and j of type k', and another state-choice (x',d') s.t.

$$\begin{cases} \widetilde{p}_i^{SRF}(x,d) = \widetilde{p}_j^{SRF}(x,d) \\ \widetilde{p}_i^{SRF}(x',d') \neq \widetilde{p}_j^{SRF}(x',d') \end{cases}$$

Assumption B3 is weaker than Assumption B1. Assumption B1 ensures that whenever two observations of different types bunch, their other SRs reveal the bunching to the researcher. Assumption B3 only requires that whenever two types bunch, some observations' SRs reveal the bunching of the types to the econometrician. Figure 3.5 gives an example which satisfy Assumption B3 but not Assumption B1. Consider the observation l in the figure. Assumption B1 would require the existence of another observation that links $p^{SRF}(x_{ih}^B, d_{ih}^B) = 0.5$ and $p^{SRF}(x', d') = 0.8$ for immediate detection of bunching types. Nevertheless, Assumption B3 is satisfied as long as the observations i and h reveal the bunching of two types at (x_{ih}^B, d_{ih}^B).

With Assumption B3 replacing Assumption B1, the linking technology under bunching now cannot guarantee to recover the exact type of each observation. For example, types of i and h in Figure 3.6 are not distinguishable. Assumption B4 deals with this issue.

Assumption B4 **(No observations with two "bunched" self-reports)** Every observation i has at least one self-report elicited at a state-choice in which there is no bunching.

The following proposition establishes one of the most important results in this paper.

Proposition 7 *Under Assumptions B2,B3,B4 and SR-No Islands, the linking technology under bunching recovers the true types.*

Proof. Given Lemma 6, the critical step is to restore the identification of bunching state-choice cells under Assumption B3 (Detection of Bunching Types), which is weaker

Figure 3.5: SRFs Allowed under Assumption B3 but not B1

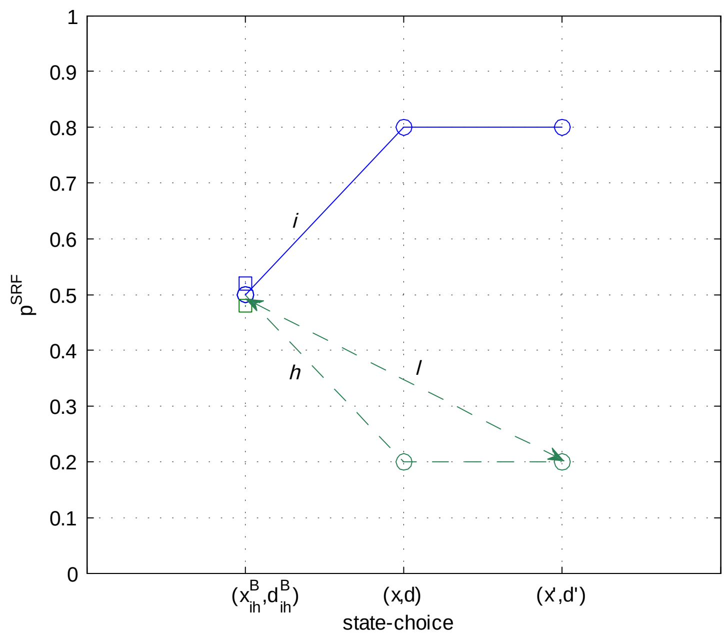


Figure 3.6: Non-identification of Types

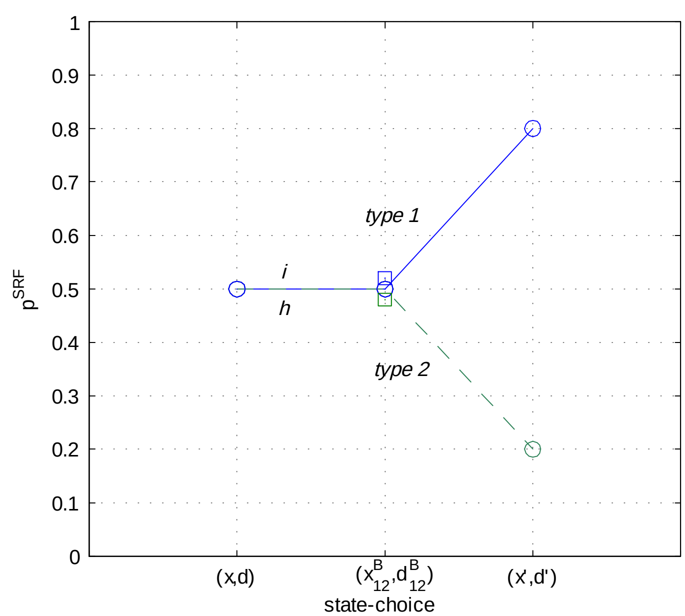
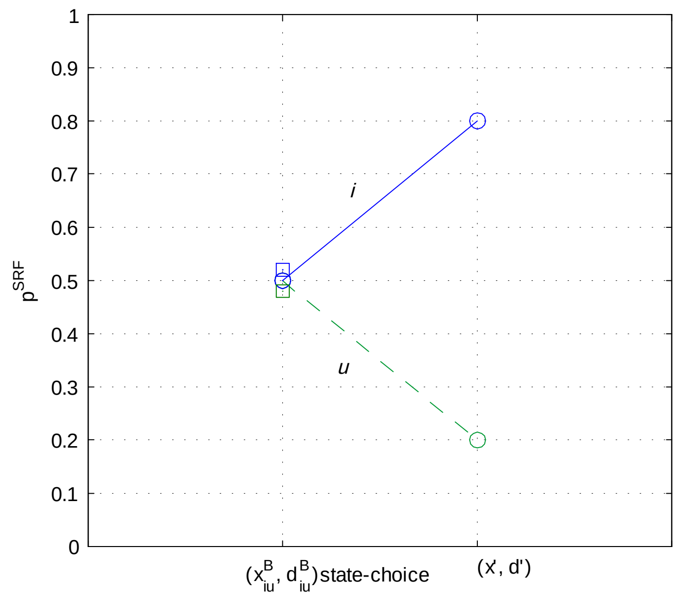


Figure 3.7: Immediate Detection of i-u Bunching



than Assumption B1 (Detection of Bunching Observations). Consider an observation i, whose SRs involve one SR in a bunching state (x, d) where her type bunches with another type. By Assumption B3, this bunching of types is detectable by two observations that do not necessarily involve i. There are thus two possibilities. One, $\exists$ an observation u, who bunches with i at (x^B_{iu}, d^B_{iu}), but differentiates itself at another state-choice (x', d'), as is depicted in Figure 3.7. Two, while i's other SR is at (x', d'), there are two observations u and v, who reveal the bunching of the types at some other state-choice (x'', d''), as is in Figure 3.9.

