CONTRIBUTIONS TO STRUCTURAL MODELING AND ESTIMATION

by

Wayne-Roy Gayle

B.Sc. in Economics and Statistics, University of the West Indies, Mona, 1996

M.Sc. in Economics, University of the West Indies, Mona, 1998

M.A. in Economics, Stony Brook University, 2001

Submitted to the Graduate Faculty of
the University of Pittsburgh in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH

ARTS AND SCIENCES

This dissertation was presented

by

Wayne-Roy Gayle

It was defended on

November 27, 2006

and approved by

Jean Francois Richard, PhD, Professor

Mehmet Caner, PhD, Associate Professor

Robert Miller, PhD, Professor

Soiliou Namoro, PhD, Assistant Professor

Holger Sieg, PhD, Associate Professor

Dissertation Director: Jean Francois Richard, PhD, Professor

CONTRIBUTIONS TO STRUCTURAL MODELING AND ESTIMATION

Wayne-Roy Gayle, PhD

University of Pittsburgh, 2006

The first chapter of my thesis develops and estimates a dynamic structural partial equilibrium model of schooling and work decisions. The estimated model explicitly accounts for the simultaneous choice of enrolling in school and working. It also allows for endogenous leisure choices, intertemporal nonseparabilities in preferences, aggregate skill specific productivity shocks, aggregate consumption price effects, and individual heterogeneity. Times spent on schooling, working, and leisure are treated as continuous choice variables. The estimated model is solved and two counterfactual simulation exercises are performed. The first is the case where a subsidy is given to individuals who enroll in school and do not participate in the labor market. The second is the case where the demands of the school curriculum are increased so that a young man enrolled in school necessarily spends more time studying. The conclusion is that the latter policy is more effective in enhancing educational achievements and future wages.

The second chapter of my thesis develops a semiparametric estimator for a dynamic nonlinear single index panel data model. Flexibility of the model is achieved by assuming that the index function is unknown. Flexibility in individual heterogeneity is achieved by assuming that the individual effect is an unknown function of some observable random variable. The paper proposes an algorithm that estimates each of the finite and infinite dimensional parameters. In particular, the full data generating process is estimated. This is important if the predicted outcomes are used as plug-in estimators, as in the multistage estimation of dynamic structural models.

The final chapter of my thesis develops a powerful new algorithm to solve single object first price auctions where bidders draw independent private values from heterogeneous distributions. The algorithm allows for the scenario in which groups of symmetric and asymmetric bidders may

collude, and for the auctioneer to set a reserve price. The paper also provides operational univariate quadratures to evaluate the probabilities of winning as well as the expected revenues for the bidders and the auctioneer. The expected revenue function is used to the compute optimal reserve under asymmetric environments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express my gratitude to Professors Hugo Benitez- Silva, Mehmet Caner, Eva Carceles- Poveda, George- Levi Gayle, Robert Miller, Holger Sieg, Soilio Namoro, and Jean-Francois Richard for numerous helpful conversations and suggestions which benefited the essays in this dissertation.

I benefited greatly from interactions with fellow graduate students. I would like to thank Wayne Archibald and Nese Yildiz for their friendship, support, discussions and spirited debates over coffee at various stages during my graduate studies. I owe special thanks to Professors Alfred Francis, Peter-John Gordon , Sung-Woo Kim and Ashu Handa Chris at the University of the West Indies, Mona for encouraging my graduate study. This thesis is particulary dedicated to the memory of Teri Waters. Without her love, patience, understanding and encouragement, this thesis would not have been possible. Finally, I would like to thank members of my family, in particular, my mother, Cutie Gayle, and my father, Fitzroy Gayle for much support and encouragement.

1.0 INTRODUCTION

This thesis simultaneously extends the literatures on econometric theory, applied microeconomics and computational economics. These extensions are guided by the increasingly recognized, yet still largely untapped complementarities between these three branches of economics. This complementarity can be explained by a simple philosophy. Developing models to analyze and solve interesting economic puzzles require understanding and appreciation of the available theoretical tools. The level of realism achieved when estimating a model with available data is constrained by the existing econometric technology and the computational feasibility of the solution. The ability to understand and compute existing estimators, and to develop new ones, therefore allows the investigator greater flexibility in thinking about and characterizing economic puzzles.

Two key economic puzzles are addressed in this thesis. The first lies essentially in the educational attainment and returns to education literature. The second lies in the literature of asymmetric first price auctions with applications to the sustainability of collusions.

The first chapter of this thesis investigates the effects of time allocation between the labor market and the classroom on educational attainment and future wages. Over the last three decades more young individuals are participating in the US labor market while actively remaining enrolled in school. Young individuals are increasing both their incidence of work, as well as the amount of hours worked while enrolled in school. This trend has generated growing interest in its possible immediate and long run effects on young individuals.

It is not obvious how working while enrolled in school affects a young individual's educational attainment and future labor market opportunities. On one hand, there is valid concern that an intensive amount of working while in school may hinder academic performance and increase drop-out rates. This is known in the literature is the crowding-out hypothesis. On the other hand, working while in school may improve a young individual's organizational skills, sense of responsibility and self esteem. This is known as the congruence hypothesis.

Along with the congruence effect other positive effects of working while enrolled in school include the human capital effect and the cash effect. The human capital effects states that working while enrolled in school gives the student immediate working experience that is directly rewarded in the labor market. The cash effect states that working while enrolled in school provides income for the student that can be used to further finance education, leading to higher educational attainment and thus increased future labor market success.

The objective of the first paper, therefore, is to disentangle these different avenues of effects of working on educational attainment and future wages. I focus on separating these different effects because their significance may vary over different groups of individuals.

The key issue then in the first paper is the optimal time allocation between time spent in the labor market, time spent on school activities, and time spent on leisure. Previous work on the effects of working on educational attainment has essentially ignored the important fact that individuals also choose the number of hours they spend on studying and other in school activities (See Eckstein and Wolpin, 1999 for example). The difficulty with such an analysis is that an hour increase in the number of time spent in the labor market is interpreted as an hour decrease in leisure. In reality, the individual could choose to reduce the time spent on school activities instead of leisure. Such analysis can therefore lead to imprecise or incorrect estimates of the effect of working on educational attainment and future labor market success.

The abstraction away from time spent on schooling is largely due to insufficient information on the time use of students. The dataset used in this study is taken from the National Longitudinal Survey of Youth (NLSY79), which is a comprehensive panel data set that follows individuals who were 14 to 21 years of age as at January 1, 1979. This dataset also contains a single wave of schooling time use data collected in 1981. This data is notorious for measurement error, and since it is a single wave, investigators have typically abstracted away from employing this data. I show that with a suitable method for controlling for measurement errors in study time and also controlling for permanent individual unobserved heterogeneity, we can indeed get precise estimates of study time. The estimation of study time is also augmented by including lagged enrollment and labor supply decisions. This helps tremendously in improving the estimates. The estimated study time function is then used to predict study time for all individuals in the sample over all applicable years. This strategy however creates a difficult econometric problem.

In a dynamic modelling framework, current decisions depend on the expectation future outcomes, including future study time. The estimation technique implemented is a modified version of the Conditional Choice Probability (CCP) estimator of Hotz and Miller [1993] and extended by Altug and Miller [1998]. In this technique, the expectation of future outcomes is captured in the probability of future decisions conditioned on the realization of future states of the world. These probabilities have to be estimated. Direct estimation as proposed by Altug and Miller [1998] would inadvertently result in conditioning on functions of the same dependent variable that is being used to estimate the probabilities. In the first paper I propose an alternative method of estimating these probabilities that avoid this severe endogeneity problem.

Another technical contribution of the first paper is the estimation method and corrected standard errors. The estimation method combines an efficient iterated GMM (GMMI) with a variation of the Nested Pseudo Likelihood Algorithm (NPL) proposed by Aguirregabiria and Mira [2002]. This method of estimation results is improved small sample properties of the estimator coming from both steps. Hansen et al. [1996] shows that iterating over the optimal weighting matrix results in improved small sample properties of the estimates. Iterating over the CCP's eliminates the initial nonparametric estimates and hence also improves the small sample properties of the estimator.

This process however results in nonstandard correct standard errors of the estimates. I propose an alternative representation of the corrected standard errors to that of Altug and Miller [1998] by employing a technique developed in Newey and McFadden [1994], and Newey [1994]. Furthermore, the structure of the state space implies that the law of iterated expectations can be used to greatly simplify the form of the standard errors. In particular, no post estimation is required to compute these standard errors. This greatly reduces the computation burden of the CCP estimator.

The estimated results indicate that crowding-out effect outweighs the positive effects for whites while the congruence and human capital effect outweighs the crowding-out effect of blacks and Hispanics. A related conclusion found in the same analysis is that modest increases in school curriculum results in significant increases in the educational attainment and hourly wage rate of whites and blacks, but only modest increases for Hispanics.

Another exercise performed with this model is to analyze the effects of equating the quality of schools of blacks and Hispanics to that of whites. The results indicate that policy would lead to significant increases the educational attainment of both minority groups. This policy also leads to significant increases in the hourly wage rate of blacks and more modest increases for Hispanics. Although this policy leads to a significant narrowing of the race education gap, it does not eliminate the gap.

The analysis of the effects of increasing the school curriculum is repeated under this new environment. The result indicates even larger increases in the educational attainment and hourly wage rate of blacks. Also in this environment we see the most significant increases in the educational attainment and hourly wage rates of Hispanics. In other words, Hispanics become significantly more responsive to policies that increases the time they spend on school activities if the quality of the school they attend is improved to the quality of the schools whites attend.

In estimating the dynamic structural model, a multistage procedure was implemented. The potential problem with the multistage procedure is that misspecification of the first stage estimator typically introduces bias in the final stage estimator. Flexible specification of pre-estimators therefore becomes an important goal. The second chapter of this thesis addresses this problem by developing a new semiparametric estimator for a dynamic nonlinear single index panel data model with small T.

In moving away from a fully parameterized nonlinear single index panel data model, there are trade-offs between which assumptions can be relaxed. In general, relaxing the parametric assumption on the unobserved heterogeneity requires maintained parametric assumptions about the index function. Under the assumption that the individual-time specific shocks are independent and if covariates are unbounded, the finite dimensional parameters can be estimated consistently with the parametric convergence rate without specifying the distribution of the individual-specific effects conditional on the covariates if and only if the distribution of the individual-time specific shocks is logistic [Magnac, 2004]. On the other hand, Manski [1987] has shown that the finite dimensional parameters can be consistently estimated with only a strict monotonicity assumption on the index function. However, this estimator does not converge at the parametric rate.

The second chapter of this theses therefore adds to this literature by showing that under the strict monotonicity assumption on the index function and a flexible assumption of the form of the individual specific effect, one can still obtain estimates of the finite dimensional parameters that converge at the parametric rate. Also the estimator proposed also produces estimates of the index function and the individual specific effects. In other words, the full data generating process is recovered. This is important if the intention of the investigator is to perform counterfactual simulations.

The assumption made on the individual specific effects is that it is an unknown function of a known random variable. This restriction extends the suggestion of Newey [1994] and arises naturally in the discrete choice framework. Our own interest however goes beyond the discrete choice framework. We prove that the resulting estimator is not only $\sqrt{n}$ consistent, but that it achieves the semi-parametric efficiency bound. Thus under the same assumption no other estimator can obtain a smaller asymptotic variance. The method used to compute the estimator is the back-fitting algorithm proposed by Buja et al. [1989]. This algorithm has the advantage that it does not depend on the type of smoother chosen to compute the estimate of the index function. The investigator can therefore implement a sieve estimator of a kernel estimator.

A small simulation exercise shows that the proposed estimator performs very well in recovering both the finite dimensional parameters and the index function. This is the case for even small numbers of observations. The method is also implemented to estimate a wage regression. An interesting result is that the function recovered resembles the exponential function, which suggests that the error made by assuming a log-linear wage regression should be relatively small.

The final chapter of this thesis proposes a powerful numerical algorithm to solve independent private values asymmetric first price auctions where the auctioneer sets a reserve price. Asymmetry arises from the specification of ex-ante heterogeneous distributions of private values, as well as from collusion among subsets of bidders. Our algorithm generalizes the seminal work of Marshall et al. [1994] who consider the special case where n players draw their values from uniform distributions on [0,1] and a subgroup of $k_1 < n$ bidders form a coalition.

We also derive operational univariate quadratures to compute the probability that the auctioneer retains the item, the probabilities that a particular bidder wins the item as well as expected revenues for bidders and auctioneer under asymmetric first and second price auctions. Embedding these calculations within a simplex optimization algorithm enables us to compute an optimal reserve price under either auction scheme.

These techniques provide us with a powerful tool to numerically investigate whether results derived under symmetry extend to the asymmetric case as well as the (single auction) viability of collusion among subsets of bidders. Illustrative examples are provided with and without the assumption of stochastic dominance.

2.0 A DYNAMIC STRUCTURAL MODEL OF LABOR SUPPLY AND EDUCATIONAL ATTAINMENT

2.1 INTRODUCTION

Over the last three decades, there has been an increasing trend of young individuals participating in the US labor market while actively enrolled in school. Young individuals are increasing their incidence of labor market participation, and the amount of hours worked while enrolled in school.[1] This trend has generated growing interest in the possible immediate and long run effects of working while enrolled in school on educational attainment and future labor market opportunities. On one hand, there is the concern that an intensive amount of working while in school may hinder academic performance and increase drop-out rates, thus jeopardizing future opportunities.[2] On the other hand, working while in school may improve a young individual's time organizational skills, sense of responsibility and self esteem, which in turn are traits that may be rewarded in the labor market in the future. Furthermore, working while in school produces immediate work experience and cash that may be used to finance their studies.[3] It is not obvious which of these two opposing effects dominate. It may be that the net effect of these opposing forces varies over different groups of young individuals.

This article develops and estimates a dynamic structural model of schooling and work decisions to investigate the process by which a cohort of young males accumulate human capital over their life cycle. The theoretical model provides a detailed treatment of the economic costs and ben-

[1] A recent documentation of this phenomena is found Bacolod and Hotz [2005].

[2] This apprehension is reflected in the article entitled "Long hours taking toll on youths, studies say," by Paloma McGregor, The Plain Dealer, March 5, 2001.

[3] This opinion was expressed in the article entitled "Teens Find Profit and Loss in Work: Part time jobs bring experience and cash, but can hinder studies," by Jacqueline Salmon, The Washington Post, March 28, 1998.

efits associated with the schooling and labor supply alternatives faced by individuals. Specifically, the estimated model explicitly accounts for the simultaneous choice of enrollment in school and labor force participation, endogenous leisure choices, intertemporal nonseparabilities in preferences, aggregate skill specific productivity shocks, aggregate consumption price effects, and individual heterogeneity.

In addition to accounting for the simultaneous choice of work and schooling, the model treats hours spent on schooling, working, and leisure as continuous choice variables.[4] This approach is in contrast to other models (see Keane and Wolpin, 1997, and Eckstein and Wolpin [1999] for examples) that treat leisure time as exogenous to the individual, where an increase in labor supply is equivalent to a decrease in time spent on schooling activities if the individual is enrolled in school. In this framework, an individual may optimally choose to sacrifice leisure and increase time spent on both schooling and labor market activities. In this sense the model is one of optimal intra- and inter-temporal allocation of time among schooling, working and leisure. The model also allows for flexible specification of preferences with respect to time allocation. The additional flexibility comes from the specification of intertemporal nonseparabilities in leisure.

Recent studies of the life-cycle models of labor supply have stressed the importance of intertemporally non-separable preferences.[5] Hotz, Kydland, and Sedlacek [1988] found that the assumption of intertemporally separable preferences for leisure is inconsistent with data for prime-age males. Given that hours schooling activities and leisure are related by the time constraint of the individual, such nonseparabilities are also likely to affect their enrollment and study patterns. The estimation results indicate that leisure choices are intertemporal complements. Increases in current hours of leisure increases the future demand of leisure. In other words, an increase in hours of current schooling activities decreases the future marginal disutility of schooling. This evidence of intertemporal complimentarity suggests habit formation by young men.

The primary data used in this study comes from the National Longitudinal Survey of Youth (NLSY79), which is a comprehensive panel data set that follows individuals who were 14 to

[4]While some studies model these alternatives as mutually exclusive [Keane and Wolpin, 1994, Cameron and Heckman, 1999], the growing trend is to allow for interior solution to choices where individuals simultaneously participate in the labor market and attend school (see D'Amico, 1984, Ruhm, 1997, Oettinger, 1999, and Eckstein and Wolpin, 1999 for examples)

[5]See Hotz et al., 1988, Eichenbaum et al., 1988, Altug and Miller, 1998, Imai, 2000, and Gayle and Miller, 2003 for examples.

21 years of age as at January 1, 1979. The estimation technique implemented is a modified version of the Conditional Choice Probability (CCP) estimator of Hotz and Miller [1993] and Altug and Miller [1998]. This estimation technique allows for unobserved individual-specific effects to be arbitrarily correlated with the observed characteristics in the model. The model employs a fixed effects method of controlling for unobserved heterogeneity. Other models of education, such as Eckstein and Wolpin [1999] control for individual-specific effects by way of a random-effects, finite mixture specification. These techniques typically require that the investigator make strong independence assumptions on the relationship between the unobserved covariates, and their observed counterparts. The cost of the flexibility allowed by a fixed effects specification is the resulting incidental parameters problem. We argue, using previous results [Altug and Miller, 1998, Gayle and Miller, 2003] and evidence from the data used in this paper that these biases are likely to be small.

The incidence of working, the number of hours worked, and the number of years that young men spend working while enrolled in school varies across races. Bacolod and Hotz [2005] documents that the number of years working while in high school increased the most for young Hispanic men, followed by young black men. Young black men experienced the largest increase in working while in college. In estimating the parameters of the model, we pay special attention to racial differences in outcomes that are not accounted for by the rich set of observed background variables found in the NLSY79, nor by estimated individual specific effects. The theoretical model provides a natural separation of these unexplained racial variations into preference differences and statistical discrimination [Altonji and Blank, 1999].

The empirical results indicate that, conditional on enrolling, young black males are likely to spend more time on school activities than white males. Young Hispanic males are likely to spend less time on school activities white males. Furthermore, young black and Hispanic males are less likely to be promoted from the grade level than young white males. These young minority males either repeat the grade level or drop out of school during the school year. These racial differences remain significant after the inclusion of the rich set of demographic variables and measures of ability that are found in the NLSY79, as well as measures of unobserved individual specific characteristics. The lower probability of grade promotion for blacks and Hispanics is interpreted as capturing race specific differences in the school environment. Specifically, in the paper we argue that this grade promotion probability gap is a measure of the differences in the quality of schools that blacks and Hispanics attend as against the quality of schools that whites attend.

Controlling for racial differences in wages, and the aforementioned racial differences in study patterns and grade promotion propensities, the results indicate that there are no race specific differences in the propensity of participate in the labor market, the propensity to enroll in school, nor in the choice of leisure. These results are in contrast to many previous results in structural estimation that find significant race indicators in their specified utility functions. The result is this paper suggests that racial disparity in outcomes are due to the racial differences in the school and work environment, and not to racial differences in tastes and preferences.

The model is solved and simulated in order to analyze the effects of various hypothetical policies. The first policy analyzed is one where the government subsidizes students who decide not to participate in the labor market. The simulated results indicate that this policy does very little in affecting the level of education, labor market experience, and wages on young men. The second policy analyzed is one where the school administration adjusts the school curriculum so that young men who enroll necessarily spend more time on school activities. Such a policy can be achieved by increasing the number of hours in school, increasing the number or difficulty of assignments, after school programs, or Saturday (Sunday) classes. The results indicate that such a policy has significant positive effects on wages and education of whites and blacks, and more modest positive effects on Hispanics.

The third simulation exercise analyzes a situation where school quality of blacks and Hispanics are equated to those of whites. The results indicate that this policy has significant positive effects on the level of education and wages of blacks. The effects of this policy on Hispanic are positive but much more modest than that for blacks. The final simulation exercise evaluates the same policy of increasing time spent on school activities of blacks and Hispanics after equating the school quality of minorities to to those of whites. It is in this environment where we find significant increases in wages and education for Hispanics. We find also significant increases in wages and education of blacks. The results indicate therefore that policies that are aimed at increasing the time minorities spend on school activities are significantly more effective if the school environment of minorities are improved to match those of white.

The rest of the paper is organized as follows. In the next section, we present the basic behavioral model. We then discuss the solution of the model in section (2.3) and describe the first order necessary conditions for optimality that will be used in estimation. Section (2.4) discusses the construction of the sample used in estimation, and Section (2.5) discusses the empirical methodology implemented in estimation of the parameters of interest. Section (2.6) describes the estimation of the consumption function and discusses the empirical findings. Section (2.7) discusses the estimation of the wage equation and the empirical findings. Section (2.8) discusses the estimation of the time spent on schooling activities and the transition probabilities. Section (2.9) presents the methodology used to estimate the conditional choice probabilities and their corresponding derivatives, which are needed to estimate the preference parameters. Section (2.10) presents the moment conditions and corresponding sample analogs that are used in estimating the preference parameters of the model, as well as discuss the empirical findings of the model. Section (2.11) presents the method of solving the dynamic programming model and discusses the policy simulations. Section (2.12) concludes.

2.2 THE THEORETICAL MODEL

This section develops the theoretical framework that is used to investigate how individuals allocate time between human capital accumulation, labor market participation, and leisure.