Now consider observations i and j, who are of the same type. We want to show that $i \sim^B j$. In the first case, by Lemma 6, we have $i \sim^B j$. In the second case, by Assumption

Figure 3.8: Detection of i-u Bunching Using v

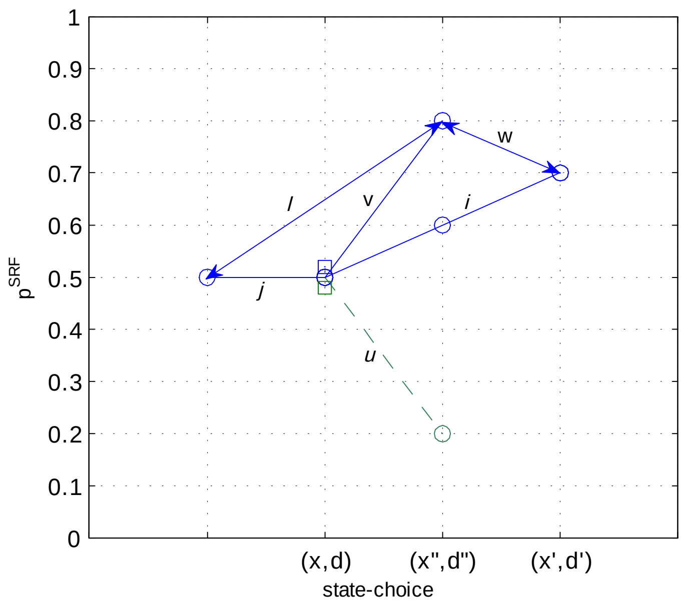


Figure 3.9:

B3, there exist two observations u and v that reveal the bunching of types at (x, d), that is, $(x, d) = (x^B_{uv}, d^B_{uv})$. By Assumption B2, there exists some observation l that bridges j and v and there exists some other observation w that bridges v and i. The linking technology under bunching gives that $j \sim^B v$ and $v \sim^B i$. By the transitivity of the equivalence relation, $j \sim^B i$.

Now comes the other direction that $j \sim^B i$ implies $k_j = k_i$. It suffices to show that $\forall m$ R^B n implies $k_m = k_n$. Suppose not. Since m R^B n, let the common state-choice cell at which m and n made a common SR be (x, d). By Assumption B3, $\exists$ two observations m' and n' and another state-choice (x', d') s.t.

$$\begin{cases} \widetilde{p}^{SRF}_{m'}(x, d) = \widetilde{p}^{SRF}_{n'}(x, d) \\ \widetilde{p}^{SRF}_{m'}(x', d') \neq \widetilde{p}^{SRF}_{n'}(x', d') \end{cases}.$$

Assumption B2 identifies through bridging that $m' \sim^B m$ and $n' \sim^B n$. Hence, $m' \sim^B n'$. Contradiction.

However, for all pairs of observations whose SRs are identical at two of their bunching state-choice cells (observations ruled out in Assumption B4), their types are not identified. Recall Figure 3.6. Assumption B3 nevertheless indicates which two types these two observations may belong to. ■

In practice, we can write a computer algorithm that implements the linking technology to determine the type of those observations whose two SRs do not bunch with those of another type simultaneously.[40]

Finally, we can relax Assumption B4. For those observations whose types are

[40] The algorithm is described in detail in a supplementary Appendix available upon request.

indeterminate, we will impute their types by finding the conditional probability of being a particular type given the observation's history of states and choices and its pair of bunching state-choices. Let i be such an observation of type k, whose SRs are $\{\widetilde{p}_i^{SRF}(x_{t'}, d_{t'}), \widetilde{p}_i^{SRF}(x_{t''}, d_{t''})\}$, where $(x_{t'}, d_{t'})$ and $(x_{t''}, d_{t''})$ are two bunching state-choice cells for types k and k'. Below we describe the procedure used to impute i's type.

First, we use the subsample where types can be correctly revealed to form a system of equations in terms of CCPs for each type and solve for the CCPs for each type. There are in general $|X| \times K$ equations and unknowns. Note that unlike the situation under SR-Precise, now even with 1-period ahead SRs the system will be non-linear. In the case of expected CCPs, the non-linearity is introduced by the double rounding-off. A typical equation of such a system will then look like

$$\begin{aligned}
& p_i^{SRF}(x_{it'}, d_{it'}, k_i) \\
= \; & p_i^{SRF}\left(d_{i,t'+1} = 1 | x_{it'}, d_{it'}, k_i\right) \\
= \; & \arg\min_{b \in B} |p_i^{SR}(d_{t+1} = 1 | x_{it}, d_{it}, k_i) - b| \\
= \; & \arg\min_{b \in B} \left\{ \left| \sum_{x_{i,t'+1}} \left[\text{FCCP}\left(x_{i,t'+1}, k_i\right)\right] f\left(x_{i,t'+1} | x_{it'}, d_{it'}\right) - b \right| \right\} \\
= \; & \arg\min_{b \in B} \left\{ \left| \sum_{x_{i,t'+1}} \left[\arg\min_{b \in B} |\Pr\left(d_{i,t'+1} = 1 | x_{i,t'+1}, k_i\right) - b|\right] f\left(x_{i,t'+1} | x_{it'}, d_{it'}\right) - b \right| \right\}
\end{aligned}$$

Note that in general the above system may not have a unique solution. Therefore we work with an approximate problem that essentially disregards the two layers of rounding-off. Given the set of focal points B, the bias introduced by the approximation will be bounded. Since the focal CCPs differ from their precise values by at most ± 0.025 at each rounding-off, our procedure leads to a bounded bias of ± 0.05

Once we solve the above system, we compute the conditional probability of i being type k given i's history of choices and states for every "problematic" observation i (i.e. every observation whose pair of self-reports does not provide enough information to uncover its type).

To that end note that for problematic observations we have

$$\begin{aligned}\Pr\left(k\left|\{x_t,d_t\}_{t\neq t',t''}\right.\right) &= \frac{\Pr\left(\{x_t,d_t\}_{t\neq t',t''}\left|k\right.\right)\Pr(k)}{\Pr\left(\{x_t,d_t\}_{t\neq t',t''}\right)} \\ &= \frac{\Pr\left(\{x_t,d_t\}_{t\neq t',t''}\left|k\right.\right)\Pr(k)}{\sum_{k'=1}^{K}\Pr\left(\{x_t,d_t\}_{t\neq t',t''}\left|k'\right.\right)\Pr(k')}\end{aligned}$$

In the RHS, $\Pr\left(\{x_t,d_t\}_{t\neq t',t''}\left|k\right.\right)$ can be computed using type k's CCPs and the estimates of the transition probabilities of the states. $\Pr(k)$ is estimated using, for example, the following equation

$$\begin{aligned}\Pr(d_t=1|x_t=5) &= \Pr(d_t=1|x_t=5,k=1)\Pr(k=1) \\ &\quad +\Pr(d_t=1|x_t=5,k=2)\left[1-\Pr(k=1)\right]\end{aligned}$$

where $\Pr(d_t=1|x_t=5)$ is estimated by simple frequency from the data and $\Pr(d_t=1|x_t=5,k=1)$ and $\Pr(d_t=1|x_t=5,k=2)$ are computed using the type specific CCPs for each type. Given that obtaining such CCPs is not feasible, we work with approximate CCPs which solve the approximate system of equations described above.[41] Among all those problematic observations who have the same SRs and the same

[41] Alternatively, the denominator in the RHS, $\Pr\left(\{x_t,d_t\}_{t\neq t',t''}\right)$, could be obtained by counting the

remaining history for $t \neq t', t''$ as $i's$. We then assign their types such that with probability $p(k|\{x_t, d_t\}_{t\neq t',t''})$, they are of type k.[42]

3.4 Montecarlo Experiments

In this section we do not discuss the precise data case because its empirical implementation is less feasible given that most subjective assessments of future choice probabilities have focal measurement error. We instead focus on the more realistic, empirically relevant, case in which there is focal measurement error in SR-CPs. We analyze two cases : a) a case in which this particular form of noise in the self-reports is innocuous and b) the more general case in which it leads to bunching.