2.2.1 Environment

The model is set in discrete time $t \in \{0, 1, \cdots, T\}$. We assume that there exists a continuum of individuals on the unit interval [0,1]. Associated with each individual is a K-dimensional vector of exogenous covariates, denoted z_{nt}, which is assumed to be independently distributed over the population with known cumulative distribution function $Q_0(z_{nt+1}|z_{nt})$. In each period, individual $n \in [0, 1]$ is endowed with a fixed amount of time normalized to one. He must choose how to allocate this unit of time between leisure l_{nt}, the time spent on labor market activities h_{nt}, and the time spent on school activities s_{nt}:

$$1 = l_{nt} + h_{nt} + s_{nt}. \tag{2.2.1}$$

Define $d^h_{nt} \equiv 1_{\{h_{nt}>0\}}$ and $d^s_{nt} \equiv 1_{\{s_{nt}>0\}}$ where $1_{\{\cdot\}}$ is the indicator function equal to one if the event in parentheses occurs and zero otherwise. There is a single composite consumption good in the economy which is consumed and traded by all individuals. Let c_{nt} denote this composite good.

We assume the model has a Markov structure, in which the individual does not need to remember the full history to solve this problem, but only a summary statistic x_{nt}, belonging to a finite vector space $\mathcal{X}$. In particular, define $(h_{nt-\rho}, \cdots, h_{nt-1})$ as the ρ-dimensional vector of past labor supply outcomes, $(s_{nt-\rho}, \cdots, s_{nt-1})$ as the ρ-dimensional vector of past time spent on schooling activities, S_{nt} as the highest grade completed by individual n as at the beginning of t, and E_{nt} as the total years of labor market experience accumulated by individual n as at the beginning of period t. Define also $(c_{nt-\rho}, \cdots, c_{nt-1})$ to be the ρ-dimensional vector of past consumption. Then the typical observed state vector for individual n at time t is given by the $(3\rho + k + 1)$-dimensional vector[6]

$$x_{nt} \equiv (h_{nt-\rho}, \cdots, h_{nt-1}, s_{nt-\rho}, \cdots, s_{nt-1}, S_{nt-\rho+1}, \cdots, S_{nt}, c_{nt-\rho}, \cdots, c_{nt-1}, E_{nt-\rho}, z'_{nt})' \quad (2.2.2)$$

Given that individual n has chosen to enroll in school, he may or may not complete that grade level. If he does complete the grade he is currently enrolled in, his level of education increases by one grade. Otherwise, his level of education remains unchanged. The probability that an individual advances a grade level given that he has enrolled in school at the beginning of period t is denoted by $F(x_{nt})$.

2.2.2 Technology

We assume that the individual has access to a sector specific production technology in each period where, if he works in sector $j = 1, \cdots, J$, he produces a quantity of the output $w_{ntj}h_{nt}$. Here, w_{ntj} is marginal product of labor of individual n at time t with skill level j. It is assumed that w_{ntj} is composed of J exogenously determined time specific aggregate skill prices ω_{tj}, an individual specific, time invariant productivity effect, μ_n, and a skill specific function of his stock of human capital, his socio-economic characteristics and other state vectors, $\gamma_j(x_{nt})$:

$$w_{ntj} = \omega_{t,j}\mu_n\gamma_j(x_{nt}), \quad (2.2.3)$$

Thus $\mu_n\gamma_j(x_{nt})$ is the number of efficiency units of labor supplied by the worker per unit of time in sector j, while $\omega_{t,j}$ is the time specific aggregate price of skill in sector j.

[6] To conserve on notation in what follows, we will use x_{nt} to denote any subset of this vector.

2.2.3 Choice Set

This model falls within the class of mixed continuous and discrete Markov decision processes. The continuous choice variables in this model are c_{nt},h_{nt},and s_{nt}. If $h_{nt} = 0$, individual n does not work at time t. Otherwise, the individual works for the fraction of time $h_{nt} > 0$. Likewise if $s_{nt} = 0$, individual n does not attend school at time t. Otherwise, the individual studies for the fraction of time $s_{nt} > 0$. Define the discrete choice variables for each individual $n \in [0,1]$ at time $t \in \{0,1,\cdots,T\}$:

$$d_{nt0} \equiv \begin{cases} 1 & \text{if } d^h_{nt} = 0 \text{ and } d^s_{nt} = 0 \\ 0 & \text{otherwise} \end{cases}, \tag{2.2.4}$$
$$d_{nt1} \equiv \begin{cases} 1 & \text{if } d^h_{nt} = 1 \text{ and } d^s_{nt} = 0 \\ 0 & \text{otherwise} \end{cases},$$
$$d_{nt2} \equiv \begin{cases} 1 & \text{if } d^h_{nt} = 0 \text{ and } d^s_{nt} = 1 \\ 0 & \text{otherwise} \end{cases},$$
$$d_{nt3} \equiv \begin{cases} 1 & \text{if } d^h_{nt} = 1 \text{ and } d^s_{nt} = 1 \\ 0 & \text{otherwise} \end{cases}.$$

2.2.4 Preferences

Similar to models such as Heckman [1976] and Eckstein and Wolpin [1999], we assume that attending school provides some consumption value to the individual. Learning may be directly valued by the individual, and social interaction within the school environment may provide positive consumption value. However, in this specification, this consumption value of attending school is not confounded with the loss in leisure due to schooling activities since leisure is modelled directly. We specify the contemporaneous utility of attending school as follows:

$$U_{nt1} = u_1(d^s_{nt}, x_{nt}). \tag{2.2.5}$$

Similarly, we assume that there is a utility associated with labor market participation. We specify this contemporaneous utility of labor force participation as follows:

$$U_{nt2} = u_2(d^h_{nt}, x_{nt}). \tag{2.2.6}$$

Preferences are assumed to be additive in consumption and leisure, but not separable with respect to leisure over time. The contemporaneous utility of leisure is therefore given by:

$$U_{nt3} = u_3(x_{nt}, l_{nt}). \tag{2.2.7}$$

The utility of leisure is specified to be dependent on current leisure level and the level of leisure consumed over the last ρ periods.[7]. We assume that u_3 is increasing and concave in l_{nt}. The utility derived from the consumption good in time t is also assumed to be increasing and concave in c_{nt} and is denoted by

$$U_{nt4} = u_4(c_{nt}, z_{nt}). \tag{2.2.8}$$

We introduce a vector of choice specific utility shifters $(\varepsilon_{nt0}, \cdots, \varepsilon_{nt3})'$, which are assumed to be independent over (n,t) and drawn from a population with a distribution function $Q_1(\varepsilon_{nt0}, \cdots, \varepsilon_{nt3})$. They are interpreted to be choice specific, time-varying characteristics that partially determine the utility associated with the corresponding alternatives and unobserved to the econometrician. Let $\beta \in (0,1)$ denote the common subjective discount factor, and E_0 denote expectation conditional on the information set at date 0. The expected discounted lifetime utility of individual n is given by:

$$E_0\left\{\sum_{t=0}^{T} \beta^t \left[\sum_{k=1}^{4} d_{ntk}(U_{nt1} + U_{nt2} + U_{nt3} + U_{nt4} + \varepsilon_{ntk})\right]\right\}. \tag{2.2.9}$$

[7] The lags in leisure are not specified explicitly here since it is a subset of the state vector x_{nt} by equation (2.2.1)

2.3 THE OPTIMIZATION PROBLEM

The inclusion of an aggregate component in marginal product of labor (2.2.3), complicates estimation. To make the model empirically tractable, we assume that markets are competitive and complete. Agents are price takers and there are no distortions in the market for the consumption good, labor supply and loans, a common interest rate facing borrowers and lenders, and that a rich set of financial securities exists to hedge against uncertainty. This assumption incorporates uncertainty in a sufficiently simple manner that leads to a tractable econometric model. Competitive and complete capital market assumption was used by Ben-Porath [1967], Blinder and Weiss [1976], Heckman [1976], and Shaw [1989] to analyze life cycle models of human capital accumulation. This assumption was also recently used by Altug and Miller [1990], Altug and Miller [1998], and Gayle and Miller [2003] to estimate life-cycle models of consumption, labor supply and fertility decisions with aggregate shock.

One key restriction that the assumption of competitive and complete markets places on the model is the lack of any binding borrowing constraint. Borrowing constraints are popular considerations in the study of educational choice. It is a widespread postulation that borrowing constraints critically restricts economically disadvantaged individuals from obtaining the level of formal education that they would have attained otherwise. However, the empirical evidence does not support this view. Cameron and Heckman [1999, 1998] conclude that it is the long-term influences of family and environment that account for ethnic and racial disparities in school attendance, and not short term liquidity constraints. Keane [2002] conclude that borrowing constraints have little effect on college attendance decisions. In the light of these and other evidences, we abstract from any considerations of liquidity constraints and thus the assumption of competitive and complete markets presents itself as an appealing approximation.

Under the assumptions of competitive and complete markets, we appeal to the fundamental welfare theorems which allows us to recast the optimization problem as a social planner problem. The objective function of the social planner is the weighted average of the expected discounted utilities of each individual n given in (2.2.9). The social weight attached to an individual is given by η_n^{-1}. The optimization problem of the social planner is subject to the time allocation constraint for each individual (2.2.1), as well as the production technology available to each individual as reflected in (2.2.3). Define L to be the lebesgue measure that integrates over the population. The aggregate feasibility condition is given by:

$$\int_0^1 [c_{nt} + a_{nt} + \pi_{nt} - w_{nt} h_{nt}] \mathrm{d}L(n) \leq 0, \qquad t \in \{0, 1, \cdots, T\}. \tag{2.3.1}$$

where a_{nt} is the individual savings at time t, or the value of claims to period $t+1$ consumption net of the claims to time t consumption. π_{nt} is the direct schooling expenses incurred by the individual if he chooses to enroll in period t.

The Pareto optimal allocations are found by maximizing

$$E_0 \Big\{ \int_0^1 \sum_{t=0}^{T} \beta^t \eta_n^{-1} \Big[\sum_{k=1}^{4} d_{ntk} (U_{nt1} + U_{nt2} + U_{nt3} + U_{nt4} + \varepsilon_{ntk}) \Big] \mathrm{d}L(n) \Big\}, \tag{2.3.2}$$

subject to (2.3.1) and (2.2.1) with respect to sequences for consumption, schooling, and labor supply $\{c_{nt}, s_{nt}, h_{nt}\}_{t=0}^{T}$ for all individuals $n \in [0, 1]$.

2.3.1 Optimal consumption

Define $\beta^t \lambda_t$ as the Lagrange multiplier associated with the aggregate feasibility constraint in equation (2.3.1). Given the assumption of an interior solution for consumption allocation, the set of necessary conditions characterizing optimal consumption allocation are given by

$$\frac{\partial u_3(c_{nt}, x_{nt})}{\partial c_{nt}} = \eta_n \lambda_t, \tag{2.3.3}$$

for all $n \in [0, 1]$ and $t \in \{0, \cdots, T\}$. Under the assumption of contemporaneous separability of consumption from education and labor supply choices, (2.3.3) can be used to solve for individuals' Frisch demand functions which determines optimal consumption allocation in terms of the time-varying characteristics x_{nt} and the shadow value of consumption $\eta_n \lambda_t$. Assume that the utility derived from consumption takes on the following augmented CRRA specification:

$$u_3(c_{nt}, x_{nt}) = g(x_{nt}) \frac{c_{nt}^{\alpha}}{\alpha}. \tag{2.3.4}$$

Then condition (2.3.3) takes the form

$$g(x_{nt}) c_{nt}^{\alpha - 1} = \eta_n \lambda_t. \tag{2.3.5}$$

Multiplying (2.3.5) by $\alpha^{-1}c_{nt}$ gives the following alternative representation of the indirect contemporaneous utility derived from consumption:

$$u_3(c_{nt}, x_{nt}) = \frac{\eta_n \lambda_t}{\alpha} c_{nt}. \tag{2.3.6}$$

The empirical strategy comprises of estimating the parameters of the utility function u_3 from (2.3.3) and (2.3.4) to obtain estimates of the individual specific weights η_n as well as the Lagrange multiplier λ_t. These estimates are then substituted in (2.3.6), which is in turn substituted into the social planner's objective function (2.3.2).

Under the assumption that none of the consumption good is wasted at the optimal allocation, the first order necessary condition with respect to the the lagrange multiplier $\beta^t \lambda_t$ gives the optimal consumption allocation for each individual

$$c_{nt} = w_{nt} h_{nt} - a_{nt} - \pi_{nt}. \tag{2.3.7}$$

2.3.2 Optimal schooling and labor supply

Characterizing the optimal labor supply, leisure and schooling decision is more complicated. The optimal schooling and work allocations are confounded by the constraint imposed by (2.2.1). In particular, in any period, increasing both schooling and labor supplied by individual n necessarily leads to a decline in the level of leisure enjoyed by that individual. Consequently, the optimal allocation of labor supply, education and leisure cannot be separately solved for as in the case of optimal consumption allocation. Following Altug and Miller [1998], the conditional valuation functions associated with the discrete choices on individual n in period t is defined as:

$$V_{ntj} + \varepsilon_{ntj} \equiv \max_{\{s_{nr}, h_{nr}\}_{r=t}^{T}} E_t \left\{ \begin{array}{l} \sum_{r=t}^{T} \beta^{r-t} [\sum_{k=0}^{3} d_{nrk} (U_{nr0} + U_{nr1} \\ + \alpha^{-1} \eta_n \lambda_r (w_{nt} h_{nt} - a_{nt} - \pi_{nt}) + \varepsilon_{nrk}) | d_{ntj} = 1] \end{array} \right\}. \tag{2.3.8}$$

Let d^0_{ntj} be the socially optimal decision by individual n in period t. The term $V_{ntj} + \varepsilon_{ntj}$ denotes the social value from individual n choosing alternative j at time t. Accordingly, individual n's choice of alternative j at time t is optimal if

$$d^0_{ntj} = \begin{cases} 1, & \text{if } V_{ntj} + \varepsilon_{ntj} > V_{ntk} + \varepsilon_{ntk} \quad \forall\, k \neq j \\ 0, & \text{otherwise} \end{cases}. \tag{2.3.9}$$

Let h^0_{nt} and s^0_{nt} be the optimal interior choice of labor supply and study time. Given that it is socially optimal for individual n to work in time t, h^0_{nt} must satisfy

$$\frac{\partial V_{ntj}}{\partial h_{nt}} = 0, \qquad \text{for} \;\; j = 1, 3. \tag{2.3.10}$$

Likewise, given that it is socially optimal for individual n to enroll in time t, s^0_{nt} must satisfy

$$\frac{\partial V_{ntj}}{\partial s_{nt}} = 0, \qquad \text{for} \;\; j = 2, 3. \tag{2.3.11}$$

In order to express the conditional valuation function recursively, define p_{ntj} to be the probability of individual n choosing option j in period t conditional on the information set available to him in period t

$$p_{ntj} \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{V_{ntj} - V_{nt0} + \varepsilon_{ntj}} \cdots \int_{-\infty}^{V_{ntj} - V_{nt3} + \varepsilon_{ntj}} \mathrm{d}Q_1(\varepsilon_{nt0}, \cdots, \varepsilon_{nt3}). \tag{2.3.12}$$

The information set available to individual n at period t is composed of the observed state vector x_{nt}, and the unobserved individual specific and aggregate shocks to productivity and consumption. Define this state vector as $\Psi_{nt} \equiv (x'_{nt}, \mu_n, \eta_n, \lambda_t, \omega_{t1}, \cdots, \omega_{tJ})'$. Define also $\mathcal{A}^i_{nt}$ to be the set of all possible realizations of the state vector for individual n at i periods after t given the realization of the state vector Ψ_{nt} at period t. Correspondingly, let $F_j(\Psi^{(i)}_{nt}|\Psi_{nt})$ is the probability that the state vector of individual n in period $t+i$ is $\Psi^{(i)}_{nt}$, given that his state vector in period t is Ψ_{nt} and he chooses alternative j in period t. Then from equation (2.3.9), the conditional probability that alternative j is chosen by n in period t in equation (2.3.12) has the following alternative representation

$$p_{ntj} \equiv p_j(\Psi_{nt}) \equiv E[d^0_{ntj}|\Psi_{nt}], \tag{2.3.13}$$

and Hotz and Miller [1993] prove the existence of a mapping $\varphi_k : [0,1] \to \Re$ such that

$$\varphi_k(p_k(\Psi_{nt})) = E[\varepsilon_{ntk}|\Psi_{nt}, d^0_{ntk} = 1], \qquad k \in 0, \cdots, 3. \tag{2.3.14}$$

Therefore, the conditional valuation function has the following recursive representation:

$$\begin{aligned} V_{ntj} &= \max_{h_{nt}>0}\Big\{U_{nt0}+U_{nt1}+\alpha^{-1}\eta_n\lambda_t(w_{nt}h_{nt}-a_{nt}-\pi_{nt}) \\ &+\beta\Big[\textstyle\sum_{\Psi_{nt}^{(1)}\in\mathcal{A}_{nt}^{1}}\Big[\sum_{k=0}^{3}p_{nt+1,k}(V_{nt+1,k}+\varphi_k(p_k(\Psi_{nt}^{(1)})))\Big]F_j(\Psi_{nt}^{(1)}|\Psi_{nt})\Big]|d_{ntj}=1\Big\}. \end{aligned} \tag{2.3.15}$$

Finally, the optimality conditions for interior solution to labor supply h_{nt}^{0} (2.3.10) and study time s_{nt}^{0} (2.3.11) are given by

$$\begin{aligned} \frac{\partial U_{nt1}}{\partial h_{nt}}+\frac{\eta_n\lambda_t}{\alpha}w_{nt} = &-\beta\Big\{\textstyle\sum_{\Psi_{nt}^{(1)}\in\mathcal{A}_{nt}^{1}}\Big[\sum_{k=0}^{3}\Big[p_{ntk+1}\frac{\partial(V_{nt+1,k}+\varphi_k(p_k(\Psi_{nt}^{(1)})))}{\partial h_{nt}} \\ &+\frac{\partial p_{nt+1,k}}{\partial h_{nt}}(V_{nt+1,k}+\varphi_k(p_k(\Psi_{nt}^{(1)})))\Big]F_j(\Psi_{nt}^{(1)}|\Psi_{nt}) \\ &+\textstyle\sum_{k=0}^{3}p_{nt+1,k}(V_{nt+1,k}+\varphi_k(p_k(\Psi_{nt}^{(1)})))\frac{\partial F_j(\Psi_{nt}^{(1)}|\Psi_{nt})}{\partial h_{nt}}\Big]|d_{ntj}=1\Big\}, \text{ and,} \end{aligned} \tag{2.3.16}$$

$$\begin{aligned} \frac{\partial U_{nt1}}{\partial s_{nt}} = &-\beta\Big\{\textstyle\sum_{\Psi_{nt}^{(1)}\in\mathcal{A}_{nt}^{1}}\Big[\sum_{k=0}^{3}\Big[p_{ntk+1}\frac{\partial(V_{nt+1,k}+\varphi_k(p_k(\Psi_{nt}^{(1)})))}{\partial s_{nt}} \\ &+\frac{\partial p_{nt+1,k}}{\partial s_{nt}}(V_{nt+1,k}+\varphi_k(p_k(\Psi_{nt}^{(1)})))\Big]F_j(\Psi_{nt}^{(1)}|\Psi_{nt}) \\ &+\textstyle\sum_{k=0}^{3}p_{nt+1,k}(V_{nt+1,k}+\varphi_k(p_k(\Psi_{nt}^{(1)})))\frac{\partial F_j(\Psi_{nt}^{(1)}|\Psi_{nt})}{\partial s_{nt}}\Big]|d_{ntj}=1\Big\}, \end{aligned} \tag{2.3.17}$$

for $j=1,3$, and $j=2,3$ respectively. The first condition in (2.3.16) says that the net current benefit from an additional hour of work is equal to the present discounted value of future utility costs of that additional hour. The current marginal utility from an additional hour of work is equal to the net of the utility cost of leisure forgone, and the consumption value of the additional goods and services produced. The future value of an additional hour of work is decomposed into three main components. The first term on the RHS captures the direct effect of an increase in hours worked on future productivity and future utitily. Future utility is directly affected because of the assumption that current and future leisure are intertemporally nonseparable. Future productivity is affected by the assumption that current labor force participation enhances human capital, which is reflected in higher future marginal productivity of labor. The second term on the RHS captures the indirect effect on future utility by current hours worked through its effect on future probability of employment. The third term on the RHS accounts for the indirect effect of current hours worked on future utility through its effect on the transition probability. The probability of being promoted a grade level given that the individual is currently enrolled is assumed to be dependent on hours worked.

2.4 DATA

The data is taken from the 1979 youth cohort of the National Longitudinal Survey of Labor Market Experience (NLSY79), a comprehensive panel data set that follows individuals over the period 1979 to 2000, who were 14 to 21 years of age as of January 1, 1979. The data set initially consisted of 12,686 individuals: a representative sample of 6,111 individuals, a supplemental sample of 5,295 Hispanics, non-Hispanic blacks, and economically disadvantaged, non-black, non-Hispanics, and a supplemental sample of 1,280 military youth. Interviews were conducted on an annual basis though 1994, after which they adopted a biennial interview schedule. This study makes use of the first 16 years of interviews, from 1979 to 1994.[8] The data is restricted to include males and to exclude respondents with missing observations on the highest grade level completed that cannot be recovered with high confidence from other data information. A list and description of the variables used in the model is presented in Table 1. Table 2 presents summary statistics of the sample used in this study. Attrition accounts for a loss of approximately 22 percent of the individuals between 1979 and 1994. However, the largest loss occurred between 1990 and 1991, late in the sample period.

2.5 ESTIMATION METHOD

The empirical analysis employs a multi-stage version of the conditional choice probability (CCP) estimator developed in Hotz and Miller [1993] and extended by Altug and Miller [1998]. We outline the estimation strategy of each stage in turn. The parameters of the model can be estimated from the optimality conditions derived in section (2.3). First, there is contemporaneous separability between consumption and labor supply in the utility function. Given that consumption is measured with error and that the measurement error is uncorrelated with the information set of the individual, the consumption function can be estimated separately from the equations characterizing optimal discrete choice to provide first stage estimates of the of the shadow price of consumption. Similarly, assuming that observed wages are noisy measures of the marginal product of labor, where

[8] Appendix 1 provides a detailed discussion of the data construction and sample restrictions.

the measurement error is assumed to be independent of the information set of the individual over time, the parameters of the marginal product of labor can be estimated separately from the other parameters of the model.

Examination of equations (2.3.15) and (2.3.10) in section (2.3) suggest that estimation of the conditional choice probabilities p_{knt} and their derivatives with respect to hours worked h_{nt} and study time s_{nt} are required. These quantities are estimated nonparametrically and substituted into the necessary conditions for optimal choice and hours allocation. The technique employed here also requires that the transition probabilities be estimated. The remaining parameters of the model are estimated by nonlinear GMM, where the moment conditions are formed as sample analogs of equations (2.3.9), (2.3.16) and (2.3.17). Since the first stage regressions are of interest in their own right, we discuss them in separate sections.

2.6 CONSUMPTION

Estimation of the marginal utility of consumption requires further parametrization of the utility of consumption given by equation (2.3.4). We assume that $g(x_{nt})$ has the following parametrization:

$$g(z_{nt}) = \exp(x'_{nt}B_1), \tag{2.6.1}$$

The first order necessary conditions for optimal consumption allocation are then given by:

$$\exp(x'_{nt}B_1)c_{nt}^{\alpha-1} = \eta_n\lambda_t. \tag{2.6.2}$$

The necessary conditions (2.6.2) and (2.3.7) provide the key equations for the estimation of the shadow value of consumption λ_t and the individual specific effect η_n. Taking the natural log of equation (2.6.2) and rearranging results in the following equation

$$ln(c_{nt}) = (1-\alpha)^{-1}x'_{nt}B_1 - (1-\alpha)^{-1}\ln(\eta_n) - (1-\alpha)^{-1}\ln(\lambda_t). \tag{2.6.3}$$

Assuming that observed consumption $\tilde{c}_{nt}$ is measured with error so that $\tilde{c}_{nt} = c_{nt}e^{\nu_{nt}}$, where c_{nt} is the true level of consumption, and $E[\nu_{nt}|x_{nt}, \eta_n, \lambda_t] = 0$. Let Δ denote the first-difference operator. Taking first difference of equation (2.6.3) and rearranging, we have that

$$\Delta\nu_{nt} \quad = \quad \Delta\ln(\tilde{c}_{nt}) - (1-\alpha)^{-1}\Delta x'_{nt}B_1 + (1-\alpha)^{-1}\Delta\ln(\lambda_t). \tag{2.6.4}$$

Equation (2.6.4) is estimated by the efficient GMM. The estimated results in Table 4 indicate that consumption increases with the size of the family, average family income, and the average age of the family. Consumption decreases with the level of unemployment local to the residence of the individual. Table 4 also suggests that for a given level of education, consumption is increasing and concave in the age of the individual. For a given age of the individual, consumption is decreasing and convex in the level of education.

The first panel of Table 6 reports the estimated log change in aggregate prices with the corresponding standard errors. The graph along with the 95% confidence interval are also presented in Figure 1. These figures show that the changes in aggregate prices are estimated precisely. The figure also show that there are significant variation in the time effects. The simple F-test reject the restriction that $(1-\alpha)^{-1}\ln(\lambda_2) = \cdots = (1-\alpha)^{-1}\ln(\lambda_T)$ at the 99% confidence level.

2.7 WAGES

Assume that the time varying component of the individuals productivity function has the representation:

$$\gamma_j(x_{nt}) \equiv \exp(x'_{nt}B_{2j}). \tag{2.7.1}$$

Observed wages are assumed to be noisy measures of the marginal productivity of labor, where the multiplicative error term is assumed to be conditionally independent over individuals, the covariates in the wage equation, and the labor supply decision

$$\tilde{w}_{ntj} = \omega_{tj}\mu_n \exp(x'_{nt}B_{2j})\exp(\varepsilon_{nt}). \tag{2.7.2}$$

The individual specific effects captures absolute advantage of the individual in the labor market [Willis, 1986]. Assume that human capital comes in two types, an unskilled type ($j = 1$) and a skilled type ($j = 2$). The skilled group is defined as having at least 16 years of formal education. All occupations in the economy are sorted across these groups according to the level of education required to carry out the task. Workers are assumed to be perfect substitutes within, but not across efficiency units. Since the model is in the panel data framework, we do not need to assume that schooling and employment choices are independent of the individual's ability as captured by the individual specific effect. This is in contrast to the model proposed in Willis [1986]. The absence of this restriction serves to eliminate the problem of sample selection caused by ability bias.

Another key consideration in the estimation of equation (2.7.2) is whether there is the need to estimate separate models for the different racial groups. The results of Neal and Johnson [1996] and Altonji and Blank [1999] indicate that the large majority of the wage gap between races in the NLSY is due to differences in measures of abilities (AFQT scores) and family background (parents education). Since these measures are time invariant, a suitable transformation of a single wage equation provides accurate estimate in the pooled data.

Taking logs of both sides of equation (2.7.2) and taking first difference gives the following equation:

$$\Delta\varepsilon_{nt} = \Delta\ln(\tilde{w}_{ntj}) - \Delta\ln(\omega_{tj}) - \Delta x'_{nt}B_{2j} \tag{2.7.3}$$

Define e_{nt1} to be equal one if individual n is belongs to efficiency unit 1 in period t. Likewise, define e_{nt2} to be equal one if individual n is belongs to efficiency unit 2 in period t. Equation (2.7.3) is estimated by the efficient GMM. The skill specific coefficients are obtained by interacting the explanatory variables with these indicator variables for each skill group. The skill specific aggregate effects are also obtained by interacting the time dummies with these indicator variables.

The estimated results for the wage equation are reported in Table 4. The positive coefficients on lagged hours indicate that there are positive returns to on the job training. Also, the effect of past hours worked on current wages decline with further lags. The declining magnitude and significance of lagged hours worked is consistent with the conjecture of depreciation in human capital. The returns to on the job training are higher for skilled workers than for unskilled workers. At 2000 hours per year, the wage elasticity of the first lagged hours is 0.04 for low skilled workers and 0.06 for high skilled workers. However, the wage elasticity of the second lagged hours is 0.01 and 0.02. These qualitative results are in line with those found in Miller and Sanders [1997], Altug and Miller [1998] and Gayle and Miller [2003].

The coefficients on the education and experience variables are all estimated highly precisely, with the exception of education squared for low skilled workers.[9] The coefficient of squared education is positive and significant at the 1% level for the high skilled group, indicating nonlinearity in marginal returns to education. We find that the coefficient on the interaction term between education and experience is positive for low skilled workers and negative for high skilled workers, both significant at the 1% level. This suggests that in terms of the productivity of young males, formal education and labor market experience are compliments in the low skilled sector, and substitutes in the high skilled sector.

The flexibility of the specification of the wage equation also allows for some heterogeneity in the returns to education. It allows for comparative advantage with respect to human capital in the labor market to be manifested through differences in patterns of schooling and employment. At first glance marginal return to education for both the skilled and unskilled sector seem very low. Indeed, the calculation would produce a marginal rate of return of 0.024 for low skilled workers and 0.069 for high skilled workers of age 30 in the sample. Table 5 of Card [1999] lists the estimated marginal returns to education found in a number of studies. The marginal returns to education found here are lower than these other estimates. However, these other studies do not account for growth in skill specific aggregate wages. When the average growth in log aggregate wages in included in the calculation, the estimated marginal return to log wages increases to 0.044 for low skilled workers and 0.217 for high skilled workers of age 30. The estimated marginal returns to education in Table 5 of Card [1999] all fall within the range.

The last two panels of Table 6 report the estimated changes in unskilled and skilled piece rates. These series are also plotted in Figures 2 and 3 along with their 95% confidence bands. The changes in unskilled piece rates ar less precisely estimated than the changes in skilled piece rates. Two separate hypothesis tests are performed. The first is an F-test of the restriction of equality of all the aggregate effects $\Delta \ln(\omega_{21}) = \cdots = \Delta \ln(\omega_{T1}) = \Delta \ln(\omega_{22}) = \cdots = \Delta \ln(\omega_{T2})$. The

[9]Because most individuals in the sample have no breaks in schooling until they have completed their total level, identification of level of schooling in a first difference model is fragile at best and is excluded from the specification. We exclude the level of experience for the same reason.

second is an F-test of the restriction of a single set of time varying aggregate effects $\Delta\ln(\omega_{21}) = \Delta\ln(\omega_{22}), \cdots, \Delta\ln(\omega_{T1}) = \Delta\ln(\omega_{T2})$. Both restrictions are rejected at the 99% level.

2.7.1 Individual-specific Effects

To estimate preference parameters of the model we need to estimate the individual specific effects η_n and μ_n. They are estimated from the residuals in the log-linear versions of consumption and wage equations (2.6.3) and (2.7.3) respectively. These estimators are subject to small sample bias when T is small. However, Hotz et al. [1988] provide Monte Carlo evidence that the small sample bias caused by using such fixed-effects estimates in computing the remaining parameters of interest are quite small for moderate to large sample sizes. Altug and Miller [1998] and Gayle and Miller [2003] estimate the parameters of their structural model under two assumptions on the fixed effects. The first is the traditional definition. The second assumes that fixed effects can be written as functionals of observed covariates. Under the second assumption, consistency of the other parameters of the model is achieved. In their studies, the resulting estimates of the structural parameters were very similar, and lead to the same conclusions. This also indicates that the bias induced by employing estimates of the traditional fixed effects is quite small in these models. The estimates of μ_n and η_n are calculated from samples where $T_1 = 15$ and $T_2 = 12$ respectively. Hahn et al. [2001] suggests that these sample sizes are actually large, implying that the bias of these estimates are expected to be small.

The fixed effects estimators of $\ln(\mu_n)$ and $\ln(\eta_n)$ are obtained as simple time averages of the estimated residuals of the consumption and wage equations.

2.8 STUDY PATTERNS AND THE PROBABILITY OF GRADE PROMOTION.

2.8.1 Study Patterns

In 1981, the NLSY79 collected information on the patterns of school activities of the respondents that are enrolled in school. In particular, the NLSY79 asked the respondent about the amount of hours they spent in school during the week before the interview date. They also asked whether or not the time they reported is typical or not, and if no, to report the typical hours spent in school. The respondents also reported the number of hours they spent studying outside of school during the week before the interview date. These responses are used to construct yearly measures of the time spent by individuals on school activities. We show that one can get reliable estimates of time spent on schooling activities from this data. We call this time spent on schooling activities study time. Clearly this includes not only the time the individual spends actually studying, but also time the student allocates to activities related to school, both during regular hours of school and outside of school.

Assume that the study time of an individual n in period t is an exponential function of observed demographic characteristics and literacy indicators of the individual, as well as unobserved individual-specific characteristics ,

$$s_{nt} \equiv \exp(x'_{nt} B_3).$$

Assume further that observed study time is a noisy measure on the true study patterns of the individual, where the measurement error is assumed to be independent of the regressors.

$$\tilde{s}_{nt} \equiv \exp(x'_{nt} B_3) \exp(\varepsilon_{nt}). \tag{2.8.1}$$

Under these assumptions, we can consistently estimate the study time of individuals enrolled in school using OLS on the log-linearized version of equation (2.8.1).

To estimate the preference parameters of the model, we need a consistent estimate of study time given that an individual has enrolled in school. Thus the issue of sample selection bias does not affect the estimation of equation (2.8.1). Another consideration is the fact that individuals were questioned about their study patterns for only one week prior to the interview period. If the interview is taken at a time where there are generally academic deadlines such as exams, then the reported time spent studying may be overstated. However, interviews were administered to different individuals at different times of the year. This makes plausible the assumption that on average, one does not expect to observe over nor under reporting of study time in the data.

Table 7 reports the regression of the time spent on school activities. The number of observations in estimation is 2253. All variables included in the specification are significant at the 5% level. The F-statistic for the model is 20.47, and the Adjusted R^2 is 11.24%. These statistics show that the instruments do well, both individually and as a group, in capturing variation in log study time. In particular there is no problem of weak instruments in this estimation of study time. This issue of weak instruments is important since the predicted values of study time serve as first stage estimates in all the estimators that follow.

The results in Table 7 show that lagged enrollment decisions are positively associated with study time, with further lags becoming less important. The size of the coefficients indicate also that lagged enrollments decisions are also quite relevant in explaining current study time. Lagged hours of work are negatively correlated with current study time, with diminishing impact for further lags. The magnitude of these effects are also considerable. Individuals with higher AFQT scores spend more time on school activities. Since the AFQT test was administered in 1980 and the data on schooling activities were collected in 1981, there is no issue of feedback effects of current study time on AFQT scores. The results also indicate that the time spent on schooling activities is approximately 11% higher for blacks and 10% lower for hispanics compared to time spent by whites. These differences are quite large, working out to be approximately 154 more hours per year for blacks and 140 less hours per year for hispanics at an average of 1400 hours, approximately what is in the sample.

2.8.2 The Probability of Grade Promotion

An individual who decides to enroll in a particular grade level may or may not be promoted from the grade. This probability of promotion is of interest in it own right, and is also a key ingredient in the final stage estimation. Assume that this probability takes the logit form:

$$F(x_{nt}) \equiv (1-d^h_{nt})\frac{\exp(x'_{nt}B_{41})}{1+\exp(x'_{nt}B_{41})} + d^h_{nt}\frac{\exp(x'_{nt}B_{42})}{1+\exp(x'_{nt}B_{42})}. \tag{2.8.2}$$

Similar to the study time regression. What is needed for consistent estimates of the preference parameters of the model is a consistent estimate of the probability of grade promotion given enrollment. Estimation of equation (2.8.2) provides us with this. In principle, if the enrollment decision is correlated with the error term defining equation (2.8.2), then the coefficient estimates obtained would be biased and inconsistent and not conducive to direct interpretation. However, the inclusion of AFQT in the regression should at least mitigate the level of biased induced by regressing only on the subset of individuals that choose to enroll.

Another issue is the choice of separate regressions for the set of students who choose to work while enrolled in school and the set who choose not to work while enrolled in school. This main reason for this specification is to improve the flexibility of the resulting estimated transition probabilities. However, if the decision to work is correlated with the error term that defines equation (2.8.2), then the coefficient estimates are expected to be biased and inconsistent. The inclusion of our measure of labor market ability, the estimated fixed effects from the wage regression are included to reduce the bias of the estimated coefficients. At the very least however, the coefficients in equation (2.8.2) can certainly be interpreted for the relevant groups of individuals.

A third issue involves the appropriateness of including current period decision variables in equation (2.8.2). The theoretical model assumes that the individual makes his schooling and employment decisions $(d^s_{nt}, s_{nt}, d^h_{nt}, h_{nt})$ at the beginning of each period conditioned on the information set available to him at that point in time. The grade promotion probability function is known by the individual, and he has control over it in so far as he has control over the decision variables. However, the uncertainty is not resolved until the beginning of the following year. The timing of the model thus makes the period t decision variables predetermined in equation (2.8.2).

Table 7 reports the result of the logit regression of the probability of completing a grade and Table 8.1 reports the corresponding average derivatives. The standard errors reported are corrected for the inclusion of predicted study time. Computation of the corrected standard errors is complicated by the nonlinear specification of the study time function and the probability of grade transformation. The details are presented in Appendix 2 for completeness. The number of observations used in estimation for the two groups ($d^h_{nt} = 0$, and $d^h_{nt} = 1$) are 2216 and 5606, the Likelihood ratio statistics are 400.65 and 1350.78, and the Pseudo R^2's are 15% and 17%. Furthermore, all coefficients except for the constant term are significant at the 10% level, and slope parameters, except for 2 are significant at the 5% level. Note that some variables are dropped from estimation in either groups because of their low precision and statistical irrelevance.

The results in Table 8 indicate that lagged labor market participation decisions are positively correlated with the probability of grade promotion. This provides evidence for the congruence hypotheses. However, the effect is a lagged effect, and the interpretation varies slightly from that proposed by D'Amico [1984]. The decision to participate in the labor market in either of the last two periods increases the current probability of grade promotion by approximately 5%. The full model will have to simulated to see exactly how large this effect turns out to be on completed education. However, at this stage it is clear that a 5% increase in the probability of completing a grade level is a significant magnitude.

We find that blacks have a lower probability of being promoted a grade level than their white counterparts. For the group that works, hispanics also have a lower probability of being promoted than their white counterparts. This result is not simply the classical drop-out story of minorities. The interpretation of these coefficients are that: given two males, one black and the other white, with the same abilities (as measured by AFQT scores and the estimated fixed effects), the same hours studied, the same hours worked, and in the same grade level, along with other conditioned covariates, the black male has a significantly lower probability of being promoted from that grade level. To understand what may be driving this result, one must also look at what in not included in the regression, that is, what factors are not controlled for and may be correlated with race. The primary excluded factor in the regression would be the quality of the schools attended. It is well known that the quality of schools attended by blacks are on average lower than those attended by their white counterparts. I argue therefore that the negative coefficient of blacks in the grade advancement regression captures the lower schooling opportunities and qualities available to these racial groups. The quality of schooling is typically measured by, among other factors, the level of funding that school receives, class size, in particular the student-teacher ratio, and the socio-economics conditions of the community surrounding the school. The available data does not contain information on these measures of school quality. However, if one is only interested in the difference in schooling opportunities across races, as this study is, and not to identify the sources of these differences, then the estimated regression is sufficient.

The results in Table 8 also indicate that the probability of grade level promotion is increasing in time spent on schooling activities for both groups, and concave for the group that works. Conversely, this probability is decreasing and convex in hours spent in the labor market. Students in grades 11 and 12 have a larger probability of being promoted than college students.

2.9 CONDITIONAL CHOICE PROBABILITIES

Estimation of conditions characterizing labor supply and schooling decisions also requires that estimates of the conditional choice probabilities defined in equation (2.3.12). Inclusion of the individual-specific effect, and time-specific effects as explanatory variables allows us to treat the sample as pooled cross-section and time series data that is independently distributed over individual and time. This implies straightforward nonparametric estimation of (2.3.13).

To estimate the preference parameters, we also need to estimate the conditional choice probabilities conditional on all the states that remain feasible. This is done by taking advantage of the finite state dependence of the model. In particular, we need to estimate the probability that individual n chooses alternative j in period $t+i$ conditional on observing state k in that period $p_j(\Psi_{ntk}^{(i)})$. We achieve this by estimating the probability that an observationally equivalent individual chooses alternative j in the current period conditional on observing the state k in the current period. The validity of this method depend on the inclusion of the individual-specific effects and the time-specific effects in these regressions. These auxiliary CCP's are estimated using nonparametric techniques. The technical details of these estimators are outlined in Appendix A.1.3.

Table 10 presents the means and standard deviations of these estimated probabilities and the required derivatives. The sample average of the CCP's are equal to the sample average of their corresponding indicator functions with 4 decimal places. This indicates that the bias in these estimates are small. The relative magnitudes of the conditional state probabilities are also plausible. The probability that an individual chooses home production given that he enrolled in school last period and did not get promoted the grade level is larger than the probability of choosing home production if he was promoted.

The average derivatives of the conditional state probabilities are also empirically plausible. An additional hour of work in the past reduces the probability that the individual will choose home production in the current period. An additional hour of school activity in the past increases the probability of choosing home production in the current period if the individual did not get promoted the grade level. On the other hand, an additional hour of school activity in the past decreases the probability of choosing home production in the current period if the individual was promoted the grade level.

2.10 SCHOOLING, PARTICIPATION, AND HOURS

2.10.1 The moment conditions.

Estimation of the remaining parameters of the model makes use of an alternative representation of the conditional valuation function derived in Hotz and Miller [1993]. This requires that parametric restrictions be placed on the utility functions. Let the components of the utility of schooling, labor supply in equation (2.2.6), and utility of leisure in equation (2.2.7) take the form

$$u_1(x_{nt}, d^s_{nt}) = d^s_{nt} x'_{nt} B_5, \tag{2.10.1}$$

$$u_2(x_{nt}, d^h_{nt}) = d^h_{nt} x'_{nt} B_6, \tag{2.10.2}$$

$$u_3(x_{nt}, g_{nt}) = l_{nt} z'_{nt} B_7 + \sum_{i=0}^{\rho} \delta_i l_{nt} l_{nt-i}. \tag{2.10.3}$$

The utility of leisure is assumed to be quadratic. Economic theory suggests that the utility of leisure is concave in leisure, $\delta_0 < 0$. The parameters $\delta_i, i = 1, \cdots, \rho$ capture intertemporal nonseparabilities in the preference for leisure. For $i > 0$, $\delta_i < 0$ implies that current leisure and leisure lagged i periods are intertemporal substitutes. On the other hand, $\delta_i > 0$ implies that current leisure and leisure lagged i periods are intertemporal complements.