Consider the model in the machine replacement example of Section 3.2. Again, note that we purposefully work with a simple toy model to be able to assess timing gains relative to a full-solution approach. However the method works equally well if we have a realistic state space that prevents estimation via full-solution. True, when the state space gets large it is likely that we will run into a "Data Curse of Dimensionality" in the sense that we will not have enough data to estimate the first-stage CCPs non-parametrically, even if we do not condition on type. This is not a limitation of our method, but one shared with the original Hotz-Miller (1993) estimator. However, there exist well known generalizations of the original Hotz-Miller strategy that preserve the initial insight while at the same time solving the "Data Curse of Dimensionality". For example, after unraveling the types we could use the estimator advanced by Hotz, Miller, Sanders and Smith (1994)

proportion of observations who have this particular history of states and choices.

[42] This procedure can be readily extended to the case where there are more than two types bunching at the state-choice cells at the time of the SRs.

that combines the alternative representation of the value function with a forward path simulation approach to greatly diminish the data requirements of the original Hotz-Miller strategy.[43]

We consider the simplest case in which there are $K = 2$ types. We simulate data on $N = 100,000$ firms and $T = 10$ periods using that model as underlying DGP with the following parameters:

$$\begin{array}{ll} \text{Type 1:} & (\theta_{11}, R_1) = (-0.4, -3) \\ \text{Type 2:} & (\theta_{12}, R_2) = (-1.2, -7) \end{array}$$

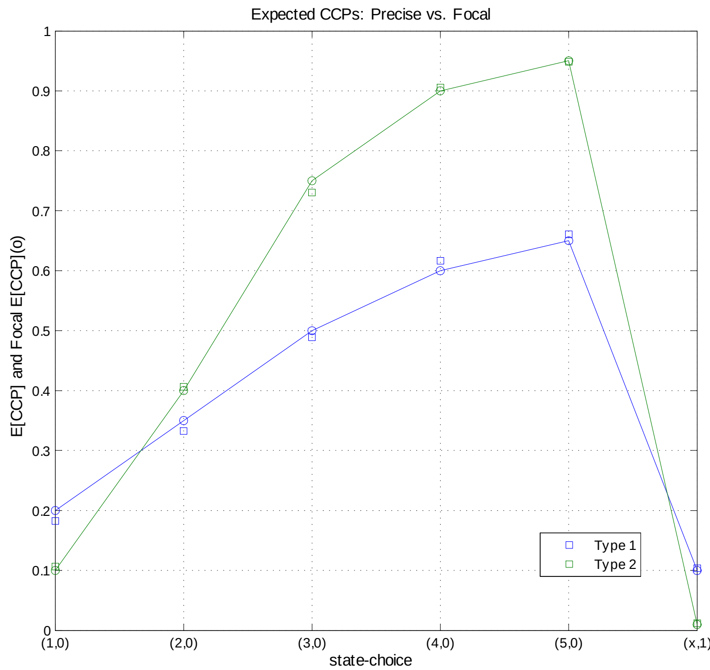
We generate inputs to the simulated self-reports using our theory of self-report and further round off self-reported choice probabilities on the simulated elicitation according to the focal points described Section $3.3.2$. In Figure 3.10 we see that despite the measurement error induced by focal self-reports, no type-bunching occurs. The squares point to the location of the precise E[CCP]s, the ones that would be elicited in the ideal case without "heaping" in focal values. The circles show the corresponding "focal" E[CCP]s

Since no type bunching occurs, the linking technology quickly establishes the number of types and type membership, and Hotz-Miller proceeds with type as an extra state variable. Table 1 describes the results of the Montecarlo simulations and illustrates that our linking technology allows quick and precise estimation of the unobserved heterogeneity in the structural model.[44]The mean estimate over the R=500 repetitions is virtually the same

43 Only states visited with positive probability in the sample at hand (as oposed to all feasible states conceptually possible in the model) are used in this estimation strategy.

44 Convergence of the entire algorithm takes on average aproximately half a minute. Almost all of the time is

Figure 3.10: Focal Self-Reports That Do Not Lead to Bunching



as the truth. The standard deviation of the Montecarlo distribution is very small.[45]

Table 3.1

	Truth	Full Solution		Hotz-Miller	
		Mean	SD	Mean	SD
θ_{11}	-0.4	-0.4000	0.0058	-0.3999	0.0058
R_1	-3.0	-2.9997	0.0198	-2.9995	0.0198
θ_{12}	-1.2	-1.2012	0.0269	-1.2008	0.0268
R_2	-7.0	-7.0071	0.0951	-7.0057	0.0949
Avg. Time	-	11 minutes		30 seconds	

We now modify our DGP to generate a more complex situation. The parameters are now:

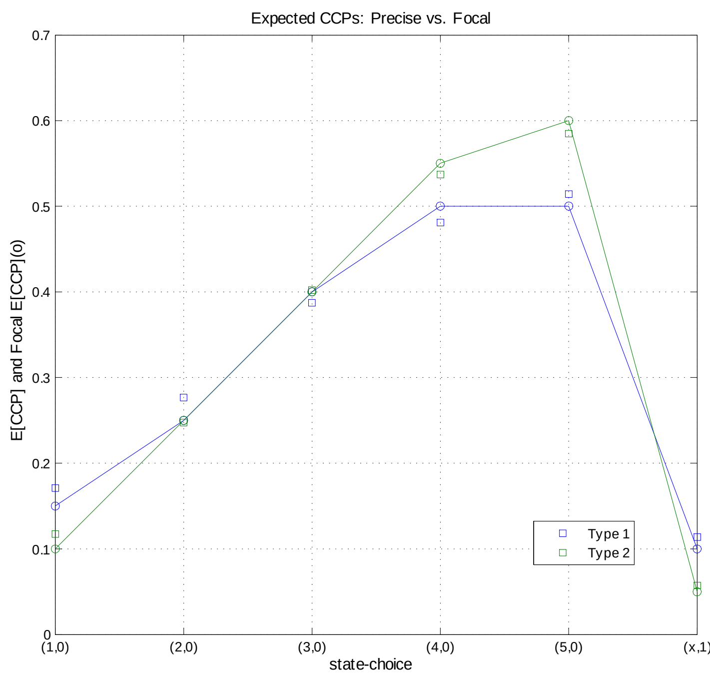
spent in the Hotz-Miller step. Indeed, preliminary type revelation and linking only takes about half a second. The montecarlo was run in a standard desktop using MATLAB.

45 Standard Deviations for the montecarlo distribution of estimates are computed for each parameter as follows: $\sqrt{\frac{1}{R}\sum_{r=1}^{R}(\theta_r - \overline{\theta})^2}$ where $\overline{\theta} = \frac{1}{R}\sum_{r=1}^{R}\theta_r$

$$\begin{array}{ll} \text{Type 1:} & (\theta_{11}, R_1) = (-0.27, -2.65) \\ \text{Type 2:} & (\theta_{12}, R_2) = (-0.40, -3.75) \end{array}$$

In Figure 3.11 we can see that focal self-reports now lead to bunching in state choice-combinations $(x, d) = (2, 0)$ and $(x, d) = (3, 0)$. Again, the squares point to the location of the precise E[CCP]s. The nearby circles show the corresponding "focal" E[CCP]s that respondents actually provide.

Figure 3.11: Focal Self-Reports That Lead to Bunching



We consider four estimation strategies for this case. In Table 2 we show the Montecarlo results for each of these.