Define $\theta \equiv (B'_5, B'_6, B'_7, \delta_0, \cdots, \delta_\rho, \alpha)'$, $\gamma \equiv (B'_1, \cdots, B'_4)'$, $P \equiv (P_{nt0}, \cdots, P_{nt3})'$. Let F denote the set of conditional state probabilities and their relevant derivatives and let $\Theta \equiv (\theta', \gamma', P', F')'$. Define also $l^{(0)}_{nt} \equiv 1$, $l^{(1)}_{nt} \equiv 1 - h_{nt}$, $l^{(2)}_{nt} \equiv 1 - s_{nt}$, and $l^{(3)}_{nt} \equiv 1 - h_{nt} - s_{nt}$. By substituting these functional forms for the utility functions into the Euler condition for hours (2.3.16), we derive the following moment condition:

$$\begin{aligned} m_{nt1}(\Theta) \equiv\ & d_{nt1}\Big[\alpha^{-1}\eta_n\lambda_t w_{nt} - z'_{nt}B_5 - 2\delta_0 l^{(1)}_{nt} - \textstyle\sum_{i=1}^{\rho}\delta_i(l_{nt-i}+\beta^i) \\ & - \textstyle\sum_{i=1}^{\rho}\beta^i p_0(\Psi^{(i)}_{nt1})^{-1}\frac{\partial p_0(\Psi^{(i)}_{nt1})}{\partial h_{nt}}\Big] \\ & + d_{nt3}\Big[\alpha^{-1}\eta_n\lambda_t w_{nt} - z'_{nt}B_5 - 2\delta_0 l^{(3)}_{nt} - \textstyle\sum_{i=1}^{\rho}\delta_i(l_{nt-i}+\beta^i) \\ & - \textstyle\sum_{i=1}^{\rho}\beta^i\Big[p_0(\Psi^{(i)}_{nt4})^{-1}\frac{\partial p_0(\Psi^{(i)}_{nt4})}{\partial h_{nt}}F(x_{nt}) + p_0(\Psi^{(i)}_{nt5})^{-1}\frac{\partial p_0(\Psi^{(i)}_{nt5})}{\partial h_{nt}}(1-F(x_{nt})) \\ & + \ln\left(\frac{p_0(\Psi^{(i)}_{nt5})}{p_0(\Psi^{(i)}_{nt4})}\right)\frac{\partial F(x_{nt})}{\partial h_{nt}}\Big]\Big]. \end{aligned}$$

Likewise, we substitute the utility functions in to the optimality condition for study time (2.3.17) to obtain the following moment condition:

$$
\begin{aligned}
m_{nt2}(\Theta) \equiv \quad & d_{nt2}\Big[-z'_{nt}B_5 - 2\delta_0 l_{nt}^{(2)} - \textstyle\sum_{i=1}^{\rho}\delta_i(l_{nt-i}+\beta^i) \\
& -\textstyle\sum_{i=1}^{\rho}\beta^i\Big[p_0(\Psi_{nt2}^{(i)})^{-1}\frac{\partial p_0(\Psi_{nt2}^{(i)})}{\partial s_{nt}}F(x_{nt}) + p_0(\Psi_{nt3}^{(i)})^{-1}\frac{\partial p_0(\Psi_{nt3}^{(i)})}{\partial s_{nt}}(1-F(x_{nt})) \\
& +\ln\left(\frac{p_0(\Psi_{nt3}^{(i)})}{p_0(\Psi_{nt2}^{(i)})}\right)\frac{\partial F(x_{nt})}{\partial s_{nt}}\Big]\Big] + d_{nt3}\Big[-z'_{nt}B_5 - 2\delta_0 l_{nt}^{(3)} - \textstyle\sum_{i=1}^{\rho}\delta_i(l_{nt-i}+\beta^i) \\
& -\textstyle\sum_{i=1}^{\rho}\beta^i\Big[p_0(\Psi_{nt4}^{(i)})^{-1}\frac{\partial p_0(\Psi_{nt4}^{(i)})}{\partial s_{nt}}F(x_{nt}) + p_0(\Psi_{nt5}^{(i)})^{-1}\frac{\partial p_0(\Psi_{nt5}^{(i)})}{\partial s_{nt}}(1-F(x_{nt})) \\
& +\ln\left(\frac{p_0(\Psi_{nt5}^{(i)})}{p_0(\Psi_{nt4}^{(i)})}\right)\frac{\partial F(x_{nt})}{\partial s_{nt}}\Big]\Big].
\end{aligned}
$$

Additional moment conditions are formed from the optimal discrete choice conditions in equation (2.3.9). In particular, we obtain the following moment conditions from the optimality condition for choosing alternatives 1,2, and 3:

$$
\begin{aligned}
m_{nt3}(\Theta) \equiv \quad & d_{nt1}\Big[\ln\left(\frac{p_{nt1}}{p_{nt0}}\right) - x'_{nt}B_6 + x'_{nt}B_7(l_{nt}^{(0)} - l_{nt}^{(1)}) + \delta_0(l_{nt}^{(0)2} - l_{nt}^{(1)2}) \\
& +\textstyle\sum_{i=1}^{\rho}\delta_i(l_{nt}^{(0)} - l_{nt}^{(1)})(l_{nt-i}+\beta^i) - \frac{\eta_n\lambda_t}{\alpha}(w_{nt}h_{nt}) - \textstyle\sum_{i=1}^{\rho}\beta^i\ln\left(\frac{p_0(\Psi_{nt0}^{(s)})}{p_0(\Psi_{nt1}^{(s)})}\right)\Big],
\end{aligned}
$$

$$
\begin{aligned}
m_{nt4}(\Theta) \equiv \quad & d_{nt2}\Big[\ln\left(\frac{p_{nt2}}{p_{nt0}}\right) - x'_{nt}B_5 + x'_{nt}B_7(l_{nt}^{(0)} - l_{nt}^{(2)}) \\
& +\delta_0(l_{nt}^{(0)2} - l_{nt}^{(2)2}) + \textstyle\sum_{i=1}^{\rho}\delta_i(l_{nt}^{(0)} - l_{nt}^{(2)})(l_{nt-i}+\beta^i) + \frac{\eta_n\lambda_t}{\alpha}\pi_{nt} \\
& -\textstyle\sum_{i=1}^{\rho}\beta^i\Big[\ln p_0(\Psi_{nt0}^{(i)}) - \ln p_0(\Psi_{nt2}^{(i)})F(x_{nt}) - \ln p_0(\Psi_{nt3}^{(i)})(1-F(x_{nt}))\Big]\Big],
\end{aligned}
$$

$$
\begin{aligned}
m_{nt5}(\Theta) \equiv \quad & d_{nt3}\Big[\ln\left(\frac{p_{nt3}}{p_{nt0}}\right) - x'_{nt}B_5 - x'_{nt}B_6 + x'_{nt}B_7(l_{nt}^{(0)} - l_{nt}^{(3)}) \\
& +\delta_0(l_{nt}^{(0)2} - l_{nt}^{(3)2}) + \textstyle\sum_{i=1}^{\rho}\delta_i(l_{nt}^{(0)} - l_{nt}^{(3)})(l_{nt-i}+\beta^i) - \frac{\eta_n\lambda_t}{\alpha}(w_{nt}h_{nt} - \pi_{nt}) \\
& -\textstyle\sum_{i=1}^{\rho}\beta^i\Big[\ln p_0(\Psi_{nt0}^{(i)}) - \ln p_0(\Psi_{nt4}^{(i)})F(x_{nt}) - \ln p_0(\Psi_{nt5}^{(i)})(1-F(x_{nt}))\Big]\Big].
\end{aligned}
$$

Define $m_{nt}(\Theta) \equiv (m_{nt1}(\Theta), \cdots, m_{nt5}(\Theta))'$ and let $\overline{T}$ denote the set of periods for which the working and schooling hours, enrollment and participation conditions are valid. Let $m_n \equiv (m'_{n1}, \cdots, m'_{n\overline{T}}$ denote the vector of empirical moments for a given individual over time. We further define the weighting matrix $\Omega \equiv E[m_n, m'_n]$ and note that this matrix is block diagonal since $E_t[m_{nt}m_{ns}] = 0$ for $s < t$.

In order to increase the finite sample precision of the estimates of the remaining parameters of the model, we implement a iterated GMM (GMMI) variation of the Nested Pseudo Likelihood

Algorithm (NPL) proposed by Aguirregabiria and Mira [2002]. This algorithm consists of two steps. The first step is where GMMI is implemented to obtain estimates of the preference parameters, give an initial estimated of the CCP's. The second step is where the CCP's are updated using the estimates of the preference parameters. To be precise, define $\Theta_1^k \equiv (\theta', \hat{\gamma}', (P^k)', \hat{F}')'$, and $\Theta_2^k \equiv ((\theta^k)', \hat{\gamma}', P', \hat{F}')'$. At iteration $K \geq 1$ of the outer algorithm, we apply the following steps

Step 1: Obtain new estimates of θ, θ^K, from the following iteration in $j \geq 1$:

$$\theta^{Kj} = \arg\max_{\theta \in \Theta} \sum_{n=1}^{N} \left[m_n(\Theta_1^{K-1}) \right]' (\Omega^{j-1})^{-1} \left[m_n(\Theta_1^{K-1}) \right], \tag{2.10.4}$$

where Ω^{j-1} is the weighting matrix evaluated at Θ_1^{K-1}, in which $\theta = \theta^{K,j-1}$. This iteration is repeated until convergence in θ is achieved, which is denoted θ^K

Step 2: Update P using the estimates θ^K as follows:

$$\begin{aligned} P_j^K &= \exp\left(V_j(\Theta_2^K) - V_0(\Theta_2^K)\right) P_0^{K-1} \\ &= \exp\left(m_{j+2}(\Theta_2^K)\right) P_0^{K-1}, \qquad j \geq 1, \\ P_0^K &= 1 - \textstyle\sum_{j=1}^{J} P_j^K. \end{aligned} \tag{2.10.5}$$

Iterate in K until convergence in P and θ is reached.

The convergence of the CCP's is stated in Proposition 1 of Aguirregabiria and Mira [2002], while the convergence of the GMMI is discussed in Hansen et al. [1996]. From our experience, it seems that the iteration in step 1 of the algorithm improves greatly the stability of the overall algorithm.

The nature of the iteration in the CCP's along with the inclusion of the pre-estimates $(\hat{\gamma}, \hat{F})'$ make the correct standard errors of the estimates of θ nonstandard. To derive the correct standard errors, we implement the technique proposed in Newey and McFadden [1994] and Newey [1994]. Interestingly, because of the structure of the state space in the model, repeated use of the law of iterated expectations results in significant simplification of the asymptotic variance. In particular, no post estimation is required to correct the standard errors. This greatly reduces the computational burden of the CCP estimator. The key effect of the iteration in the CCP's is an alternative specific re-weighting of the influence functions of the pre-estimators. This re-weighting is such that a larger weight is assigned to alternatives with a higher probability of occurring. The asymptotic properties of this estimator are discussed in appendix A.1.5.

2.10.2 Consumption Value of School Attendance

Table 11 reports the estimated psychic value of enrollment. The results indicate that the consumption value of schooling is increasing and concave in the level of education. For a given age, the consumption value is decreasing in level of education. These signs capture the decreasing rate of enrollment in school for higher levels of education and older individuals.

The coefficients on *BLACK* and *HISPANIC* in the consumption value of schooling are positive by not significantly different from zero. This result holds with and without the inclusion of AFQT. This implies that after controlling for racial differences in wages, hours worked, time spent of schooling, and school quality, black and Hispanic males are no more likely to enroll in school than their white counterparts.

2.10.3 Fixed Utility of Participation

Table 12 presents the estimate fixed utility of participating in the labor market. We find that the consumption value of labor force participation is increasing and concave in the level of labor market experience. However, these coefficients are imprecisely estimate. We find also that for a given age, the consumption value of labor force participation is decreasing in the level of labor market experience. The coefficients on *BLACK* and *HISPANIC* in the consumption value of labor force participation are negative, but imprecisely estimated. This results the racial disparity in the employment rates is not explained by differences in the propensity of participate in the labor market.

2.10.4 Utility of Leisure

The estimates of the utility of leisure are reported in Table 13. The results indicate that the utility of leisure is (weakly) decreasing and convex in age. This results is also found in Altug and Miller [1998] and Gayle and Miller [2003]. The results also indicate that the utility of leisure in increasing and concave in leisure. However, the parameter capturing the concavity is imprecisely estimated. We find also that the coefficients on the black and Hispanic indictors are not statistically different from zero. In other words we find no evidence of racial differences in the utility of leisure. In other words, the observed racial differences in hours worked and study time are not explained by racial differences in the preferences for leisure.

2.10.5 Intertemporal Nonseparabilities in Leisure

The results in table 14 indicate that preferences are intertemporally nonseparable in leisure. The positive coefficients on the interaction between current and lagged leisure in the utility of leisure indicate that for males in the sample, current and future leisure are complements in intertemporal preferences. This indicates a habit formation pattern where increases in current hours worked decreases the future marginal disutility of work. Likewise, increases in current hours spent on school activities decreases the future marginal disutility of studying.

Intertemporal nonseparabilities in leisure is estimated in, among others, Eckstein and Wolpin [1989], Miller and Sanders [1997], Altug and Miller [1998], and Gayle and Miller [2003]. The results concerning the intertemporal substitutability of complimentarity of leisure varies across these studies. Altug and Miller [1998] conjecture that employing data sampled over shorter time intervals result in the finding of complementarity between current and past leisure choices, while data sampled over longer (yearly) intervals result in the finding of substitutability between current and past leisure. However, the results in table 11 run in contrast to this conjecture, since in this study, hours are measured annually.

2.11 SOLUTION AND SIMULATION EXERCISES

2.11.1 Solving the model

Given the estimated parameters, the model is solved by means of backward induction from age 65 to age 15. Ideally, we would like to treat hours worked and studied completely symmetrical, as done in the estimation. However, solving for both hours worked and studied on a fine enough grid is infeasible. To bypass this problem, we use the estimated function for study time to approximate optimal study time in the solution. This approximation makes solution of the model tractable. However, this function is valid only for males that choose to enroll. While this was not a problem for estimating the model, it may cause biases in the simulation results.

With the use of the study time function, optimal hours can then be solved for on a fine grid. The problem of interpolating off this grid then arises. Interpolation is carried out by a third order polynomial regression of the value at each point of the grid on the corresponding state space. The parametric regression is preferred over nonparametric kernel techniques because it allows for a finer grid on hours and avoids the corresponding curse of dimensionality that nonparametric techniques face. In solving the baseline model, the smallest R^2 at age 40 is 0.994, indicating that the third order polynomial approximation is expected to provide very precise approximations of the value functions off the grid of hours. We also assume in the solution that nobody enrolls in school after the age of 36. This is justified as in the data only a very small fraction of the sample enrolls in school past the age of 36.

The baseline model is solved assuming that the economy is in equilibrium where aggregate components grow at an equilibrium rate of the average in the sample period. These aggregate components are the shadow price of consumption, the skill specific piece rates, and tuition costs. The assumption of zero growth rate in aggregate skill prices would result in unrealistic predictions of wages over the life cycle. The baseline model is solved for 10,000 replications separately for whites, blacks, and Hispanics. Table 14 reports the baseline simulation by age along with the corresponding sample averages from the data. The baseline model under-predicts the level of labor market experience and the average hourly wage rate. It may be possible to improve the fit of the model to the data by adding dummies to capture the large drop-off in enrollment and increase in working of 18 and 19 year old males that is found in the data. However, there is no economic intuition for such dummies, and they are not necessary for the analysis to come. Furthermore, given that we we do not have the full profile of the growth in the aggregate variables, the simulation results are not expected to closely fit the sample averages at any rate. Not withstanding this, the model predicts remarkably well the general patterns within each race group. Moreover, the model also gets exactly the relative patterns in the reported outcomes across races.

The first two counterfactual simulations performed evaluate policies that are aimed at affecting working while enrolled in school. First the government subsidizes individuals who choose to enroll in school and not participate in the labor market. Second the government increases the school curriculum so that individuals who enroll in school necessarily spend more time on school activities. The Third set of counterfactual simulations addresses the issue of equating the quality of schooling across races. The final set addresses the issue an increase in time spent on school activities when school quality is held constant across races.

2.11.2 Cash Subsidy

For the first counterfactual simulation exercise, we consider a subsidy of 1000 dollars, which grows yearly at the same rate as the aggregate component of the marginal utility of consumption (which is the same as the growth rate in tuition). The results from this simulation exercise are reported in table 15 under the column labeled "Pol. 1". The baseline simulation results are included for comparison under the column labeled "Base".

The results indicate that this policy does very little in affecting the outcomes of young men. We see very modest increases in education, and reductions in experience. There are also modest overall increases in wages due to this policy. The effect of the policy is the same for all races.

2.11.3 Increased time spent on school activities

In practice, the second policy can be achieved by increasing the number of hours school is in session for, summer classes, or Saturday (or Sunday) classes. This can also be achieved my increasing the number of, or level of difficulty of homework assignments and projects. In the simulation exercise, this policy is achieved by increasing the study time function. The amount by which the constant is increased is chosen to make the magnitudes of this policy and the subsidy policy above comparable. In particular, if at age 16, the individual was to work for $1000 at $4 hourly wage rate, he would work for 250 hours. The study time function is therefore increased by 250. Since the average wage at age 16 in the baseline simulations is approximately $3.50, the results from this simulation are considered to be lower bound comparisons to the above simulation exercise. The findings are reported in table 16 under the columns labeled "Pol 2".

The findings indicate that this policy significantly increases education and wages for white and black men, with moderate increases for Hispanics. By the age 35, the completed level of education increases by 15% for whites and 12.3% for blacks, but only by 1% for Hispanics. Also we find that the level of labor market experience for whites and black decreases as a result of the policy, while it increases for Hispanics.

Analysis of the change in the choices young men make due to the policy shows that Hispanics are the least responsive. Further more, while the fraction of the population that enroll in school and not work increase significantly for whites and black (21.8% and 16.4%), it increases only modestly for Hispanics (1%). Another difference in the patterns of choices is that while the faction of the white population that works and attends school decreases (by 5%), it increases for blacks (0.5%) and Hispanics (1.8%). Furthermore, Hispanics are the only group in which the decline in young men where the percentage increase in those working and attending school outweighs the percentage decline in those who choose to exclusively participate in the labor market.

The conclusion therefore is that the crowding-out hypothesis holds most significantly for whites, followed by blacks and Hispanics. This conclusion comes from the fact that a mandatory increase in the time spent on school activities has the most significant negative effect on the employment rate of whites, and the most significant positive effect on completed education and future wages of whites. This result is also empirically bolstered by the fact that in the data a larger fraction of whites enroll in school and work at the same time. Hence intuitively, one would expect that they may be most subject to the crowding out effect of working while attending school. Hence policies that are aimed at increasing the time students spend on school activities has significant positive effects on whites and blacks, but less so on Hispanics.

2.11.4 Equating school quality

The next policy experiment equalizes the quality of schools across races. Technically, this is done by setting the coefficients of *BLACK* and *HISPANIC* in the grade transition probability equation to zero. The results from this exercise are presented in table 15 under the columns labeled “Pol 3”. We also present the results from the baseline simulation under the columns labeled “Base”.

The results in table 17 indicate that the policy has significant impacts on both blacks and Hispanics. For blacks, by the age of 35, the completed level of education increases by 11%, the years of labor market experience increase by 1%, and the hourly wage rate increases by 15%. For Hispanics, by the age of 35, the completed level of education increases by 7%, the years of labor market experience increase by 3%, and the hourly wage rate increases by 4%.

For both blacks and Hispanics, the policy has the effect of increasing enrollment rates. However, the pattern of enrollment is quite different for both groups. For blacks, the policy has an effect of increasing the fraction of those who enroll exclusively in school by 12%, and 13% for those who enroll and work. For Hispanics however, the policy only increases the fraction of those who enroll exclusively in school by 2%, but by 14% for those who enroll and work. Since the chances of completing a grade level is smaller if the student is also working, this results in a more modest increase in completed education, and thus a more modest increase in hourly wage rate.

We conclude therefore that policies aimed at improving the quality of schools for minorities results in significantly increased education for both groups, but a more modest increase in hourly wage rates for Hispanics.

2.11.5 Equating school quality and increasing time spent on school activities

Given that equating school quality results in a significant increase in the education level of Hispanics, it is interesting to know if the magnitude of the effect of an increase in study time changes in magnitude under this new environment. Therefore we simulate this environment and the results are reported in table 18 under the columns labeled "Pol 4". Again, the baseline simulation results are presented for comparison under the columns labeled "Base".

The results under the new environment, the choices and outcomes for Hispanics are far more responsive to the exogenous increase in study time. The simulated completed level of education increases by 23% and the hourly wage rate increases by 29% for Hispanics by age 35. Furthermore, the fraction of the Hispanic population that exclusively enroll in school increase by 23% and the fraction that enroll and work increase by 18%. Thus under the environment where the quality of schools are equated across race, the responsiveness of Hispanics to an exogenous increase in study time increases significantly.

For blacks in this new environment, the exogenous increase in study time increases the completed level of education by 28% and the hourly wage rate by 79% by age 35. The fraction on blacks that enroll in school exclusively increases by 38%, and the fraction that enrolls and work increases by 8%.

These results indicate that policies aimed that increasing the time spent on school activities has a positive effect on minority students; magnitudes that are comparable to their white counterparts.

2.12 CONCLUSIONS

The paper has developed and estimated a dynamic structural model of educational attainment and labor supply. The main focus of the analysis has been to study the allocation of time between labor supply, formal schooling activities and leisure, both within a year and over the life cycle. The model allows for skill specific productivity and piece rates, as well as intertemporal nonseparabilities in the utility of leisure. It also allows for racial variation in wages, consumption, school quality, study patterns, the fixed cost of labor market participation, the fixed utility of schooling, and the utility of leisure. The estimated results indicate that current and future leisure choices are intertemporal complements. The results also indicate that the observed racial differences in outcomes come from a variety of sources that interact in a highly nonlinear fashion, but not from racial differences in tastes.

The estimated model is used to evaluate two policies that are aimed at affecting the allocation of time between schooling and working. The first policy subsidizes young students that do not participate in the labor market. The results indicate that this subsidy does little in changing the patterns of enrollment and labor supply on either the extensive or the intensive side. The second policy increases the school curriculum so that young men who choose to enroll in school necessarily spend more time on schooling activities. The results indicate that this policy would have significant positive effects on white and blacks, but more modest effects on Hispanics.

A third exercise was performed to evaluate the effects of equating school qualities of blacks and Hispanics to that of whites. The results indicate that such a policy would have a large positive effect on education and wages for blacks, but a smaller positive effect on Hispanics. We also show that under this environment, Hispanics become significantly more responsive to policies aimed at increasing the school curriculum.

This study was motivated by the increasing number of students that decide to also participate in the labor market. The results indicate that the effect of this trend varies across races. Policy focused on changing this trend to improve the level of education and labor market outcomes may have only modest effects on some racial groups. As a matter of policy, the results indicate that equating school quality across races may be a more productive first step for improving the outcomes of minorities. Of course, our measure of school quality is agnostic about exactly what are the parameters in the school system that needs to be addressed. This would require an understanding of the key variables that affect students' grade promotion probabilities.

One of the main limitations of the model presented in this paper is that it is set in a partial equilibrium framework. In a general equilibrium framework, one would expect that the aggregate skill specific wages will also be affected by a policy that changes the distribution of the labor force over these groups. A policy that increases the level of education will result in more labor supplied to the high skilled sector and less to the low skilled sector. In a general equilibrium framework, this will drive down the price of high skilled labor and push up the price of low skilled labor, thus reducing the incentive to acquire higher education. Since this general equilibrium effect is not accounted for in the model presented in this paper, the effects of policies that increase the level of education may be overstated. How far the partial equilibrium effects are from the general equilibrium effects is an important issue for future research.

Table 1: List and Description of Variables Used

d^s_{nt}	Indicator variable equal to 1 if individual n enrolls in year t
d^s_{nt}	Indicator variable equal to 1 if individual n works in year t
s_{nt}	Fraction of time spent on school activities in year t
h_{nt}	Fraction of time spent working in year t
S_{nt}	Completed level of education
E_{nt}	Level of experience
AGE_{nt}	Age at year t
$WHITE$	Indicator variable equal to 1 if White and 0 otherwise
$BLACK$	Indicator variable equal to 1 if Black and 0 otherwise
$HISPANIC$	Indicator variable equal to 1 if Hispanic and 0 otherwise
FAM_INC_{nt}	level of family income at year t
FAM_SIZE_{nt}	size of n's household at year t
FAM_AGE_{nt}	average age of n's household at year t
$SIBLINGS$	number of siblings of n as at age 14
US_BORN	indicator variable equal to 1 if n was born in the US
$AFQT$	The Armed Force Qualification Test score for individual n
$ASSETS$	Level of asset holdings by the household of n in year t
$UNEMP$	Level of the unemployment rate local to n in year t
$RURAL$	Indicator variable equal to 1 if n lives in a rural area in year t
$TUITION$	Level of college tuition that individual n is subject to in year t

Table 2: Summary Statistics

Year	1979	1980	1981	1982	1983	1984	1985	1986
Observations	3749	3512	3595	3575	3594	3549	3504	3413
d_0	0.0205	0.0529	0.1115	0.1325	0.1719	0.1541	0.1435	0.1300
d_1	0.0381	0.1452	0.2842	0.4215	0.5158	0.6198	0.6889	0.7380
d_2	0.5644	0.3809	0.2439	0.1367	0.0951	0.0617	0.0345	0.0240
d_3	0.3769	0.4208	0.3602	0.3090	0.2170	0.1642	0.1329	0.1078
d^s	0.9413	0.8018	0.6041	0.4458	0.3121	0.2259	0.1675	0.1318
s	1436.5	1354.6	1276.0	1203.3	1149.7	1139.3	1114.6	1077.3
S	9.7967	10.730	11.335	11.842	12.198	12.416	12.578	12.708
d^h	0.4150	0.5660	0.6445	0.7306	0.7328	0.7841	0.8219	0.8458
h	710.90	972.82	1080.5	1159.8	1310.0	1477.6	1577.7	1694.5
E	1.2107	1.6136	2.1655	2.8036	3.5166	4.2310	4.9877	5.8025
w^1	4.3872	4.1601	4.3383	4.6541	4.8560	5.1220	5.5749	6.0788
AGE	16.743	17.653	18.695	19.697	20.706	21.699	22.690	23.688
$WHITE$	0.5727	0.5769	0.5713	0.5757	0.5759	0.5711	0.5736	0.5722
$BLACK$	0.2625	0.2640	0.2651	0.2626	0.2613	0.2646	0.2606	0.2625
$HISPANIC$	0.2648	0.1592	0.1635	0.1617	0.1627	0.1643	0.1658	0.1653
FAM_INC^1	17647	19086	20011	21168	21398	21785	23577	25319
FAM_SIZE	4.8434	4.5948	4.3171	3.9625	3.7045	3.3722	3.1726	2.9856
FAM_AGE	26.225	26.823	26.978	26.665	26.699	26.653	26.538	26.175
$SIBLINGS$	3.6220	3.5899	3.6069	3.6204	3.6165	3.6238	3.6204	3.6024
US_BORN	0.9306	0.9328	0.9310	0.9311	0.9315	0.9323	0.9326	0.9326
$AFQT$	42.024	43.186	42.793	42.835	42.774	42.606	42.545	42.565
$ASSETS^1$						4141.2	4278.8	4998.8
$UNEMP$	2.5646	2.8476	3.1652	3.7848	4.1978	3.4356	3.2919	3.1693
$RURAL$	0.2125	0.20871	0.1997	0.1932	0.1830	0.1718	0.1680	0.1614
$TUITION^1$	813.19	793.04	809.79	865.54	916.18	960.77	1029.0	1087.4

[1] In 1981 dollars

Table 3: Summary Statistics (Contd.)

Year	1987	1988	1989	1990	1991	1992	1993	1994
Observations	3338	3357	3389	3328	2931	2936	2937	2896
d_0	0.1207	0.0965	0.0994	0.0943	0.1044	0.1226	0.1113	0.1142
d_1	0.8001	0.8394	0.8574	0.8647	0.8614	0.8474	0.8593	0.8649
d_2	0.0155	0.0071	0.0023	0.0006	0	0	0	0
d_3	0.0635	0.0568	0.0407	0.0402	0.0341	0.0299	0.0292	0.0207
d^s	0.0790	0.0640	0.0430	0.0408	0.0341	0.0299	0.0292	0.0207
s	1043.2	977.74	970.54	962.93	976.60	1006.7	1118.6	1128.3
S	12.833	12.890	12.917	12.962	13.050	13.049	13.073	13.08
d^h	0.8636	0.8963	0.8982	0.9050	0.8955	0.8773	0.8886	0.8857
h	1836.4	2016.8	2078.7	2025.0	2072.1	2126.6	2076.2	2111.7
E	6.6363	7.4566	8.2912	9.1908	10.022	10.853	11.676	12.548
w^1	7.0968	7.6098	7.6038	8.0964	7.7159	7.8402	8.2973	8.4466
AGE	24.680	25.684	26.686	27.687	28.624	29.620	30.621	31.611
WHITE	0.5733	0.5737	0.5716	0.5736	0.5165	0.5150	0.5138	0.5162
BLACK	0.2657	0.2654	0.2653	0.2644	0.2972	0.2973	0.3006	0.2987
HISPANIC	0.1609	0.1609	0.1632	0.1620	0.1863	0.1877	0.1856	0.1851
FAM_INC[1]	26572	29047	46666	34705	36938	59830	41624	43778
FAM_SIZE	2.8406	2.7768	2.7722	2.7641	2.8161	2.8692	2.9240	2.9229
FAM_AGE	26.154	25.624	25.707	25.814	26.108	26.231	24.292	24.610
SIBLINGS	3.6096	3.6136	3.6208	3.6283	3.6349	3.6294	3.6275	3.6339
US_BORN	0.9340	0.9368	0.9350	0.9353	0.9344	0.9335	0.9342	0.9350
AFQT	42.789	42.565	42.270	42.422	42.089	41.905	41.869	41.965
ASSETS[1]	7107.8	7132.9	20246	10064	11688	13922	13488	12195
UNEMP	2.9331	2.6094	2.3865	2.4002	2.9512	3.1757	3	2.9499
RURAL	0.1791	0.1805	0.1844	0.1850	0.1641	0.1665	0.1722	0.1833
TUITION[1]	1153.1	1170.5	1181.5	1234.6	1351.1	1404.9	1490.2	1504.5

[1] In 1981 dollars

Table 4: The Consumption Equation.

Variable	Parameter	Estimate	Std. Err.
Demographic Variables			
$\Delta FAM\ SIZE_{nt}$	$B_{1,1}$	0.1466	0.0022
$\Delta FAM\ INC_{nt}$	$B_{1,2}$	8.00E-06	0.08E-06
$\Delta FAM\ AGE_{nt}$	$B_{1,3}$	4.00E-06	2.00E-06
$\Delta UNEMP_{nt}$	$B_{1,4}$	-0.0010	0.0005
ΔS_{nt}	$B_{1,5}$	-0.0091	0.0008
$\Delta(AGE \times S_{nt})$	$B_{1,6}$	0.0089	0.0008
ΔAGE^2_{nt}	$B_{1,7}$	-0.0008	0.0004

Table 5: The Wage Equation.

	Low Skill		High Skill	
Variable	Parameter	Estimate	Parameter	Estimate
Lags of Enrollment				
Δd^s_{nt-1}	$B_{2,1,1}$	-0.0309	$B_{2,2,1}$	-0.0701
		(0.0382)		(0.0266)
Δd^s_{nt-2}	$B_{2,1,2}$	-0.0198	$B_{2,2,2}$	-0.01239
		(0.0421)		(0.02707)
Lags of Participation				
Δd^h_{nt-1}	$B_{2,1,3}$	0.0198	$B_{2,2,3}$	-0.1513
		(0.0431)		(0.0175)
Δd^h_{nt-2}	$B_{2,1,4}$	0.0319	$B_{2,2,4}$	-0.1272
		(0.0460)		(0.0193)
Lags of Hours Worked				
Δh_{nt-1}	$B_{2,1,5}$	0.20E-04	$B_{2,2,5}$	0.28E-04
		(0.02E-04)		(0.01E-04)
Δh_{nt-2}	$B_{2,1,6}$	0.07E-04	$B_{2,2,6}$	0.10E-04
		(0.02E-04)		(0.01E-04)
Socio-Economic Variables				
ΔS^2_{nt}	$B_{2,1,8}$	-0.29E-04	$B_{2,2,8}$	0.0040
		(1.37E-04)		(0.0001)
ΔE^2_{nt-2}	$B_{2,1,7}$	-0.0010	$B_{2,2,7}$	-0.0011
		(0.0003)		(0.0002)
$\Delta(S_{nt} \times E_{nt-2})$	$B_{2,1,9}$	0.0027	$B_{2,2,9}$	-0.0072
		(0.0003)		(0.0002)

Table 6: Estimated changes in aggregate prices and wages[1]

Year	Aggregate Prices $(1-\alpha)^{-1}\Delta \ln(\lambda_t)$	Aggregate Wages Unskilled $(\Delta \ln \omega_{t,1})$	Skilled $(\Delta \ln \omega_{t,2})$
1984	0.0345	0.0287	0.1127
	(0.0199)	(0.0393)	(0.0162)
1985	-0.0423	0.0449	0.2320
	(0.0200)	(0.0381)	(0.0177)
1986	0.0288	0.0526	0.2303
	(0.0206)	(0.0402)	(0.0204)
1987	0.0713	0.0584	0.2831
	(0.0218)	(0.0384)	(0.0212)
1988	-0.0102	0.0556	0.1421
	(0.0226)	(0.0363)	(0.0210)
1989	0.1111	-0.0228	0.1781
	(0.0228)	(0.0375)	(0.0221)
1990	-0.0186	0.0133	0.1652
	(0.0232)	(0.0366)	(0.0219)
1991	0.0230	-0.0360	0.1610
	(0.0237)	(0.0368)	(0.0219)
1992	0.2044	-0.0101	0.1713
	(0.0246)	(0.0392)	(0.0237)
1993	-0.0260	0.0290	0.1770
	(0.0250)	(0.0411)	(0.0252)
1994	-0.0056	0.0120	0.1587
	(0.0251)	(0.0351)	(0.0218)

[1] Standard errors in parentheses

Table 7: Estimate of time spent on school activities.

Variable	Parameter	Estimate	Std.Err
Constant	$B_{3,0}$	7.2383	0.1829
Lags of Enrollment			
d^s_{nt-1}	$B_{3,1}$	0.2602	0.0463
d^s_{nt-2}	$B_{3,2}$	0.2037	0.0789
Lags of Hours Worked			
h_{nt-1}	$B_{3,3}$	-0.77E-04	0.17E-04
h_{nt-2}	$B_{3,4}$	-0.50E-04	0.26E-04
Socio-Economic Variables			
$BLACK$	$B_{3,5}$	0.1063	0.0265
$HISPANIC$	$B_{3,6}$	-0.0996	0.0304
$AGE_{nt} \times S_{nt}$	$B_{3,7}$	-0.0045	0.0013
$(AGE_{nt} \times S_{nt})^2$	$B_{3,8}$	0.76E-05	0.26E-05
$US\ BORN$	$B_{3,10}$	-0.1261	0.0417
$FAM\ SIZE_{nt}$	$B_{3,11}$	0.0135	0.0050
$RURAL$	$B_{3,12}$	0.0647	0.0250
$UNEMP_{nt}$	$B_{3,13}$	-0.0244	0.0100
$AFQT$	$B_{3,15}$	0.0037	0.0004
$\ln(\mu)$	$B_{3,17}$	-0.1435	0.0273

Table 8: Probability of Grade Promotion.

	$d_{nt}^h=0$		$d_{nt}^h=1$	
Variable	Parameter	Estimate	Parameter	Estimate
Constant	$B_{4,1,0}$	0.0307	$B_{4,2,0}$	0.0482
		(0.7734)		(0.5499)
Time Use Variables				
s_{nt}	$B_{4,1,1}$	0.0025	$B_{4,2,1}$	0.0036
		(0.0003)		(0.0008)
s_{nt}^2			$B_{4,2,2}$	-0.15E-05
				(0.03E-05)
h_{nt}			$B_{4,2,3}$	-0.0006
				(0.0001)
h_{nt}^2			$B_{4,2,4}$	0.10E-06
				(0.03E-06)
Participation Variables				
d_{nt-1}^h			$B_{4,2,8}$	0.2185
				(0.0873)
d_{nt-2}^h	$B_{4,1,4}$	0.2771		
		(0.1203)		
Socio-Economic Variables				
$BLACK$	$B_{4,1,5}$	-0.2296	$B_{4,2,9}$	-0.3751
		(0.1305)		(0.0925)
$HISPANIC$			$B_{4,2,10}$	-0.4627
				(0.0928)
AGE_{nt}	$B_{4,1,6}$	-0.1468	$B_{4,2,11}$	-0.0824
		(0.0268)		(0.0147)
$AFQT$	$B_{4,1,7}$	0.0058	$B_{4,2,13}$	0.0100
		(0.0027)		(0.0017)

Table 9: Marginal Effects Probability of Grade Promotion.

	$d^h_{nt} = 0$		$d^h_{nt} = 1$	
Variable	Parameter	Estimate	Parameter	Estimate
Time Use Variables				
s_{nt}	$B_{4,1,1}$	0.0005	$B_{4,2,1}$	0.0008
s^2_{nt}			$B_{4,2,2}$	-0.26E-06
h_{nt}			$B_{4,2,3}$	-0.0001
h^2_{nt}			$B_{4,2,4}$	0.02E-06
Enrollment Variables				
d^s_{nt-1}			$B_{4,2,5}$	0.0915
GRADE 11	$B_{4,1,2}$	0.1136	$B_{4,2,6}$	0.0717
GRADE 12	$B_{4,1,3}$	0.1109	$B_{4,2,7}$	0.2235
Participation Variables				
d^h_{nt-1}			$B_{4,2,8}$	0.0487
d^h_{nt-2}	$B_{4,1,4}$	0.0542		
Socio-Economic Variables				
$BLACK$	$B_{4,1,5}$	-0.0449	$B_{4,2,9}$	-0.0837
$HISPANIC$			$B_{4,2,10}$	-0.1032
AGE_{nt}	$B_{4,1,6}$	-0.0365	$B_{4,2,11}$	-0.0184
S_{nt}			$B_{4,2,12}$	-0.0232
$AFQT$	$B_{4,1,7}$	0.0011	$B_{4,2,13}$	0.0022
$\ln(\eta)$	$B_{4,1,8}$	-0.1247	$B_{4,2,14}$	-0.0985
$\ln(\mu)$	$B_{4,1,9}$	-0.0508	$B_{4,2,15}$	-0.1216

Table 10: Sample Averages of Nonparametric Estimates

Variable	Sample Mean	Sample Std. Dev.	Variable	Sample Mean	Sample Std. Dev
p_{nt0}	0.1197	0.2145	$\frac{\partial p_0(\Psi^{(1)}_{nt1})}{\partial h_{nt}}$	-0.1988	2.0533
p_{nt1}	0.7076	0.3427	$\frac{\partial p_0(\Psi^{(2)}_{nt1})}{\partial h_{nt}}$	-0.3520	5.2983
p_{nt2}	0.0489	0.1303	$\frac{\partial p_0(\Psi^{(1)}_{nt4})}{\partial h_{nt}}$	-0.6092	4.5189
p_{nt3}	0.1232	0.2307	$\frac{\partial p_0(\Psi^{(1)}_{nt5})}{\partial h_{nt}}$	-0.5044	5.2893
$p_0(\Psi^{(1)}_{nt0})$	0.3870	0.2398	$\frac{\partial p_0(\Psi^{(1)}_{nt2})}{\partial s_{nt}}$	-0.0391	4.1811
$p_0(\Psi^{(2)}_{nt0})$	0.5709	0.1835	$\frac{\partial p_0(\Psi^{(1)}_{nt3})}{\partial s_{nt}}$	0.4081	6.9457
$p_0(\Psi^{(1)}_{nt1})$	0.0995	0.1503	$\frac{\partial p_0(\Psi^{(1)}_{nt4})}{\partial s_{nt}}$	0.8412	5.5360
$p_0(\Psi^{(2)}_{nt1})$	0.3659	0.2446	$\frac{\partial p_0(\Psi^{(1)}_{nt5})}{\partial s_{nt}}$	-0.5360	6.3767
$p_0(\Psi^{(1)}_{nt2})$	0.0283	0.1095			
$p_0(\Psi^{(1)}_{nt3})$	0.2616	0.3736			
$p_0(\Psi^{(1)}_{nt4})$	0.0370	0.1504			
$p_0(\Psi^{(1)}_{nt5})$	0.1436	0.3166			

Table 11: Psychic Value of School Attendance.

Variable	Parameter	Estimate	Std.Err.
Constant	B_{50}	-20.8502	10.0810
S_{nt}	B_{51}	3.6935	1.6900
S_{nt}^2	B_{52}	-0.0654	0.0619
$AGE_{nt} \times S_{nt}$	B_{53}	-0.0635	0.0093
$BLACK$	B_{54}	1.4361	1.3736
$HISPANIC$	B_{55}	0.0667	1.8812
$AFQT$	B_{56}	0.0165	0.0343

Table 12: Fixed Utility of Labor Force Participation.

Variable	Parameter	Estimate	Std.Err.
Constant	B_{60}	-0.8174	2.3807
E_{nt}	B_{61}	1.2834	1.2741
E_{nt}^2	B_{62}	-0.0270	0.2294
$AGE_{nt} \times E_{nt}$	B_{63}	-0.0645	0.0176
$BLACK$	B_{64}	-0.4961	1.4026
$HISPANIC$	B_{65}	-0.0351	2.4603

Table 13: Utility of Leisure and the CRRA parameter.

Variable	Parameter	Estimate	Std.Err.
l_{nt}	B_{70}	0.0043	0.0114
$AGE_{nt} \times l_{nt}$	B_{71}	-0.0009	0.0010
$AGE_{nt}^2 \times l_{nt}$	B_{72}	0.27E-04	0.24E-04
$BLACK \times l_{nt}$	B_{73}	0.0009	0.0008
$HISPANIC \times l_{nt}$	B_{74}	0.0003	0.0021
l_{nt}^2	δ_0	-0.58E-07	0.68E-07
$l_{nt} l_{nt-1}$	δ_1	2.87E-07	1.15E-07
$l_{nt} l_{nt-2}$	δ_2	3.86E-07	0.11E-07
CRRA parameter	α	0.1067	0.0060

Table 14: Results from baseline simulation by race.