(1) **Discarded:** In this strategy, we just drop from the sample those observations whose type cannot be determined. Column 2 shows the mean estimates. While the maintenance costs, θ_1 are estimated very precisely for both types, there is a small bias in the estimates of replacement costs R_1 and R_2. In both cases we tend to underestimate replacement costs. This makes sense. Since the two bunching

state choice combinations $(2,0)$ and $(3,0)$ involve non-replacement decisions, when we discard observations we tend to disproportionately eliminate from the sample observations that do not replace machines. Therefore the sample becomes more dominated by observations that do replace machines. The structural parameter estimates rationalize this behavior in the data by making machine replacement decisions less costly than they really are.

(2) **Infeasible A:** In this case we pretend we know each observation's type. Then we estimate $p\left(k\left|\{(x_t,d_t)\}_{t=t',t''},\{x_t,d_t\}_{t\neq t',t''}\right.\right)$ by simple frequency and assign types to "problematic" observations such that they (as a group) are consistent with this estimated probability. Here we are back to the scenario of our first Montecarlo without bunching. Not surprisingly the performance is excellent.

(3) **Infeasible B:** Here we no longer pretend we know each observation's type but instead claim we know the precise CCPs. Then we compute $p\left(k\left|\{(x_t,d_t)\}_{t=t',t''},\{x_t,d_t\}_{t\neq t',t''}\right.\right)$ using the Bayesian update described above and again assign types to "problematic" observations.[46] Again results are extremely good.

(4) **Feasible:** Our feasible estimation strategy follows the same protocol as Infeasible B, but now using the approximate type-specific CCPs derived from the approximate system of equations based on focal E[CCP]s. The performance here is also excellent and virtually the same as the one achieved by Infeasible B, which uses the (usually unavailable) precise CCPs.

Table 3.2

	Truth	"Discarded"		Infeasible A		Infeasible B		Feasible	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
θ_{11}	-0.27	-0.2735	0.0016	-0.2707	0.0016	-0.2713	0.0016	-0.2709	0.0015
R_1	-2.65	-2.6227	0.0088	-2.6538	0.0085	-2.6566	0.0085	-2.6561	0.0082
θ_{12}	-0.40	-0.4006	0.0024	-0.3985	0.0022	-0.3968	0.0022	-0.3977	0.0022
R_2	-3.75	-3.7024	0.0138	-3.7414	0.0134	-3.7304	0.0131	-3.7329	0.0134
t	-	24.1 seconds		27.4 seconds		30.2 seconds		76.6 seconds	

3.5 Extensions

We first consider in some detail three important extensions in the ideal case in which we have self reports that are precisely elicited. We then briefly outline other directions for future research.

[46] In the actual implementation there is a trade-off when choosing how much information to condition on when doing the Bayesian update. If we condition on all the history $\{x_t,d_t\}_{t\neq t',t''}$,the number of observation in each cell might be very small so in practice it might be better to condition on a subset of the available history.

3.5.1 S-periods ahead Self-Reported Cumulative Choice Probability

In some cases the question explicitly specifies a longer planning horizon and the elicited subjective probability then refers to the probability of the action being taken at some point during the given planning horizon. In this case the linking technology still works and types can be revealed in the same fashion as in the 1-period ahead case. Then we proceed via standard Hotz-Miller using the revealed types as an extra observed state in the first stage.

Still, we can attempt to recover the underlying type-specific CCPs. Consider the 2-periods ahead SR-Cumulative CP. Then we have a nonlinear equation given by

$$
\begin{aligned}
& p^{SR}\left(d=1 \text{ at some point during the next two periods}\right) \\
= \;& \Pr\left(d_{t+1}=1 \cup d_{t+2}=1 | x_t, d_{t,} k\right) \\
= \;& 1-\Pr\left(d_{t+1}=0 \cap d_{t+2}=0 | x_t, d_{t,} k\right) \\
= \;& 1-\sum_{x_{t+1}}\left[\Pr\left(d_{t+1}=0 \cap d_{t+2}=0 | x_{t+1,} d_{t,} k\right)\right] f\left(x_{t+1} | x_t, d_t, k\right) \\
= \;& 1-\sum_{x_{t+1}}\left[\Pr\left(d_{t+2}=0 | x_{t+1}, d_{t+1}=0, k\right) \Pr\left(d_{t+1}=0 | x_{t+1}, k\right)\right] f\left(x_{t+1} | x_t, d_t, k\right) \\
= \;& 1-\sum_{x_{t+1}}\left[\xi_{t+1}\right] f\left(d_{t+1} | x_t, d_t, k\right)
\end{aligned}
$$

where

$$
\xi_{t+1}=\left(\sum_{x_{t+2}} \Pr\left(d_{t+2}=0 | x_{t+2}, k\right) f\left(x_{t+2} | x_{t+1}, d_{t+1}=0, k\right)\right) \Pr\left(d_{t+1}=0 | x_{t+1}, k\right)
$$

Again if assumption F4 holds, in this equation the $f\left(x_{t+1}|x_t, d_t\right)$ and $p^{SR}\left(\right)$ are known, whereas the $\Pr\left(d=0|x,k\right)\ \forall x$ are unknown. Here we can also link the two self-reports of the same respondent and form additional equations with other self-reports from other respondents of the same type until we have a system that can be solved.[47]

3.5.2 Continuous Distribution of Unobserved Heterogeneity

Allowing for a continuous distribution of unobserved heterogeneity in dynamic programming models is very complicated. The only computationally feasible attempt we are aware of is the importance sampling strategy proposed by Ackerberg (1999,2009). Let's assume we have the full set of precise CCPs or, alternatively, we have precise E[CCP]s at every possible state choice combination (x, d) . In our machine replacement example, this would amount to having SR-CPs requested at each and every one of the six state-choice combinations.[48] In this case we can estimate the model allowing for a nonparametric continuous distribution of unobserved heterogeneity as follows: With all the possible E[CCPs] in hand we can solve a system of equations and recover the individual specific CCPs at every possible state point. Then we can "plug-in" the individual-specific set of CCPs to compute individual specific W_i for each $i = 1, ..., N$ exploiting Hotz & Miller's alternative representation. If we redefine structural parameter heterogeneity as deviations η_i from a common mean,

$$\begin{aligned} \theta_{1i} &= \mu_{\theta_1} + \eta_{1i} \\ R_i &= \mu_R + \eta_{Ri} \end{aligned}$$

[47] Unlike the 1-period ahead case, this system of equations is non-linear (even under precise elicitation) and the computational advantage of this strategy should be evaluated on a case by case basis for each specific model. This system of non-linear equations grows with the state-space so while our basic linking technology still works, recovering the underlying CCPs directly from the self-reports becomes more computationally demanding in realistic models. Still, it should be kept in mind that this whole step needs to be done only once so one can easily afford some computational cost.

[48] Note that given the renewal structure of our model, we need only consider the following six state-choice combinations: $(1, 2), (2, 0), (3, 0), (4, 0), (5, 0)$ and $(x, 1)$
for any x. No matter what the state is, if a replacement decision is made, the state variable next period is "reset" to one with probability one.

the rest of the estimation protocol follows as in Hotz-Miller. The estimation routine searches for the mean of the structural parameters and, at each iteration, we solve out for the structural parameter deviations $\{\eta_i\}_{i=1}^N = \left\{ \begin{pmatrix} \eta_{1i} \\ \eta_{Ri} \end{pmatrix} \right\}_{i=1}^N$ that are consistent with the data (given current trial parameter vector for the mean, $\begin{pmatrix} \mu_{\theta_1} \\ \mu_R \end{pmatrix}$). We do so by picking two CCPs implied by the self-reports ($\text{CCP}_i^{SR}(x_{it})$) and solving the linear system based on

$$Z\eta_i = E$$

where

$$\begin{aligned} Z &= \left(\widetilde{z}^{\widehat{P}}(1, x_{it}) - \widetilde{z}^{\widehat{P}}(0, x_{it})\right) \\ E &= -\log\left(\frac{1}{\text{CCP}_i^{SR}(x_{it})} - 1\right) - \widetilde{e}^{\widehat{P}}(1, x_{it}) + \widetilde{e}^{\widehat{P}}(0, x_{it}) - \widetilde{z}^{\widehat{P}}(1, x_{it})\,\theta_u + \widetilde{z}^{\widehat{P}}(0, x_{it})\,\theta_u \end{aligned}$$

which can be derived by noting that[49]

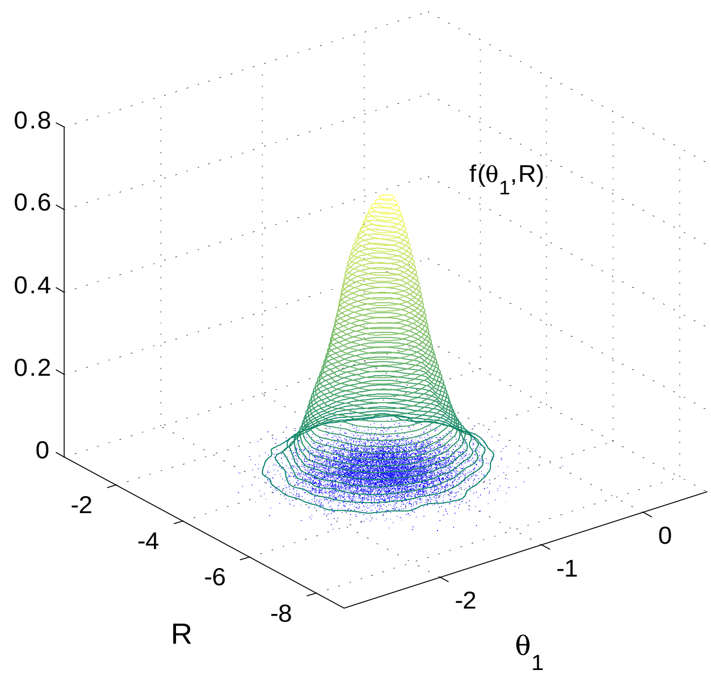
$$\text{CCP}_i^{SR}(x_{it}) = \frac{\exp\left\{\widetilde{z}^{\widehat{P}}(1, x_{it})(\theta_u + \eta_i) + \widetilde{e}^{\widehat{P}}(1, x_{it})\right\}}{\exp\left\{\widetilde{z}^{\widehat{P}}(0, x_{it})(\theta_u + \eta_i) + \widetilde{e}^{\widehat{P}}(0, x_{it})\right\} + \exp\left\{\widetilde{z}^{\widehat{P}}(1, x_{it})(\theta_u + \eta_i) + \widetilde{e}^{\widehat{P}}(1, x_{it})\right\}}$$

After convergence we can "plot" the nonparametric joint distribution of the structural parameters and recovers its moments.

---

[49] If current utility is given by $U(d, x_{it}, \varepsilon_{it}) = u(d, x_{it}) + \varepsilon_{it}(d)$ and we let $\theta = (\theta_u, \theta_f)$ be the vector of structural parameters. Consider a linear-in-parameters utility $u(d, x_{it}) = z(d, x_{it})'\,\theta_u$ where $z(d, x_{it})$ is a Dim(θ_u) $\times$ 1 vector. State variables evolve according to $f_x(x_{t+1}|d_t, x_t, \theta_f)$. The choice specific value functions can be re-written as $v(d, x, \theta) = \widetilde{z}(d, x, \theta)\,\theta_u + \widetilde{e}(d, x, \theta)$ where $\widetilde{z}(d, x_t, \theta) = z(d, x) + \sum_{s=1}^{T-t} \beta^s E_{x_{t+s}|d_t=d, x_t}\left[\sum_{d'=0}^{D} P(d'|x_{t+s}, \theta)\, z(d', x_{t+s})\right]$ and $\widetilde{e}(d, x_t, \theta) = \sum_{s=1}^{T-t} \beta^j E_{x_{t+s}|d_t=d, x_t}\left[\sum_{d'=0}^{D} P(d'|x_{t+s}, \theta)\, e(d', x_{t+s})\right]$. The policy function is $\alpha(x_t, \varepsilon_t) = \arg\max_d \{v(d, x_t, \theta) + \varepsilon_t(d)\}$ and the expected error conditional on optimal choice is $e(d, x_t) = E[\varepsilon_t(d)\,|x_t, \alpha(x_t, \varepsilon_t) = d]$ .Hotz & Miller (1993) show that $e(d, x_t) = f(d, P(\cdot|x_t, \theta), G_\varepsilon)$ .For example, if $\varepsilon_{it}(d)$ are iid Extreme Value, then $e(d, x_t) = \gamma - \log[P(d|x_{it}, \theta)]$

Figure 3.12: Figure 3.11 Distribution of Structural Parameter Population Heterogeneity



For example, consider the following DGP.

$$\begin{pmatrix} \theta_{1i} \\ R_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\theta_1} \\ \mu_R \end{pmatrix}, \Sigma\right)$$

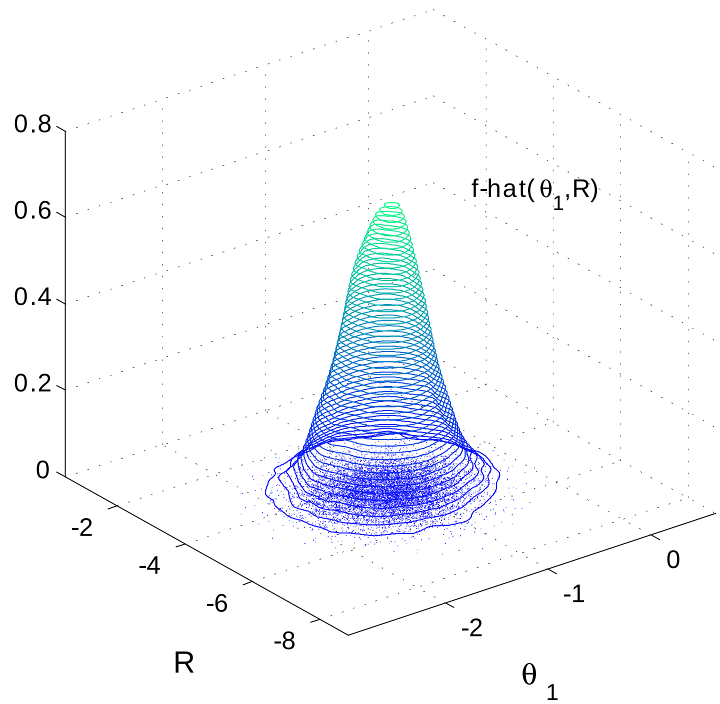
with true parameter values given by $\mu_{\theta_1} = -1.15$, $\mu_R = -4.45, \Sigma = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.5 \end{bmatrix}$.
Figure 3.12 depicts the true distribution of structural parameter values in the population.

Our estimates are obtained in less than one minute. The estimated population means for both structural parameters are

$$\begin{aligned} \widehat{E[\theta_1]} &= -1.1496 \\ \widehat{E[R]} &= -4.4451 \end{aligned}$$

which are almost exactly equal to the truth.