Age	Education		Experience		Hours		Wages	
	Actual	Sim.	Actual	Sim.	Actual	Sim.	Actual	Sim.
				White				
20	11.96	10.37	3.32	2.90	1257	1708	4.89	3.77
25	13.16	12.21	6.96	5.19	1957	1812	9.37	6.71
30	13.52	13.43	10.70	7.50	2198	2092	13.77	11.57
35		14.37		9.93		2338		15.85
				Black				
20	11.71	9.69	2.67	2.65	1129	1521	4.35	3.35
25	12.36	10.90	5.90	4.61	1830	1711	7.38	5.73
30	12.53	11.58	9.62	6.52	1963	2023	10.36	8.84
35		11.91		8.60		2275		11.67
				Hispanic				
20	11.33	9.69	3.04	2.84	1320	1773	5.00	3.82
25	11.99	10.89	6.71	5.03	1817	1960	9.15	6.57
30	12.28	11.56	10.57	7.20	2107	2219	12.26	10.03
35		11.90		9.61		2403		13.20

Table 15: Effect of cash subsidy to students who do not work.

Age	Education		Experience		Hours		Wages	
	Base	Pol 1	Base	Pol 1	Base	Pol 1	Base	Pol 1
	White							
20	10.37	10.38	2.90	2.88	1708	1709	3.77	3.77
25	12.21	12.24	5.19	5.14	1812	1810	6.71	6.71
30	13.43	13.48	7.50	7.41	2092	2094	11.57	11.68
35	14.37	14.44	9.93	9.78	2338	2336	15.85	16.08
	Black							
20	9.69	9.71	2.65	2.63	1521	1522	3.35	3.35
25	10.90	10.95	4.61	4.55	1711	1708	5.73	5.72
30	11.58	11.63	6.52	6.41	2023	2020	8.84	8.85
35	11.91	11.98	8.60	8.41	2275	2274	11.67	11.78
	Hispanic							
20	9.69	9.71	2.84	2.83	1773	1771	3.82	3.81
25	10.89	10.92	5.03	4.99	1960	1958	6.57	6.56
30	11.56	11.61	7.20	7.10	2219	2222	10.03	10.02
35	11.90	11.96	9.61	9.44	2403	2402	13.20	13.21

Table 16: Effects of mandatory increases in time spent on school activities.

Age	Education		Experience		Hours		Wages	
	Base	Pol 2	Base	Pol 2	Base	Pol 2	Base	Pol 2
	White							
20	10.37	10.66	2.90	2.89	1708	1722	3.77	3.78
25	12.21	12.95	5.19	5.09	1812	1827	6.71	7.05
30	13.43	14.78	7.50	7.23	2092	2185	11.57	14.99
35	14.37	16.52	9.93	9.38	2338	2412	15.85	23.28
	Black							
20	9.69	9.96	2.65	2.64	1521	1528	3.35	3.35
25	10.90	11.62	4.61	4.52	1711	1704	5.73	5.84
30	11.58	12.68	6.52	6.35	2023	2070	8.84	10.50
35	11.91	13.38	8.60	8.39	2275	2325	11.67	15.15
	Hispanic							
20	9.69	9.76	2.84	2.84	1773	1771	3.82	3.81
25	10.89	10.98	5.03	5.04	1960	1958	6.57	6.58
30	11.56	11.66	7.20	7.23	2219	2222	10.03	10.07
35	11.90	12.00	9.61	9.67	2403	2402	13.20	13.24

Table 17: Equating school quality.

Age	Education		Experience		Hours		Wages	
	Base	Pol 3	Base	Pol 3	Base	Pol 3	Base	Pol 3
	Black							
20	9.69	10.03	2.65	2.65	1521	1501	3.35	3.34
25	10.90	11.62	4.61	4.60	1711	1654	5.73	5.82
30	11.58	12.60	6.52	6.65	2023	2000	8.84	9.80
35	11.91	13.20	8.60	8.71	2275	2297	11.67	13.47
	Hispanic							
20	9.69	9.97	2.84	2.86	1773	1771	3.82	3.82
25	10.89	11.39	5.03	5.11	1960	1958	6.57	6.63
30	11.56	12.21	7.20	7.40	2219	2222	10.03	10.35
35	11.90	12.68	9.61	9.94	2403	2402	13.20	13.67

Table 18: Effects of mandatory increases in time spent on school activities after equating school quality.

Age	Education		Experience		Hours		Wages	
	Base	Pol 4	Base	Pol 4	Base	Pol 4	Base	Pol 4
	Black							
20	9.69	10.28	2.65	2.63	1521	1509	3.35	3.34
25	10.90	12.38	4.61	4.49	1711	1648	5.73	6.06
30	11.58	13.96	6.52	6.26	2023	2118	8.84	13.50
35	11.91	15.26	8.60	8.31	2275	2393	11.67	20.91
	Hispanic							
20	9.69	10.35	2.84	2.84	1773	1737	3.82	3.81
25	10.89	12.29	5.03	5.04	1960	1884	6.57	6.82
30	11.56	13.58	7.20	7.25	2219	2215	10.03	12.45
35	11.90	14.56	9.61	9.70	2403	2432	13.20	17.05

Figure 1: Changes in Shadow Price of Consumption $\Delta((1-\alpha)^{-1}\ln\lambda_t)$

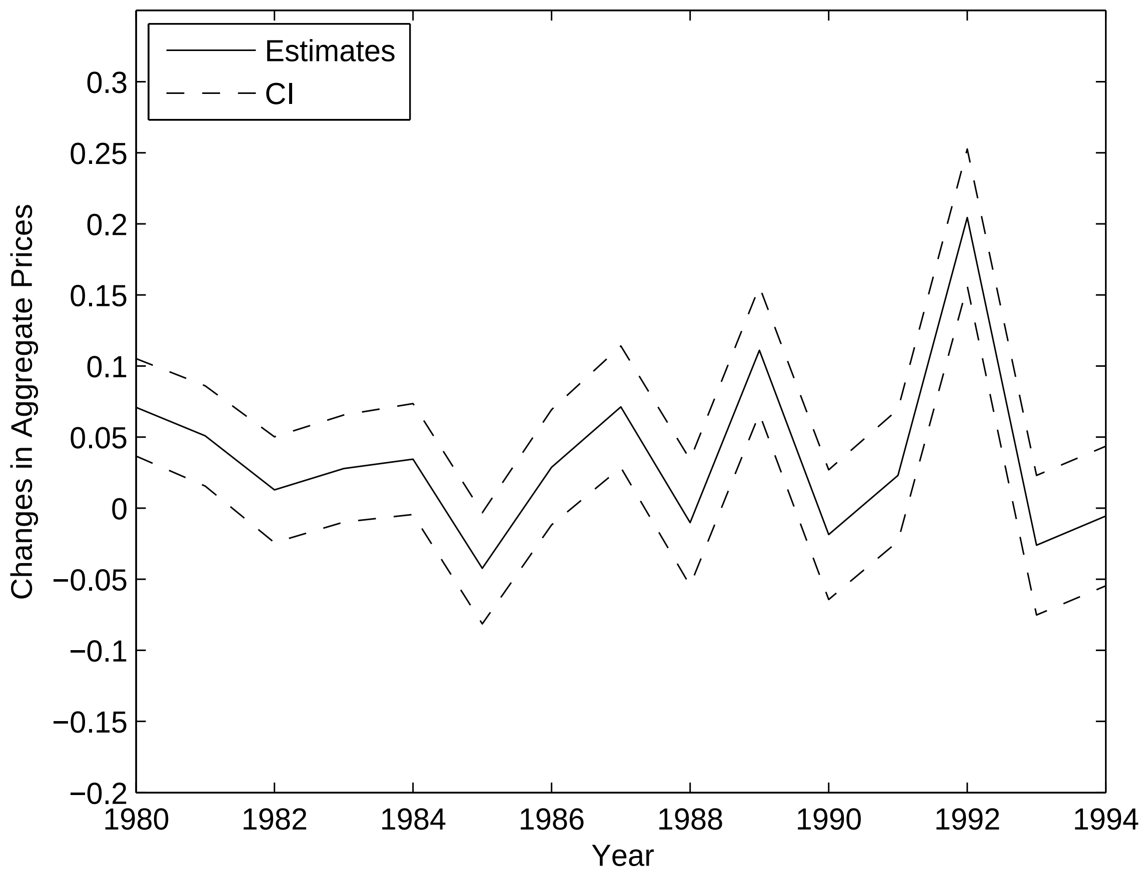


Figure 2: Changes in Unskilled Aggregate Wage $\Delta(\ln \omega_{t1})$

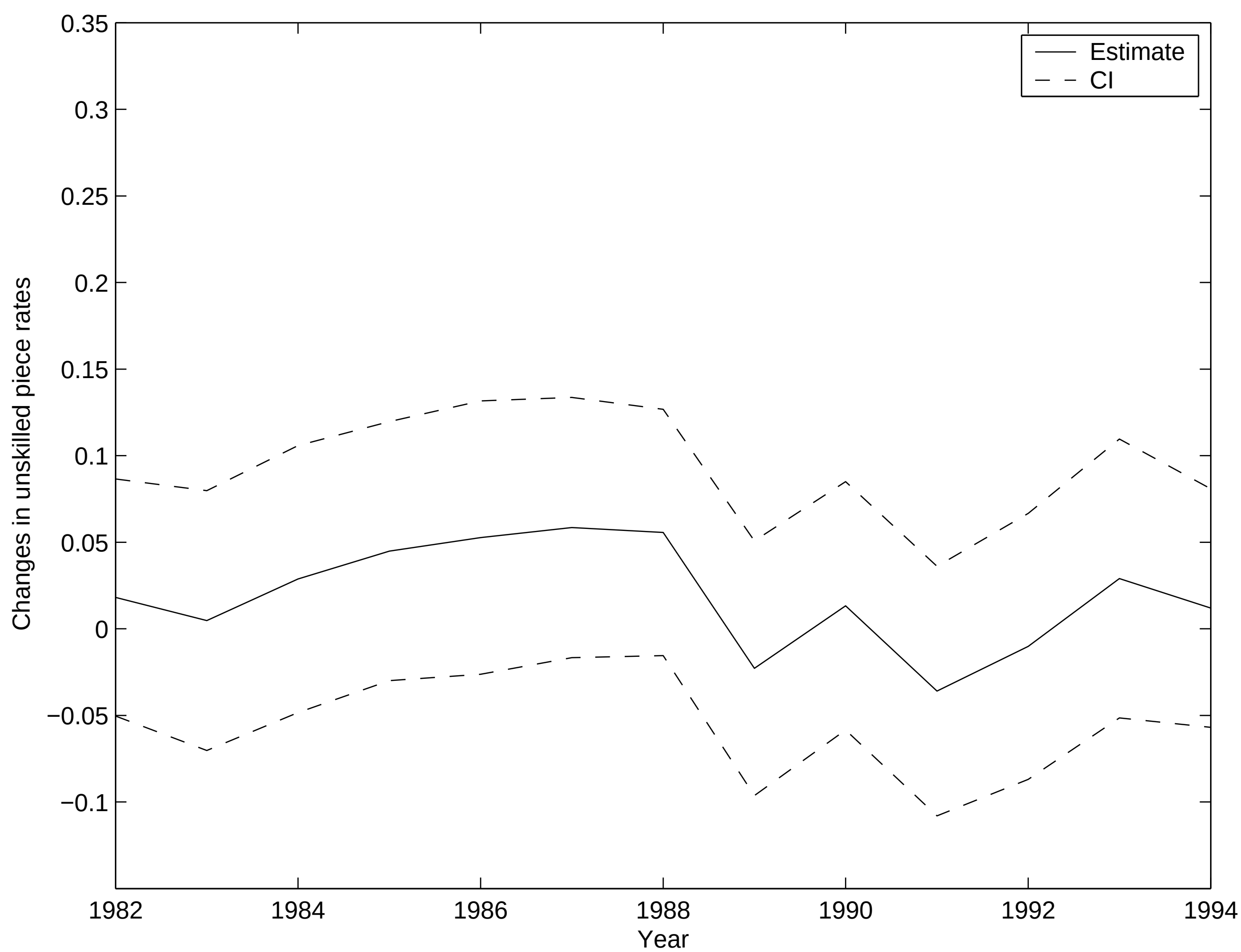
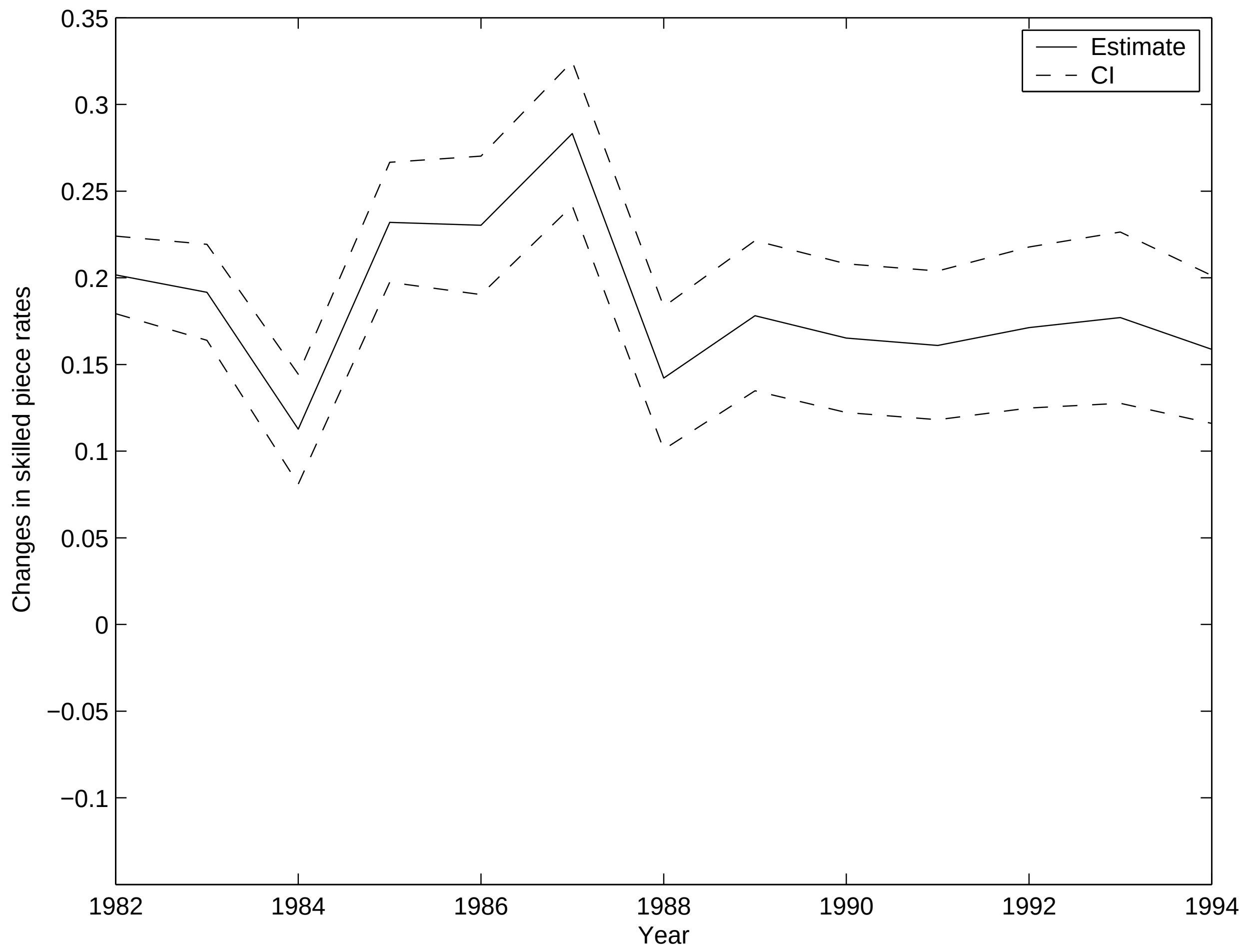


Figure 3: Changes in Skilled Aggregate Wage $\Delta(\ln \omega_{t2})$



3.0 SEMIPARAMETRIC ESTIMATION OF A NONLINEAR PANEL DATA MODEL WITH PREDETERMINED VARIABLES AND SEMIPARAMETRIC INDIVIDUAL EFFECTS (WITH SOILIOU NAMORO)

3.1 INTRODUCTION

Panel data models are important in econometrics, primarily because of their capacity to capture facets of agent behavior in ways that cannot be accounted for in cross-sectional and time-series data models. Furthermore, detailed and relatively reliable panel data sets have become increasingly available. As a consequence, there is a growing demand for more sophisticated panel data models by applied researchers.

A major advantage of the linear panel data model is its ability to jointly account for permanent unobservable individual effects, time specific aggregate effects, and (structural) dynamics in agent behavior.[1] Though at a slower rate, progress has been made by econometricians in developing nonlinear panel data model that allow for individual specific effects, aggregate time effects, and dynamics in behavior. Indeed, the most significant development in nonlinear panel data models has been spurred by the limited dependent variables framework (see Honoré, 1992 and Honoré and Kyriazidou, 2000 for examples). Typically however, the estimation of these nonlinear models rely heavily on the logit specification of the index function. The model we consider in this paper is complementary to these nonlinear index models in that we impose a stronger restriction on the form of the unobserved individual specific effect, but we relax the assumption that the index function is known.

[1] For comprehensive summary of the advances in linear panel data models and estimation techniques, see Chamberlain [1984] and Arellano and Honoré [2001].

This paper considers semiparametric estimation of a nonlinear single index panel data model of the following form:

$$y_{it} = \Phi(\alpha y_{it-1} + x_{it}'\beta + f(z_i) + \lambda_t) + \varepsilon_{it} \tag{3.1.1}$$

where x_{it} is a $K \times 1$ vector of strictly exogenous and/or predetermined variables, z_i is an $L \times 1$ vector of exogenous individual specific time invariant random variables,[2] β is a $K \times 1$ vector of unknown parameters, $\Phi(\cdot)$ is a real valued unknown function, $f(\cdot)$ is an unknown real valued function, λ_t is an unobservable time-specific effect, and ε_{it} is an unobservable error term assumed to have a conditional mean of zero. The purpose of this paper is to find an estimator of β with the usual parametric convergence rate $n^{-1/2}$ without assuming that Φ and f belong to some parametric class of functions. We are particularly interested in constructing estimators for Φ and f because the goal is to be able to simulate and predict the dependent variable y_{it}.

This restriction on the individual-specific effects extends the suggestion of Newey [1994], pp. 1354-1355. In the binary choice framework, the model presented in equation (3.1.1) arises naturally under the assumption that the individual specific effect is of the form $\mu_i = f(z_i) - u_i$. Papers that provide estimators of the finite dimensional parameter in these binary choice models include Chen [1998] and Gayle and Miller [2003]. The former paper suggests implementing a series estimator of the index function and estimating by OLS, where the latter suggests estimating by GMM. Our own interest goes beyond the discrete choice framework, and our estimator is an efficient semi-parametric least squares estimator that can be implemented using either series expansions or the investigators favorite Kernel estimator. Furthermore, the index function is easily recoverable in our estimation framework. This is important since we are specifically interest in estimating the full data-generating process for the purpose of prediction and simulation.

A variety of models used in empirical studies fall within this class of single index models. The model proposed here is in some sense an extension of the single index models proposed in Ichimura [1993] and Klein and Spady [1993] to panel data and pre-determined variables. The cost of this extension relative to these models is that we assume that the index function is strictly increasing over its support. In many cases, the assumption of strict monotonicity of the index function may

[2] Note that z_i could be made of the vector of strictly exogenous random variable $(x_{i1}, \cdots, x_{iT})'$, in which case this is a generalization of the Mundlak specification.

be informed by economic theory. One such example is in cases where the index function is a cumulative distribution function (cdf) as in the case of probability models. Indeed, discrete choice models also falls within the class of models specified in equation (3.1.1).

In the sub-class of discrete choice models, the literature has developed by taking two dominant paths: the case where the index function Φ is assumed to belong to a parametric class of cdf's, and the case where the nonparametric assumption in placed on the cdf. In the former case, most progress have been made under the assumption that the index function Φ is the cdf of the logistic distribution. Rasch [1960], Anderson [1970], and Chamberlain [1980] show that these models can be estimated by conditional maximum likelihood for $T \geq 2$. Chamberlain [1985] and Magnac [1997] show that this model can be estimated with both individual-specific effects and lagged dependent variables, but without any other explanatory variables. Honoré and Kyriazidou [2000] expand the estimation of these models to include explanatory variables.

Despite these rapid advancements, the method of identification used in these studies relies crucially on the logit assumption. Indeed, under the assumption that the individual-time specific shocks are independent and if covariates are unbounded, the finite dimensional parameters can be estimated consistently with a $\sqrt{n}$ convergence rate without specifying the distribution of the individual-specific effects conditional on the covariates if and only if the distribution of the individual-time specific shocks is logistic (Magnac, 2004). However, the logit assumption raises the question of robustness of these estimators to violation of that crucial assumption. This leads us then to find other methods of estimating these dynamic panel data models that allow for sufficiently general individual heterogeneity.

Another class of estimators for discrete choice models are those that do not make parametric assumptions on index function. Manski [1987] derives a maximum score estimator for the single-index model with exogenous regressors, and individual-specific effects based on that, under weak regularity conditions, the sign of difference in the first conditional probabilities is equal to the sign of the first difference in the index. Horowitz [1992] extends this model by maximizing a smoothed version of Manski's score function. This modification results in Horowitz being able to prove asymptotic normality coefficient β, a property that Manski's model does not enjoy. Honoré and Kyriazidou [2000] further extends this estimation technique to include lagged dependent variables. They show that this estimator is consistent, but did not derive the asymptotic distribution for this estimator. These estimators typically converge at a rate slower than $n^{-1/2}$. More importantly, since the index function is not estimated in these models, the full data generating process is not estimated and thus these models are incapable of performing predictions.