Figure 3.13: Figure 3.12 Estimated Distribution of Structural Parameter Heterogeneity



The estimated variance covariance matrix for the distribution of structural parameter heterogeneity in the population is

$$\widehat{Var}\left(\theta_1, R\right) = \left[\begin{array}{cc} 0.1002 & 0.0028 \\ 0.0028 & 0.5043 \end{array}\right]$$

which is consistent with the underlying bivariate normal DGP. Note however, that our approach is non-parametric in the sense that we could have recovered any kind of distribution since at no point did we use normality, other than to simulate the data.

Figure 3.13 presents the plot for the non-parametric estimate of the distribution of structural parameters in the population.

Note that in more complex models, having the entire set of E[CCPs] might be unrealistic. However, if we have three self-reports, we can use a variant of the linking technology that supplements our original linking strategy with an interpolation step. Indeed, if we assume Finite Dependence[50], Modal CCP self-reports and Assumption F4, then we can recover the continuous distribution of unobserved heterogeneity in structural parameters by interpolating CCPs using nearest neighbors to complete each individual's full set of self-reports.[51]

3.5.3 Time-Varying Unobserved Heterogeneity

As we discussed in the Introduction, recent efforts, to which our paper contributes, show that under some conditions, it is possible to accommodate permanent unobserved heterogeneity in two-step estimation strategies. The work of Arcidiacono & Miller (2008) pushes the frontier forward by not only allowing for unobserved heterogeneity but by letting it evolve in systematic ways over time.

In this subsection we briefly note that our linking approach can be modified to accommodate unobserved heterogeneity that evolves over time. To continue with our machine replacement example, we now think of firms as being in one of two possible unobserved states. These unobserved states are not permanent, but rather can change over time. We focus on cases in which we have two self-reports taken in consecutive periods.

In this scenario k is no longer fixed but becomes k_t, a random variable that exogenously evolves over time as in assumption F2. The key idea can be grasped with $K = 2$ (i.e. there

[50] The concept of finite dependence was originally developed by Altug & Miller (1998) and further generalized by Arcidiacono & Miller (2008). It sharpens the insight of the Hotz-Miller original result by showing that for certain class of models only the 1-period (or, in general for finite $\rho > 0$, the $\rho-$period) ahead CCPs are needed in the alternative representation of the value functions.

[51] Nearest neighbors are those observations that report similar SR-CPs in two of the same state-choice cells.

are two possible unobserved states $k = 1, 2$). Suppose that at time t' the first self-report is collected. Among those with the same $(x_{t'}, d_{t'})$ we can identify those who give different SR-CPs and, following the reasoning of previous sections, those who have different $k_{t'}$. Without loss of generality we can assign one group to $k_{t'} = 1$ and the others to $k_{t'} = 2$. We can track this group into the next period. Suppose that next period, $t' + 1$, the second self-report is collected. We can see how the answers of each group split at the time of the second self-report. These splits give information on unobserved state transitions π_{jl} for $j, l = 1, 2$.

How do we know which % transitioned into which state ? We can compare self-reports in the second period across different states and check against the self-reports collected at those same states in the first period as long as the model is stationary. We are then able to identify those that remained in their previous unobserved state and those who transitioned into a new one.

Finally, note that all of the above can be generalized to: a) any $K > 2$, b) self-reports collected in any two, not necessarily consecutive time periods (t', t''), c) cases in which the first period collecting self-reports, t', is not the first sample period and d) models with choice-dependent transitions for unobserved states.[52]

[52] Choice dependent transitions for unobserved states accomodate the following case: instead of firms differing in k_t, machine differences are the underlying source of unobserved heterogeneity. Suppose when a firm replaces an old machine, the new machine may turn out to be an "easy maintenance, easy replacement" machine or a "problem" machine which is difficult to maintain and difficult to replace. Here the unobserved state may evolve over time but only if the renewal action is taken. To handle this case we relax Assumption F2.

3.5.4 Other Extensions

(1) In some special cases we can entertain the possibility of estimating the model off SR-CPs alone, rather than using actual revealed choice data.

(2) Using Other Probability Questions:

- As seen in Section 3.5.1, sometimes we have SR-Cumulative CPs like "What's the probability that you will choose $d = 1$, at least once, at some point during the next S years?". In some cases, we may have two of these questions at the same time t', eliciting the cumulative probability that an action is taken or a state is reached within two or more time horizons. We may have, for example,
$$\left\{ \begin{pmatrix} p_i^{SR}\left(\cup_{s=1}^{5}\left\{d_{t+s}=1\right\}|x_t,d_t,k\right) \\ p_i^{SR}\left(\cup_{s=1}^{10}\left\{d_{t+s}=1\right\}|x_t,d_t,k\right) \end{pmatrix} \right\}_{i=1}^{N}$$
eliciting the 5- and 10-period ahead cumulative CP. This eliminates the need of a panel of self-reports.
- "What's the probability that you will have $x_{t+S} = x$ in S years?"

(3) Using Other Types of Expectations: Some questions don't ask about the probability of making a given choice or reaching a given state but rather ask whether the respondent "expects" to make that choice or reach a given point in the state space.

- "Do you expect to have $d_{t+S} = d$ in S years?" This case which severely limits the informational content of the self-report arguably asks whether $\Pr\left(d_{t+S} = d|x_t, d_t, k\right) > 50\%$
- "What value of x do you expect to end up having over your planning horizon?" In these cases x may refer to the number of children or completed years of education that a person will have over their remaining lifetime, i.e. over the next $T - t$ years. Here, respondents could arguably be rounding-off the expected maximum, $E\left[x_T|x_t, d_t, k\right]$ to the nearest integer or providing the mode, $Mode\left[x_T|x_t, d_t, k\right] = \arg\max_{x_T}\left\{f\left(x_T|x_t, d_t, k\right)\right\}$

(4) Application to multiple agent models.

(5) Multinomial choice and use of GEV unobservables following McFadden (1978) , Breshnahan, Stern & Trajtenberg (1997), Arcidiacono (2005) and Arcidiacono & Miller (2008).

3.6 Conclusions

We have introduced a new approach to allow for unobserved heterogeneity in two-step, CCP-based estimation strategies for discrete choice dynamic programming models such as those pioneered by Hotz & Miller(1993). Our strategy exploits the availability of expectations data. Since subjective expectations data about future choice probabilities integrate the future temporary idiosyncratic shocks, they are extremely powerful and they become a valuable resource to identify and estimate unobserved heterogeneity. We believe that if and when such data is available, our approach should be attractive given that identification requires mild assumptions and estimation can proceed with very light data. Indeed, the method can be implemented with only two unconditional self-reports about future choice probabilities per respondent. Our Montecarlo experiments show that computational burden is essentially the same as that of the (already fast) original Hotz-Miller estimator. The method can be applied in combination with variants of the original Hotz-Miller estimator that reduce its onerous data requirements in models with rich state spaces. While our focus has been on single agent models of dynamic discrete choice, we believe that our approach can be generalized to the many other contexts discussed in the introduction as long as subjective expectations data is available to supplement traditional data on observed choices and states. We leave these and other extensions for future research. We believe this is a first step in a fruitful research program that leverages new forms of available data to be more flexible about the specification of unobserved heterogeneity in structural estimation.

3.6 Appendix: Proofs of Lemmas

Lemma 1 Under Assumption SR-Precise and SR-No Islands, the linking technology recovers the true number of types and type membership for each observation.