The proposed method to estimate the nonlinear panel data model presented in this paper essentially mimics that of the linear case. It starts with the inversion of the unknown function that links the conditional expectations (or the predicted outcome) to the explanatory variables. In fact this alternative representation (3.1.1) can be viewed as a generalized linear model (GLM) with the link function given by the index function Φ^{-1} (see for example Chen, 1995). In this literature, the link function is typically assumed to be known. In this respect the proposed model can be seen as a extension of the GLM.

The inversion is then followed by a differentiation, which suppresses the fixed effects from the regressors. The predicted outcomes are themselves estimated nonparametrically, prior to the computation of the estimator. The method proceeds with an iterative back fitting algorithm, which yields the estimates of the slopes as well the unknown index function. Estimates of the fixed effects are readily obtained from the first estimates.

The rest of paper is organized as follows: the following section describes the class of models considered in this paper. Section (3) discusses identification while section (4) presents the estimator. Section (5) presents the algorithm used to compute the estimate, and section (6) discusses the large sample properties of the estimator. Section (7) is devoted to the monte carlo simulations and section (8) concludes. All the proofs, as well as the lemmas on which these proofs are based, are to be found in the appendix of the paper.

3.2 THE MODEL

The underlying data is a vector valued cross-section

$$(y_i, x_i, z_i) \in M(T \times 1) \times M(T \times K) \times M(L \times 1) =: \mathcal{X},$$

where $M(a \times b)$ denotes the set of real-entry matrices of a rows and b columns. More precisely, we have $y_i := (y_{i1}, \ldots, y_{iT})'$, $x_i := (x_{i1}, \ldots, x_{iT})'$, where $x'_{it} := (x_{it,1}, \ldots, x_{it,K})$ and z_i is an L dimensional vector of time invariant regressors. Since we consider panel data with predetermined variables, we have $T \geq 3$. The basic assumption regarding the data is that the sequence (x_i, y_i, z_i) is an independent and identically distributed $\mathcal{X}$-valued random process, where $\mathcal{X}$ is endowed with its Borel sigma-field $\mathcal{B}$. We shall denote the probability law of the vector (y_i, x_i, z_i) by Q. We assume that at least one column of x_i contains a random variable that is strictly exogenous. Notice that for notational convenience, we have suppressed the explicit representation of the lagged dependent variable and the aggregate shock. Since we allow for x_{it} to include predetermined variables and discrete variables, we can assume that these lagged dependent variable and the aggregate shocks are indeed included in the vector x. We define the conditioning vector w_{it} as

$$w'_{it} := (x'_{it}, z'_i).$$

The model considered in this paper is given by:

$$y_{it} = \Phi(x'_{it}\beta_0 + f_0(z_i)) + \varepsilon_{it}. \tag{3.2.1}$$

The following assumption will be maintained through the paper:

Assumption 3.2.1. $\Phi : \Re \longrightarrow \Re$ is a strictly increasing function.

This assumption arises naturally in discrete choice models where Φ is a cdf. For purposes of estimating the finite dimensional parameter, this assumption can be weakened to the assumption that Φ is strictly increasing on an interval of its index, and that the number of observations within that interval of the support increases with the sample size.

By taking the conditional expectation of y_{it} we obtain:[3]

$$P_{it0} := E(y_{it} \mid w_i) = \Phi(x'_{it}\beta_0 + f_0(z_i)), \ i = 1, \cdots, N, \ t = 1, \cdots T. \tag{3.2.2}$$

Assumption 3.3 allows us to express the relation (3.2.2) as

$$\Phi^{-1}(P_{it0}) = x'_{it}\beta_0 + f_0(z_i), \ i = 1, \cdots, N, \ t = 1, \cdots T, \tag{3.2.3}$$

[3]We shall assume that the unconditional and conditional expectations that we write are all defined and we shall most of the time omit to add the label “almost surely” to relations involving conditional expectations, unless we want to stress the underlying probability.

which in turn implies

$$\Delta[\Phi^{-1}(P_{it0})] := \Phi^{-1}(P_{it0}) - \Phi^{-1}(P_{it-1,0}) = \Delta x_{it}'\beta_0 := (x_{it}'\beta_0 - x_{it-1}'\beta_0) \qquad (3.2.4)$$

Relation (3.2.4) will be the starting point in the definition of our semiparametric (SP) estimators. In particular, defining $\varphi_0 := \Phi^{-1}$ and $\varphi(P_i) := [\varphi(P_{i1}), \cdots, \varphi(P_{iT})]$, relation (3.2.4) can be written as

$$\Delta x_i'\beta_0 - \Delta[\varphi_0(P_{i0})] = 0. \qquad (3.2.5)$$

Our estimation technique is to find the couple (β, φ) that minimizes the mean squared deviation between $\Delta x_{it}'\beta$ and $\Delta[\varphi(P_{it0})]$. This, of course, relies on β_0 and φ_0 being the unique solution to equation (3.2.5). Therefore, we first impose identification restrictions and state the identification theorem.

3.3 IDENTIFICATION

We make the following assumptions:

Assumption 3.3.1. 1. $\|\beta_0\| = 1$.

2. The random vector x_i contains at least one continuous regressor that is not contained in z_i.
3. $E[\Delta x_{it}\Delta x_{it}']$ is invertible.
4. The unconditional mean of the nonparametric individual effect is zero: $E[f_0(z_i)] = 0$

Assumption (3.3.1.1) is frequent in single index models (see Manski, 1987 for example). An alternative normalization (see Horowitz, 1992 and Ichimura, 1993) is to assume that the fist component of x_{it}' has a probability distribution conditional on the remaining components that is absolutely continuous with respect to the Lebesgue measure, and then assume that $|\beta_1| = 1$. In our case, under the assumption that the index function is strictly increasing, assumption (3.3.1.1) allows us to determine the signs of all the coefficients as in linear regression models. Assumption (3.3.1.3) is the traditional full rank condition on the regressors. Of course this condition can be relaxed by considering pseudo-inverses. Assumption (3.3.1.4) is a limit version of the traditional zero average assumption in fixed effects models. This assumption is what gives location identification of the nonparametric functions in our model. This assumption is appealing in applications, but is not necessary for the theoretical model because in most nonlinear and limited dependent variables theoretical models, identification up to location is usually sufficient.

The model (3.2.3) introduced in section 3.2 is specified by the triplet $\pi_0 = (\beta_0, \varphi_0, f_0(z_i))$. Consider another model $\pi_1 = (\beta_1, \varphi_1, f_1(z_i))$. We say that the models π_0 and π_1 are observationally equivalent if π_1 also satisfies:

$$\varphi_1(P_{it0}) = x_{it}'\beta_1 + f_1(z_i),\ i = 1, \cdots, N,\ t = 1, \cdots T. \tag{3.3.1}$$

Then under assumptions (3.3.1.1)-(3.3.1.4) we can prove the following theorem.

Theorem 3.3.2. (Identification) If

i. π_0 and π_1 are observationally equivalent,

ii. π_1 satisfies assumptions (3.3.1.1) - (3.3.1.3), and

iii. φ_1 is strictly increasing,

then

$$\beta_0 = \beta_1 \tag{3.3.2}$$

$$\varphi_0 = \varphi_1 + c \tag{3.3.3}$$

$$f_0(z_i) = f_1(z_i) + c \tag{3.3.4}$$

for some constant c. Furthermore, if assumption (3.3.1.4) also holds, then $c = 0$.

3.4 THE ESTIMATOR

In this section, we define the estimator and describe the algorithm. For ease of exposition we first define the unfeasible estimator and discuss the properties of such an estimator. Then we discuss the feasible estimator. The following estimator is unfeasible because of the fact that the predicted outcome, P_0 is not observed.

Definition 3.4.1. The unfeasible estimator (β^*, φ^*) of (β_0, Φ^{-1}) is the solution to the minimization problem

$$\inf_{(\beta,\varphi)\in\{\beta \mid \|\beta\|=1\}\times\mathcal{S}} \frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\Delta x_{it}'\beta - \Delta[\varphi(P_{it0})]\right)^2, \tag{3.4.1}$$

$$s.t. \quad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\varphi(P_{it0}) - x_{it}'\beta) = 0, \tag{3.4.2}$$

where $\mathcal{S}$ is the set of strictly increasing real-valued functions.

The constraint in (3.4.2) fixes the location of the estimate of φ_0. It imposes that the unfeasible estimator of the fixed effect is of mean zero. This is similar to the restriction imposed in linear models (see Baltagi, 2001).

Typically, the predicted outcomes $P_{it0} = E[y_{it}|w_{it}]$ is unknown and must be estimated. Non-parametric procedures can be used to estimate these quantities. Since this is a conditional expectation, the density of the data, that is found in the denominator, must be bounded away from zero. We therefore impose a fixed trimming condition by defining a closed and bounded subset $\mathcal{W}$ of the support of the density and assume that the density only affects the estimator through its values on this set. Define the function, $J_\delta(w_{nt}) := \delta^{-D_w}J(\delta^{-1}w_{nt})$, where D_w is the dimension of w, and J is a Kernel which integrates to 1 over $\mathcal{W}$. The scalar $\delta \in R^+$ is the band-width associated with the kernel estimator. Then the kernel estimator for the predicted outcomes is given by:

$$\hat{P}_{nt} = \frac{\sum_{m=1,m\neq n}^{N}\sum_{r=1,r\neq t}^{T} y_{mr} J_\delta\left(w_{mr} - w_{nt}\right)}{\sum_{m=1,m\neq n}^{N}\sum_{r=1,r\neq t}^{T} J_\delta\left(w_{mr} - w_{nt}\right)} \tag{3.4.3}$$

Substituting $\hat{P}_{it}$ for P_{it0} in equations (3.4.1), the feasible estimator is obtained as follows:

Definition 3.4.2. The feasible estimator $(\hat{\beta}, \hat{\varphi})$ of (β_0, Φ^{-1}) is the solution to the minimization problem

$$\inf_{(\beta,\varphi)\in\{\beta \mid \|\beta\|=1\}\times\mathcal{S}} \frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\Delta x_{it}'\beta - \Delta[\varphi(\hat{P}_{it})]\right)^2, \tag{3.4.4}$$

$$s.t. \quad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\varphi(\hat{P}_{it}) - x_{it}'\beta) = 0, \tag{3.4.5}$$

Once the problem (3.4.4) is solved, and denoting the solution by $(\hat{\beta}, \hat{\varphi})$, an estimator $\widehat{f(z_i)}$ for the individual-specific effects $f(z_i)$ is given by the following definition:

Definition 3.4.3. The estimator $\widehat{f(z_i)}$ for the individual-specific effects $f(z_i)$ is given by

$$\widehat{f(z_i)} = \frac{\sum_{t=2}^{T}[\hat{\varphi}(\hat{P}_{it}) - x'_{it}\hat{\beta}]}{T} \tag{3.4.6}$$

However, these estimates are not useful if the goal is to perform simulation exercises or to simply make out of sample predictions. A straightforward solution to this problem is a simple kernel estimator of the projection of $\widehat{f(z_i)}$ on z_i as in the estimation of $\hat{P}_{it}$. This then gives a smooth estimator $\hat{f}(z_i)$ of the function $f(z_i)$.

3.5 COMPUTATION OF THE SP ESTIMATORS

3.5.1 The Algorithm

An analytic solution of the problem (3.4.4) hardly exists, due to the presence of a functional component φ. The computation of the SP estimator requires, therefore, the use of a numerical algorithm. Several such algorithms are conceivable. The one that we present here is a back fitting algorithm (Buja, Hastie, and Tibshirani, 1989). It starts with an arbitrarily chosen function φ, and computes and estimate of β, say $\hat{\beta}$. The algorithm proceeds by setting the value of β to $\hat{\beta}$ and then updating the previous estimate of φ, and so on, in a cyclical way until convergence. The SP estimator is chosen to be any couple $(\hat{\beta}, \hat{\varphi})$ that corresponds to the asymptotic fixed point.

The algorithm involves two additional complications above what is discussed in Buja et al. [1989]. The first is that the estimate of β is in fact a constrained estimate. The second is that the estimate of φ involves an (inner) contraction mapping. Define

$$\begin{aligned}
\varphi(P) &:= (\varphi(P_{12}), \cdots, \varphi(P_{NT}))' \\
\varphi(P)_t &:= (\varphi(P_{13}), \cdots, \varphi(P_{NT}))' \\
\varphi(P)_{t-1} &:= (\varphi(P_{12}), \cdots, \varphi(P_{NT-1}))'. \\
\Delta[\varphi(P)] &:= \varphi(P)_t - \varphi(P)_{t-1}
\end{aligned}$$

The following back fitting algorithm can therefore be used:

1. *Initialization*. The parameters to be initialized are below.

   a. φ_1: An initial value of φ can be arbitrarily chosen. For example, one may choose the mapping $\varphi_1(x) = x^3$.

   b. $\hat{P}_{it}$: As stated above, the algorithm requires that estimates of P_{it} be obtained before hand. These empirical quantities can readily be obtained from equation (3.5.4).

   c. $(\varepsilon_1, \varepsilon_2)$: Two small positive numbers to be used in the evaluation of our convergence criteria.

2. *Numerical Evaluation*. Given φ_s at iteration s of the algorithm, approximate values for the pairs $(\beta_{s+1}, \varphi_{s+1})$ are computed recursively as follows:

   a. Compute the constrained regression of $\Delta[\varphi_s(\hat{P})]$ on Δx to obtain the approximate value β_{s+1}.

   To perform this estimation, we present a general technique which is probably standard, but we describe it in detail here since we are unable to find a reference for it. So to the best of our knowledge, this constrained estimation technique is novel. Consider the standard problem of estimating the $(K \times 1)$ dimensional parameter β in the model:

$$y_i = x_i'\beta + \varepsilon_i, \ E[\varepsilon_i | x_i] = 0 \tag{3.5.1}$$

   under the constraint that $||\beta|| = 1$. The solution to this problem can be written as follows:

$$\hat{\beta} = arg \min_{\{\beta:||\beta||=1\}} (-y + x\beta)'(-y + x\beta). \tag{3.5.2}$$

   To solve this problem, we propose solving the auxiliary problem for estimates of the $(K+1) \times 1$ dimensional parameter parameters $\delta = (\delta_1, \delta_2')'$:

$$\begin{aligned} \hat{\delta} &= arg \min_{\{\delta:||\delta||=1\}} (-y\delta_1 + x\delta_2)'(-y\delta_1 + x\delta_2) \\ \Leftrightarrow \widehat{\frac{\delta}{||\delta||}} &= arg \min_{\frac{\delta}{||\delta||}} \frac{\delta}{||\delta||}' B \frac{\delta}{||\delta||} \end{aligned} \tag{3.5.3}$$

   where $\delta_1 > 0$, the $(K+1)$ dimensional square matrix B is given by $B = C'C$ and we define $C := (-y, x)$. Then it is well known that the solution to this problem is the (normalized) eigenvector corresponding to the smallest eigenvalue of B. There are numerous efficient softwares available for computing these eigenvectors. We adopt the subroutine “jacobi” from Press et al. [1996] for our purposes. This algorithm computes the solution in milliseconds. Once we have the solution $\frac{\delta}{||\delta||} = (\tilde{\delta}_1, \tilde{\delta}'_2)'$ our solution for the original problem is easily recovered by the equality $\hat{\beta} = (1 - \tilde{\delta}_1^2)^{-1/2}\tilde{\delta}_2$. This process is very fast on standard computers, even for quite large values of K. Indeed, there is no observable difference in the time this process takes to estimate β and the time that would be taken by OLS. This makes this constrained estimation technique quite appealing.

b. Perform the regression of the vector $\Delta x'_{it}\beta_{s+1} + \varphi_s(\hat{P})_{t-1}$ on $\hat{P}$ to obtain the approximate value φ_{s+1}.

Recall that we assume that the index function φ is strictly increasing. This assumption is not necessary for the algorithm to converge. Indeed it is sufficient that the index function be in the class of functions of bounded variation. Furthermore, the current technology in isotonic regression when the data set is large is unsatisfactory, because of the computational time required to implement any of these techniques. As such, we relax the strict monotonicity assumption on φ in the algorithm. Besides the ease of computation, dropping this constraint allows one to test the assumption of monotonicity of the index function.

The regression of $\Delta x'_{it}\beta_{s+1} + \varphi_s(\hat{P})_{t-1}$ on $\hat{P}$ is itself a fixed point algorithm. The algorithm accommodates either kernel or series estimators, but we present the kernel estimator here. The procedure goes as follows. Given $\varphi_s = \varphi_{sj}$, we Construct $\Delta x'_{it}\beta_{s+1} + \varphi_{sj}(\hat{P}_{it-1})$. Then, adopting the notation of the kernel estimator given in section (3.4), for any value $\hat{P}_{nt}$ the kernel estimator for $\varphi_{sj+1}(\hat{P}_{nt})$ is given by:

$$\varphi_{sj+1}(\hat{P}_{nt}) = \frac{\sum_{m=1,m\neq n}^{N}\sum_{r=3,r\neq t}^{T} J_\delta\left(\hat{P}_{mr} - \hat{P}_{nt}\right)\left(\Delta x'_{mr}\beta_{s+1} + \varphi_{sj}(\hat{P}_{mr-1})\right)}{\sum_{m=1,m\neq n}^{N}\sum_{r=3,r\neq t}^{T} J_\delta\left(\hat{P}_{mr} - \hat{P}_{nt}\right)}. \qquad (3.5.4)$$

This process is repeated (in j) until convergence. The proof that equation (3.5.4) defines a contraction mapping is presented in Appendix A.2.6. The convergence criterion for this inner contraction mapping is provided by the following inequality:

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\varphi_s(\hat{P}_{it}) - \varphi_{s-1}(\hat{P}_{it})\right)^2 \leq \varepsilon_1^2. \qquad (3.5.5)$$

If (3.5.5) holds at iteration J then we take φ_{sJ} to be our projection estimate φ_{s+1} for iteration $s+1$ of the outer contraction. Our experience is that this part of the algorithm converges within 6 iterations and is extremely stable.

3. *Outer Convergence*. Step 2. is repeated until the process converges. The convergence criterion is provided by the following inequality:

$$\|\beta_s - \beta_{s-1}\|_K^2 \leq \varepsilon_2^2 \tag{3.5.6}$$

If (3.5.5) and (3.5.6) hold, then the corresponding values (β_s, φ_s) constitute our numerical solution $(\hat{\beta}, \hat{\varphi})$. Otherwise, step 2. is repeated until the conditions (3.5.5) and (3.5.6) are simultaneously satisfied.

The corresponding estimates of the individual-specific effects is then computed from equation (3.4.6).

An important point to note is that if the investigator is willing to assume the form of the index function Φ, for example, a logit or probit specification, then the computation of the estimator $\hat{\beta}$ is simply to Regress $\Phi^{-1}(\hat{P})$ on Δx. The asymptotic properties of this estimator will typically be the same as that of the more general estimator which assumes that the index function is unknown. This property illustrates the power of the estimator presented in this paper.

3.6 ASYMPTOTIC PROPERTIES OF THE SP ESTIMATOR

In order to derive the asymptotic properties of the SP estimator, some regularity conditions must be imposed. We turn first to the nuisance parameter, the first stage kernel estimator of $P_{it0} = E[y_{it}|w_{it}]$. Following Newey and McFadden [1994] we impose conditions that ensures uniform convergence of the nonparametric estimate $\hat{P}_{it0}$. Following the notation of Newey and McFadden [1994], define $\gamma := (\gamma_1, \gamma_2)$ where $\gamma_1 := f(w_{it})$ and $\gamma_2 := f(w_{it})E[y_{it}|w_{it}]$. Clearly $P_{it} = \gamma_2/\gamma_1$. Define also $q_{it} := (1, y_{it})'$. Then the numerator and denominator of the first stage kernel estimator can be conveniently written as $\hat{\gamma(w)} = \sum_{i=1}^{N}\sum_{t=3}^{T} q_{it} J_\delta(w - w_{it})$. We make the following assumptions

Assumption 3.6.1. 1. $J(u)$ is differentiable of order d, the derivatives d are bounded, $J(u)$ is zero outside a bounded set, $\int J(u)du = 1$, there is a positive integer m such that for all $j < m$, $\int J(u)[\otimes_{\ell=1}^{j}]du = 0$

2. There is a version of $\gamma_o(w)$ that is continuously differentiable to order d with bounded derivatives on an open set containing $\mathcal{S}$, a set contained in the support of w.
3. There is $r \geq 4$ such that $E[||q||^r] < \infty$ and $E[||q||^r|w]f_0(w)$ is bounded.
4. The bandwidth $\delta = \delta(N)$ satisfies
$N^{1-(2/r)}\delta^k/\ln N \longrightarrow \infty$, $\sqrt{N}\delta^{2m} \longrightarrow 0$, and $\sqrt{N}\ln N/(N\delta^{r+2d}) \longrightarrow 0$

Under these assumptions, Newey and McFadden [1994] shows that

$$\sqrt{N}||\hat{\gamma} - \gamma_0|| \xrightarrow{p} 0, \qquad (3.6.1)$$

where the norm here is the Sobolev norm.

We now impose conditions for consistency of our estimates of the the pair (β_0, φ_0). First, the fixed trimming condition along with assumptions 3.6.1 and 3.6.2 imply that there is a compact set $\mathcal{K}$ in which all the P's lie. We therefore define the restriction of the set $\mathcal{S}$ to $\mathcal{K}$ as $\mathcal{S}_{\mathcal{K}}$. Define also the distance d on the cartesian product $B_K[0,1] \times \mathcal{S}_{\mathcal{K}}$ (where B_K is the close unit ball in $\Re_K$) as follows:

$$d[(\beta,\phi),(\alpha,\psi)] := ||\beta - \alpha||_K + \sup_P |\phi(P) - \psi(P)|$$

where $||.||_K$ is the Euclidean norm on $\Re^K$. In what follows we assume that the conditions on the kernel in assumption are also satisfied by the kernel used to estimate φ.