Proof. First, we establish that under Assumption SR-Precise, the linking technology implies that for all $i \sim j$, $k_i = k_j$. By definition of the linking technology, $i \sim j$ iff $\exists$ a possibly empty subset of observations $\{i_1, ...i_n\} \subseteq I$, such that $i\ R\ i_1\ R...R\ i_n\ R\ j$. By transitivity of "$=$", it is enough to show that $\forall m\ R\ n$, $k_m = k_n$. By definition, $m\ R\ n$ iff

$$\{\widetilde{p}_m^{SR}(x_{mt'}, d_{mt'}), \widetilde{p}_m^{SR}(x_{mt"}, d_{mt"})\} \cap \{\widetilde{p}_n^{SR}(x_{nt'}, d_{nt'}), \widetilde{p}_n^{SR}(x_{nt"}, d_{nt"})\} \neq \emptyset$$

Since under Assumption SR-Precise the probability of two different types having exactly the same report on the same state-choice cell is zero, it holds that $m\ R\ n \Rightarrow k_m = k_n$. Hence, $k_i = k_{i_1} = ... = k_{i_n} = k_j$.

Second, we need to show that $\forall i, j$ with $k_i = k_j$, $i \sim j$. Suppose not. Consider observations i, j, who are of type k, but $i \nsim j$. Let $\Sigma^{[i]}$ ($\Sigma^{[j]}$) be the set of all state-choice cells at which the equivalent class $[i]$ ($[j]$) gives SRs. Then, $\Sigma^{[i]} \cap \Sigma^{[j]} = \emptyset$. Otherwise, $\exists (x, d) \in \Sigma^{[i]} \cap \Sigma^{[j]}$ and $\exists$ observations $i' \in [i], j' \in [j]$, who share a common SR at (x, d) and this implies that $i'\ R\ j'$. Hence, $i \sim i'\ R\ j' \sim j$, contradicting our assumption that $i \nsim j$. So $\Sigma^{[i]} \cap \Sigma^{[j]} = \emptyset$. But it further contradicts Assumption SR-No Islands, because $\Sigma^{[i]} \cap \Sigma^{[j]} = \emptyset$, together with $i \nsim j$, implies that $\forall (x, d) \in \Sigma^{[i]} \subseteq \Sigma^k, (x', d') \in \Sigma^{[j]} \subseteq \Sigma^k$, there does not exist two observations m and n, such that m gives a SR at (x, d) and n at (x', d') and $m \sim n$. ■

Lemma 2 Under Assumptions B1, B2 and SR-No Islands, the linking technology under bunching recovers the true types exactly.

Proof. First, we want to prove that under Assumption B1, the linking technology under bunching implies that for all $i \sim^B j$, $k_i = k_j$. Definition of linking under bunching gives $i \sim^B j$ iff $\exists$ a possibly empty subset of observations $\{i_1, ...i_n\} \subseteq I$, such that $i\ R^B\ i_1\ R^B...R^B\ i_n\ R^B\ j$. By transitivity of "$=$", it is enough to show that $\forall m\ R^B\ n$, $k_m = k_n$.

Consider such $m\ R^B\ n$. They must satisfy

(1) $\{\widetilde{p}_m^{SRF}(x_{mt'}, d_{mt'}), \widetilde{p}_m^{SRF}(x_{mt''}, d_{mt''})\} \cap \{\widetilde{p}_n^{SRF}(x_{nt'}, d_{nt'}), \widetilde{p}_n^{SRF}(x_{nt''}, d_{nt''})\} \neq \emptyset$,
(2) if $\exists$ observation h,
$\{\widetilde{p}_m^{SRF}(x_{mt'}, d_{mt'}), \widetilde{p}_m^{SRF}(x_{mt''}, d_{mt''})\} \cap \{\widetilde{p}_n^{SRF}(x_{nt'}, d_{nt'}), \widetilde{p}_n^{SRF}(x_{nt''}, d_{nt''})\} = \{\widetilde{p}_m^{SRF}(x_{mh}^B, d_{mh}^B)\}$
then

$$\begin{aligned} & \exists l, \{\widetilde{p}_l^{SRF}(x_{lt'}, d_{lt'}), \widetilde{p}_l^{SRF}(x_{lt''}, d_{lt''})\} \\ = \; & \{\widetilde{p}_m^{SRF}(x_{mt'}, d_{mt'}), \widetilde{p}_m^{SRF}(x_{mt''}, d_{mt''})\} \triangle \{\widetilde{p}_n^{SRF}(x_{nt'}, d_{nt'}), \widetilde{p}_n^{SRF}(x_{nt''}, d_{nt''})\}, \end{aligned}$$

where $\triangle$ denotes the set difference.

Proceed by contradiction. Suppose that $k_m \neq k_n$. Then by Assumption B1, m and n bunching at the state-choice cell of their common SR is immediately detected: $\{\widetilde{p}_m^{SRF}(x_{mt'}, d_{mt'}), \widetilde{p}_m^{SRF}(x_{mt''}, d_{mt''})\} \cap \{\widetilde{p}_n^{SRF}(x_{nt'}, d_{nt'}), \widetilde{p}_n^{SRF}(x_{nt''}, d_{nt''})\} = \{\widetilde{p}_m^{SRF}(x_{mn}^B, d_{mn}^B)\}$. In this case, n qualifies as the observation h in the second condition, so

$$\begin{aligned} & \exists l, \{\widetilde{p}_l^{SRF}(x_{lt'}, d_{lt'}), \widetilde{p}_l^{SRF}(x_{lt''}, d_{lt''})\} \\ = \; & \{\widetilde{p}_m^{SRF}(x_{mt'}, d_{mt'}), \widetilde{p}_m^{SRF}(x_{mt''}, d_{mt''})\} \triangle \{\widetilde{p}_n^{SRF}(x_{nt'}, d_{nt'}), \widetilde{p}_n^{SRF}(x_{nt''}, d_{nt''})\}, \end{aligned}$$

contradiction to the non-existence of such an observation who reports two different probabilities at one state-choice cell. Hence, $k_m = k_n$ and $k_i = k_{i_1} = k_{i_n} = k_j$.

Second, to show that for any pair of observations of the same type, they must belong to the same equivalence class, we proceed by contradiction. Consider two observations i and j of the same type k, but $i \nsim^B j$. Define $\Sigma^{[i]}, \Sigma^{[j]}$ as in the proof of Lemma 1. Now for any $(x, d) \in \Sigma^{[i]} \subseteq \Sigma^k$ and any $(x', d') \in \Sigma^{[j]} \subseteq \Sigma^k$, by Assumption SR-No Islands, $\exists$ two observations m and n of type k, with m reporting at (x, d) and n at (x', d'), and $m \sim^B n$. Since $(x, d) \in \Sigma^{[i]}$, $\exists$ observation $i' \in [i]$ who reports at (x, d) and some other $(x'', d'') \in \Sigma^{[i]}$. Obviously, i' and m share a common SR at (x, d). If (x, d) is a bunching state-choice cell, Assumption B1 immediately identifies this and Assumption B2 makes sure that $\exists$ an observation l that bridges i' and $m's$ non-bunching SRs. Linking technology under bunching implies $i' \sim^B m$. If (x, d) is not a bunching state-choice cell, the linking technology under bunching directly gives $i' \sim^B m$. By the same argument, $\exists j' \in [j]$ such that $j' \sim^B n$. Therefore, $i \sim^B i' \sim^B m \sim^B n \sim^B j' \sim^B j$. Therefore, $[i] \cap [j] = \{i', j', m, n\} \neq \emptyset$. A contradiction to the definition of equivalence class. ■

Ackerberg (1999) Importance Sampling and the Method of Simulated Moments, mimeo.

Ackerberg (2009) A New Use of Importance Sampling to Reduce Computational Burden In Simulation Estimation, Quantitative Marketing and Economics, forthcoming

Arcidiacono (2005), Affirmative Action in Higher Education: How Do Admission and Financial Aid Rules Affect Future Earnings?, Econometrica, Vol. 73, No. 5, September, 1477–1524

Arcidiacono, P., A. Khwaja and L. Ouyang (2007) Habit persistence and teen sex: Could increased access to contraception have unintended consequences for Teen pregnancies? mimeo. Duke University.

Arcidiacono & Miller (2008), CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity, mimeo.

Aguirregabiria & Mira (2009), Dynamic Discrete Choice Structural Models: A Survey. Journal of Econometrics.

Aguirregabiria & Mira (2002), Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models, Econometrica

Aguirregabiria & Mira (2007), Sequential Estimation of Dynamic Discrete Games, Econometrica

Aguirregabiria, Mira & Roman (2007), An Estimable Dynamic Model of Entry, Exit, and Growth in Oligopoly Retail Markets, American Economic Review

Altug & Miller (1998), The Effect of Work Experience on Female Wages and Labor Supply, Review of Economic Studies

Attanasio, O. P. (2009): Expectations and Perceptions in Developing Countries: Their Measurement and Their Use" American Economic Review, Papers and Proceedings.

Bajari, Hong, Krainer & Nekipelov (2009), Estimating Static Models of Strategic Interactions, mimeo.

Bajari, Benkard & Levin (2007), Estimating Dynamic Models of Imperfect Competition, Econometrica

Bajari, Fox, Kim & Ryan (2009), A Simple Nonparametric Estimator for the Distribution of Random Coefficients, mimeo.

Blass, Lach & Manski (2010), "Using Elicited Choice Probabilitites to Estimate Random Utility Models: Preferences for Electricity Reliability", International Economic Review, Vol. 51, No. 2, May

Blau, D. and D. Gilleskie (2008) The Role of Retiree Health Insurance in the Employment Behavior of Older Males. International Economic Review.

Blevins (2009) “Sequential MC Methods for Estimating Dynamic Microeconomic Models", Duke University, mimeo.

Bresnahan, Stern & Trajtenberg (1997), Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980s, RAND

Buchinsky, Hahn & Hotz (2005), Cluster Analysis: A Tool for Preliminary Structural Analysis.

Carro, J. and P. Mira (2006), A Dynamic Model of Contraceptive Choice of Spanish Couples. Journal of Applied Econometrics 21, 955-980.

Choo & Siow (2005): Dynamic Two-Sided Matching Games, mimeo

Dube, Fox and Su (2009), "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation" mimeo

Golan & Levy-Gayle (2008), Estimating a Dynamic Adverse-Selection Model: Labor-Force Experience and the Changing Gender Earnings Gap 1968-93, mimeo Carnegie Mellon University

Eckstein and Wolpin (1989) "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey" Journal of Human Resources, Fall, 24, 4

Eckstein and Wolpin (1999) "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities" Econometrica, 67(6): 1295-339

Heckman, J., and B. Singer, (1984) A method for minimizing the impact of distributional assumptions in economic models for duration data. Econometrica 52, 271-320

Hill, Perry and Willis (2004), Estimating Knightian Uncertainty from Survival Probability Questions on the HRS, mimeo

Hotz & Miller(1993), Conditional Choice Probabilities and the Estimation of Dynamic Models, Review of Economic Studies

Hotz, Miller, Sanders & Smith (1994), A Simulation Estimator for Dynamic Models of

Discrete Choice, Review of Economic Studies

Hu & Shum (2008) Identifying Dynamic Games with Serially-Correlated Unobservables, mimeo

Hu & Shum (2009) Nonparametric Identification of Dynamic Models with Unobserved State Variables, mimeo

Jofre-Bonet & Pesendorfer (2003), Estimation of a Dynamic Auction Game, Econometrica

Karni (2009) "A Mechanism for Eliciting Probabilities" Econometrica, Vol. 77, No. 2, March, 603–606

Kasahara & Shimotsu (2008), Pseudo-likelihood Estimation and Bootstrap Inference for Structural Discrete Markov Decision Models, Journal of Econometrics, 146(1), 92-106

Kasahara & Shimotsu (2009a), Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices, Econometrica

Kasahara & Shimotsu (2009b), Sequential Estimation of Structural Models with a Fixed Point Constraint, mimeo

Keane & Wolpin (1994) "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence" Review of Economics and Statistics

Keane & Wolpin (1997) "The Careers of Young Men" Journal of Political Economy

Keane & Wolpin (2009) "Empirical applications of discrete choice dynamic programming models" Review of Economic Dynamics

Todd & Wolpin (2009) "Structural Estimation and Policy Evaluation in Developing Countries" Annual Review of Economics, September 2010, Vol. 2

Keane, Todd & Wolpin (2010) "The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications, Handbook of Labor Economics, forthcoming.

Liu, Mroz & van der Klaauw (2009), Mother's Work, Migration, and Child Achievement, Journal of Econometrics, forthcoming.

Magnac & Thesmar (2002) "Identifying Dynamic Discrete Decision Processes", Econometrica, Vol. 70, No 2, March 801-816

McFadden (1978) " Modelling the Choice of Residential Location", in A. Karlqvist et al., eds., Spatial Interaction: Theory and Planning Models. Amsterdam. North Holland

Publishing.

Mira (2007) "Uncertain Infant Mortality, Learning and Life-Cycle Fertility". International Economic Review 48, 809-846

Pakes, Ostrovsky & Berry (2008), Simple estimators for the parameters of discrete dynamic games (with entry/exit examples), RAND

Pesendorfer & Schmidt-Dengler (2008), Asymptotic Least Squares Estimators for Dynamic Games, Review of Economic Studies

Pesendorfer & Schmidt-Dengler (2008), Sequential Estimation of Dynamic Discrete Games: A Comment, mimeo

Rust (1987), Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher, Econometrica

Siebert, R., and C. Zulehner (2008): “The Impact of Market Demand and Innovation on Market Structure” Purdue University, mimeo.

Su, Che-Lin & Judd, Kenneth (2007), "Constrained Optimization Approaches to Estimation of Structural Models", mimeo van der Klaauw (1996) Female Labor Supply and Marital Status Decisions: A Life Cycle Model, Review of Economics Studies, 1995, 63(2): 199-235.

van der Klaauw (2000) On the Use of Expectations Data in Estimating Structural Dynamic Models: An Analysis of Career Choices, mimeo, UNC-Chapel Hill

van der Klaauw & Wolpin (2008) Social Security and the Retirement and Savings Behavior of Low Income Households, Journal of Econometrics

Walker (2003), "Pregnancy and Fertility Expectations: Estimates of Bounded Rationality and Unintended Births", mimeo.

Wolpin (1984), An Estimable Dynamic Stochastic Model of Fertility and Child Mortality, Journal of Political Economy,1984, Vol. 92: 852-875

Wolpin (1999), Commentary on “Analysis of Choice Expectations in Incomplete Scenarios”, by C.F. Manski” Journal of Risk and Uncertainty

Wolpin & Gonul (1985) "On The Use of Expectations Data in Micro Surveys: The Case of Retirement", Report to the Employment and Training Administration, U. S. Department of Labor, Washington, DC, March.