Assumption 3.6.2. 1. $||\Delta X_i||_2^T \leq R_0 > 0 \ \forall i \geq 1$ Q-almost surely.

2. Each element of $\mathcal{S}$ is differentiable.
3. There exist (unknown) γ such that for every $\varphi \in \mathcal{S}_{\mathcal{K}}$, and for any $x \in \mathcal{K}^\circ$, where $\mathcal{K}^\circ$ denotes the interior of $\mathcal{K}$, $\varphi'(x) \leq \gamma < \infty$.
4. There is a (unknown) function η in $\mathcal{S}$ such that for all $x \in \Re$

$$\sup_S |\varphi(x)| = |\eta(x)|$$

Assumption 3.6.2.1 is weaker than assuming that the covariates are uniformly bounded almost surely. We now state the consistency and asymptotic normality theorems.

Theorem 3.6.3. Let the assumptions 3.3.1, 3.6.1, and 3.6.2 be satisfied. Then

$$\hat{\beta} \xrightarrow{p} \beta_0$$
$$\sup_{P \in \mathcal{K}} |\hat{\varphi}(P) - \varphi_0(P)| \xrightarrow{p} 0$$

Theorem 3.6.4. If the assumptions 3.3.1, 3.6.1, and 3.6.2 are satisfied, then

$$\sqrt{N}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V),$$

where $V = E[\Delta x' \Delta x]^{-1} E[\Delta x' R] \Omega E[R' \Delta x] E[\Delta x' \Delta x]^{-1}$, $\Omega = Var(\varepsilon)$, and

$$R_i = \begin{pmatrix} -\varphi_0'(P_{i1}) & \varphi_0'(P_{i2}) & 0 & \cdots & 0 & 0 \\ 0 & -\varphi_0'(P_{i2}) & \varphi_0'(P_{i3}) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\varphi_0'(P_{iT-1}) & \varphi_0'(P_{iT}) \end{pmatrix}.$$

3.6.1 Semiparametric Efficiency Bound

We now tackle the question of whether the proposed estimator of the finite dimensional parameter β is efficient. The model for which we compute the efficiency bound is the implied model given in equation (3.2.5), and not the conditional independence model of equation (3.1.1). In general these bounds are different, and in many cases that of equation (3.1.1) may not be sharp, in that there may be no estimator that can attain the bound. The variance bound that we compute for equation (3.2.5) is the one that would be attained within an GMM framework. Thus our estimation framework is as efficient as any competing extremum estimator for the condition given in (3.2.5), but retains the property that it is independent of the choice of smoother, and that a consistent estimator of the infinite dimensional nuisance parameters are computed immediately, and ready for simulations and predictions. We state the following theorem:

Theorem 3.6.5. The estimator of the finite dimensional parameter β developed in section 3.4 is semiparametric efficient with variance bound given in theorem 3.6.4.

3.7 MONTE CARLO RESULTS

In this section we examine the small sample properties of the proposed estimator via a Monte Carlo experiment. Consider the following data generating process:

$$y_{it} = \Phi(\alpha y_{it-1} + \beta x_{it} + f(z_i)) + v_{it}, \qquad i = 1, \cdots, n, \quad t = 1, 2, 3.$$

In this model the index function is chosen to be asymmetric about 0 with range between 0 and 10. Specifically the index function is given by:

$$\begin{aligned} \Phi(x) &= \frac{10}{1 + e^{-\lambda(x)x}}, \\ \lambda(x) &= 0.5 - \frac{0.35}{1 - e^{-5x}}. \end{aligned} \tag{3.7.1}$$

The individual specific effect is given by:

$$f(z_i) = 4\left(\frac{e^{-z_i}}{1 + e^{-z_i}} - \frac{1}{N}\sum_{i=1}^{N}\frac{e^{-z_i}}{1 + e^{-z_i}}\right) \tag{3.7.2}$$

This specification results in the the function f to be of mean zero and ranges between -2 and 2. The strictly exogenous random variable x_{it} is distributed N(1,7), and z_i is distributed N(0,3). The error term v_{it} is distributed N(0,0.5). The initial values y_{i0} are distributed N(0,6). Finally, (α, β) = (0.6,0.8). We perform 50 Monte carlo replications of the model for four sample sizes n: 200, 500, 1000, and 1500. The mean bias and the root mean squared error (RMSE) are calculated for each sample size.

The computation was done on a 3GHz Pentium 4 laptop computer. The algorithm take 30 seconds of CPU time to compute the estimates for a sample size of 1500. Table reft181 reports the results from the Monte Carlo study. The results indicate that the estimator performs remarkably well, even so for the index function Φ.

Table 19: Small sample properties of estimator. Monte Carlo simulation of 50 trials.

	α		β		Φ
Sample Size	Mean Bias	RMSE[a]	Mean Bias	RMSE[a]	RMSE[a]
200	0.0235	0.0355	-0.0189	0.0283	0.2735
500	0.0133	0.0201	-0.0104	0.0157	0.1721
1000	0.0153	0.0185	-0.0118	0.0143	0.1435
1500	0.0128	0.0159	-0.0098	0.0123	0.1165

[a] Root mean square error.

Figures 4 to 8 presents plots of the estimated and true index function Φ for sample the four sample sizes 200, 500, 1000, and 1500. The estimated index function tracks very well the true one even for the sample size of 200. For the sample size of 1500, the two plots are largely indistinguishable. This again shows that the proposed cyclical projection algorithm performs remarkably well.

3.8 EMPIRICAL EXAMPLE

In this section, we implement the algorithm developed in the paper to estimate a wage equation.[4] The wage equation has the following specification

$$w_{it} = \Phi(x_{it}'\beta + f(z_i)) + \varepsilon_{it}, \tag{3.8.1}$$

where the assumptions on the data are as in section 3.4. The vector x_{it} is composed of the first two lags of hours worked (h_{it-1}, h_{it-2}) and labor force participation (d_{it-1}, d_{it-2}), highest grade level completed (S_{nt}), and squared age (AGE^2). As is well established in the literature on returns to education, this equation is subject to selectivity bias which is typically called ability bias. The idea

[4]The executables panel.exe for WINDOWS and panel.out for UNIX used for estimation of these models is available upon request from the authors.

is that both wage and level of education are partially determined by the ability of the individual, which is unobserved, creating a correlation with the explanatory variable S_{it} and the error term. In this exercise however, we control for ability bias by including AFQT scores as a time invariant explanatory variable in z_i. The other time invariant covariates included in z_i are the indicators BLACK and HISPANIC.

The data is taken from the 1979 youth cohort of the National Longitudinal Survey of Labor Market Experience (NLSY79), a comprehensive panel data set that follows individuals over the period 1979 to 2000, who were 14 to 21 years of age as of January 1, 1979. The data set initially consisted of 12,686 individuals: a representative sample of 6,111 individuals, a supplemental sample of 5,295 Hispanics, non-Hispanic blacks, and economically disadvantaged, non-black, non-Hispanics, and a supplemental sample of 1,280 military youth. This study makes use of 9 years of interviews, from 1982 to 1990. The data is restricted to include males.

The estimates are presented in Table 1. The signs and relative magnitude of lagged hours worked are consistent with the hypothesis of returns to on the job training, and depreciation in human capital. Furthermore, the coefficient on S_{it} is positive and significant at the 5 percent level. The only coefficient that does not conform to a-priori expectation is the coefficient of AGE^2, which is positive and significant. However, considering that the maximum age in the sample is 37, it is unlikely that the declining effect on wages would be captured in this estimation, since this typically begins in early to mid forties.

The isotonic estimate of the index function $\hat{\Phi}$ is presented in figure 5. It is interesting to note that the shape of the index function roughly resembles that of the exponential function. This is notable since it is common practice to express the wage equation in log linear form.

3.9 CONCLUSION

Over recent years, the specifications of econometric models have undergone a rapid increase in complexity. An important stimulus for this transformation is that applied economists have become more sensitive to the issue of specification bias and robustness of the estimation techniques used in practice. The relaxation of parametric assumptions however comes at the cost of less tractable estimators, increased computational time, and limited ability to perform post-estimation analysis such as out of sample predictions and Monte-Carlo studies. Hence, there is a need for the development of estimators that allow for flexible specification, but at the same time estimates the full data generating process at cheap computational cost.

This paper attempts to contribute to the semi-parametric single index panel data framework by presenting an estimator and algorithm to achieve the above goals. In particular, we develop an efficient semiparametric method for estimating nonlinear panel data index models with small-T. The estimation technique allows for the inclusion of predetermined variables, in particular lagged dependent variables, aggregate time-specific unobserved effects, and a semiparametric specification of the individual-specific effects. The paper provides a root-N consistent, asymptotically normal and efficient estimator for the slope parameter, a consistent nonparametric estimator of the index function as well as its convergence rate, and an estimator of the individual specific effects. Thus with our estimator, one can predict and simulate the dependent variable. The algorithm presented is straightforward and is found to be quite stable in practice. It immediately provides an estimate of the index function, and the the investigator may implement it his favorite series or kernel smoother in estimation. To the best of our knowledge, this property is novel in this framework. Furthermore, the algorithm can estimate an extension of the generalized linear model (GLM) where the link function is unspecified and not assumed to be monotone. Therefore our implied model (i.e. assuming that the ananlysis begins with equation (3.2.4)) can be used to test the assumption of monotonicity, linearity, or simply that the index function belongs to a specific parametric family such as the logistic. Excellent references for testing of monotonicity in nonparametric regression include Bowman et al. [1996], van der Vaart et al. [1998], and Gibjels [2003].

Table 20: The Semiparametric Wage Equation

Variable	Estimate
Lags of Hours Worked	
Δh_{it-1}	0.000282
	(0.0000006)
Δh_{it-2}	0.000163
	(0.0000001)
Lags of employment	
Δd_{it-1}	-0.8528
	(0.0017)
Δd_{it-2}	-0.4451
	(0.0010)
Education	
ΔS_{it}	0.2732
	(0.0015)
Demographic Variable	
ΔAGE_{it}^2	0.0035
	(0.00004)

Figure 4: True and estimated index function for sample size of 200

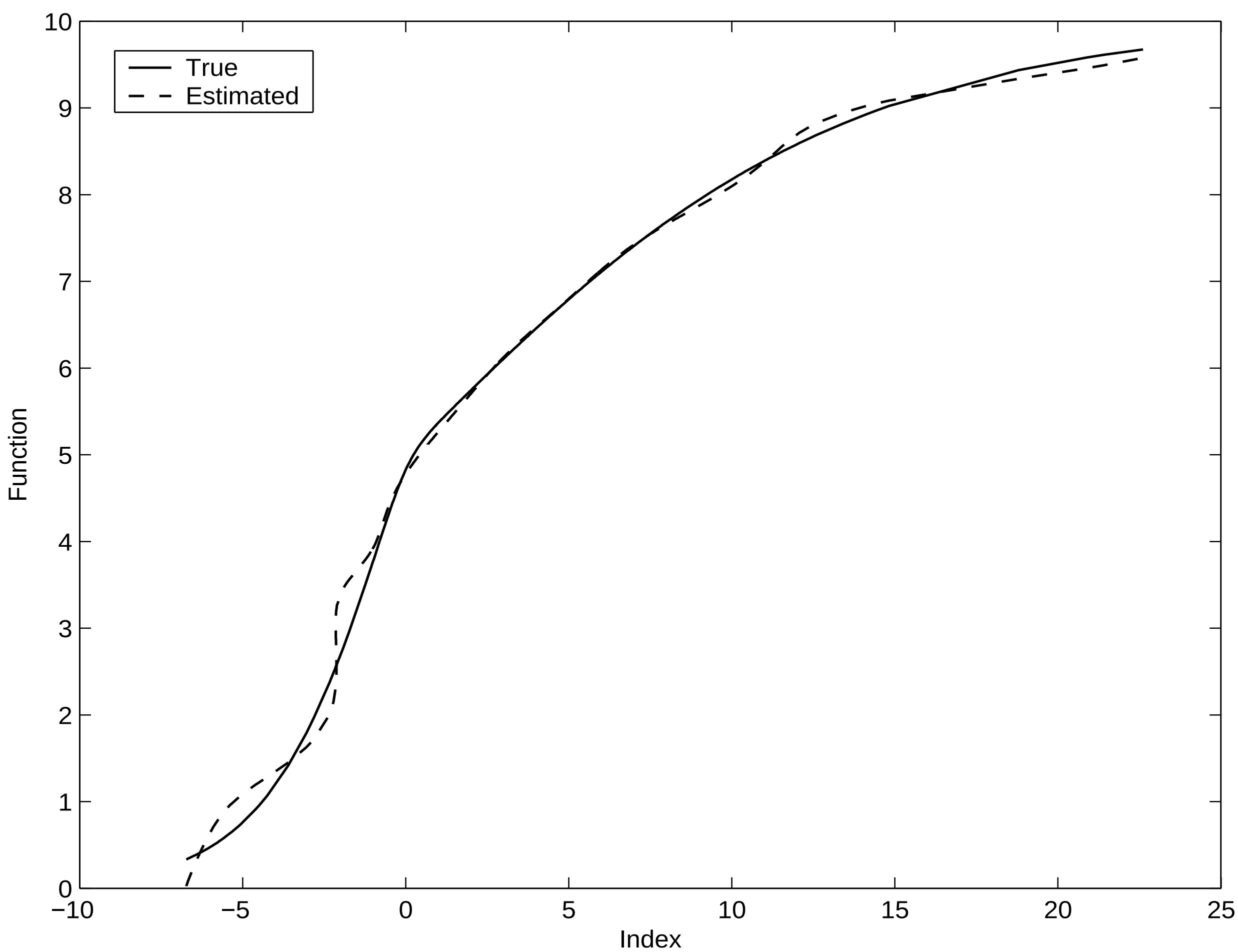


Figure 5: True and estimated index function for sample size of 500

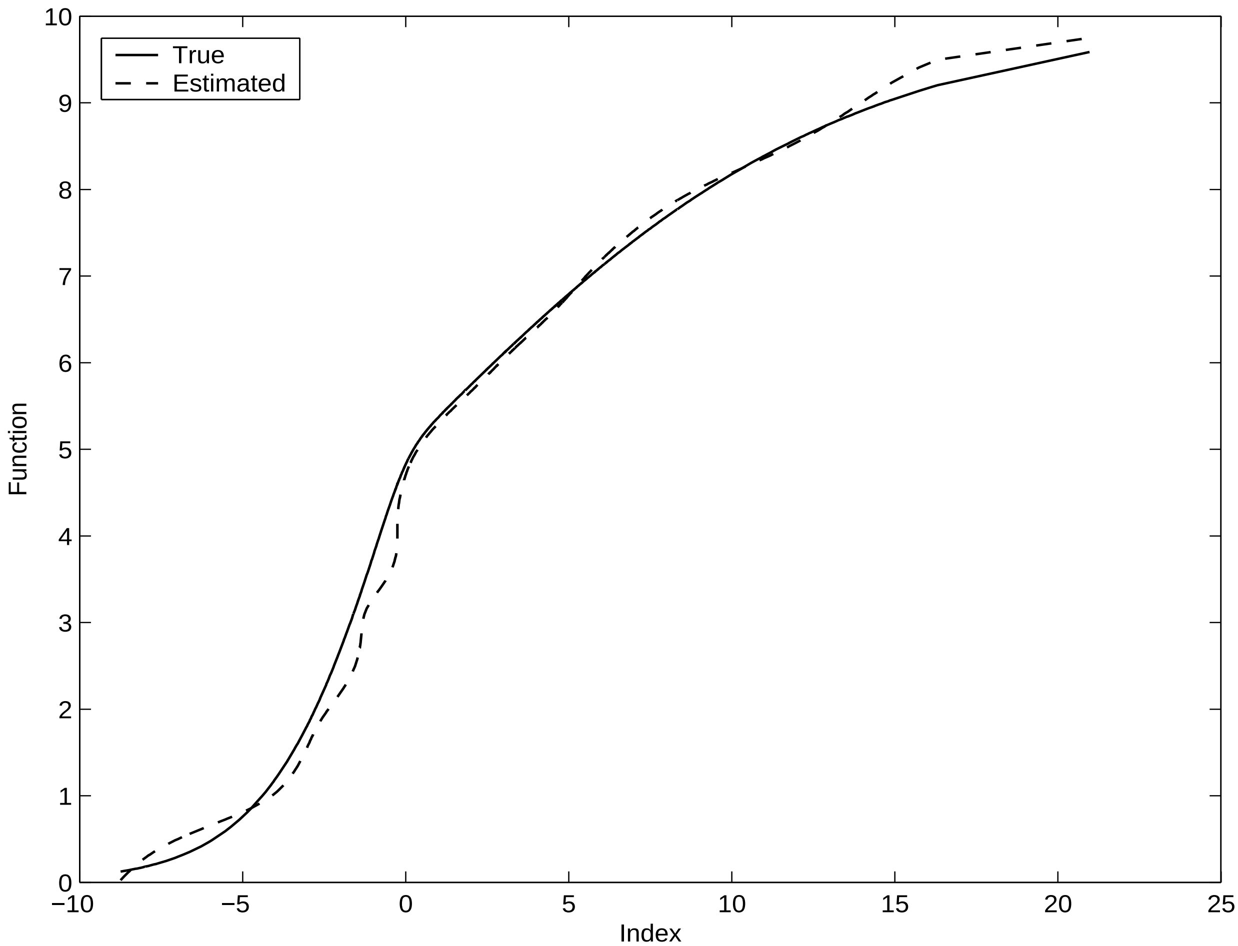


Figure 6: True and estimated index function for sample size of 1000

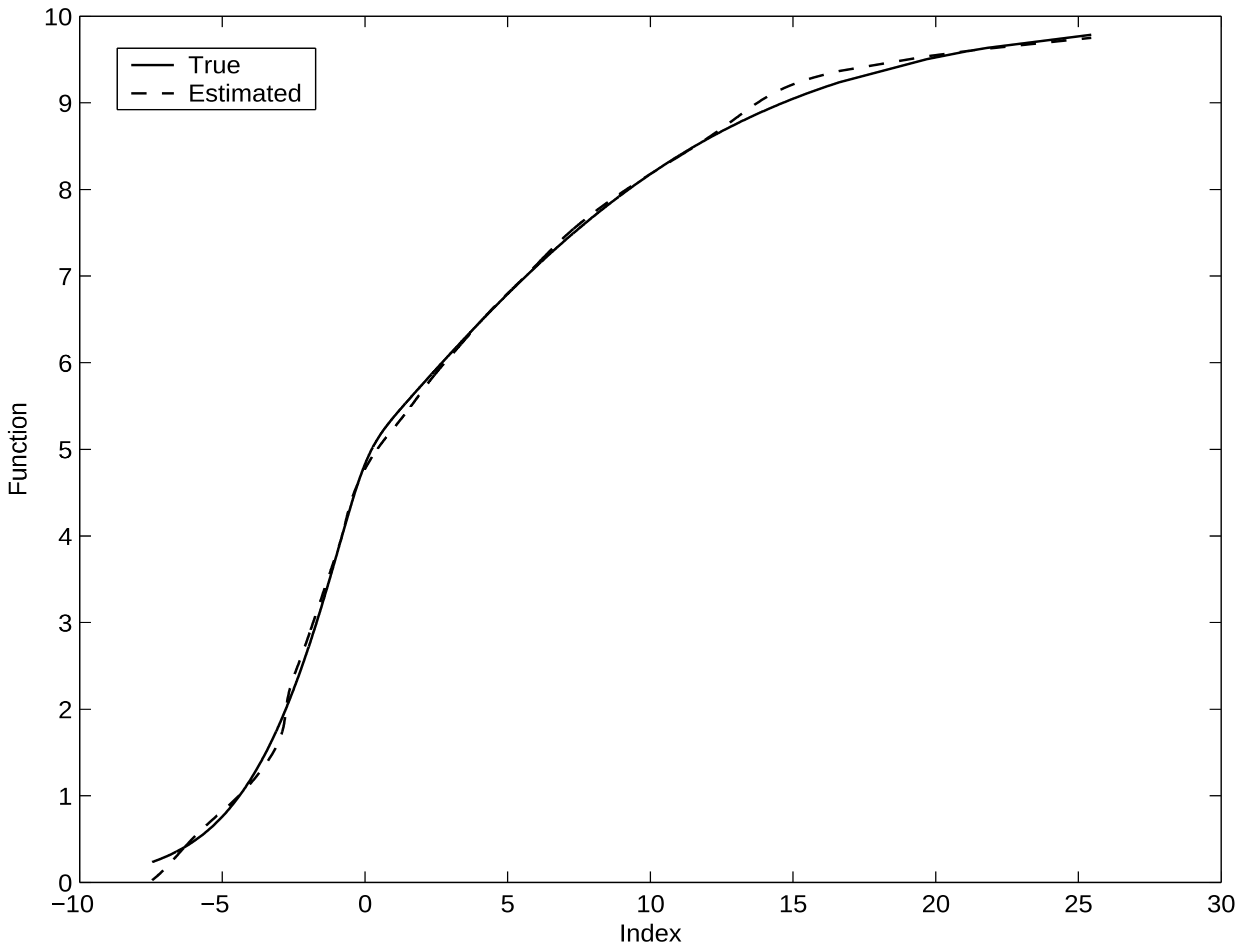


Figure 7: True and estimated index function for sample size of 1500

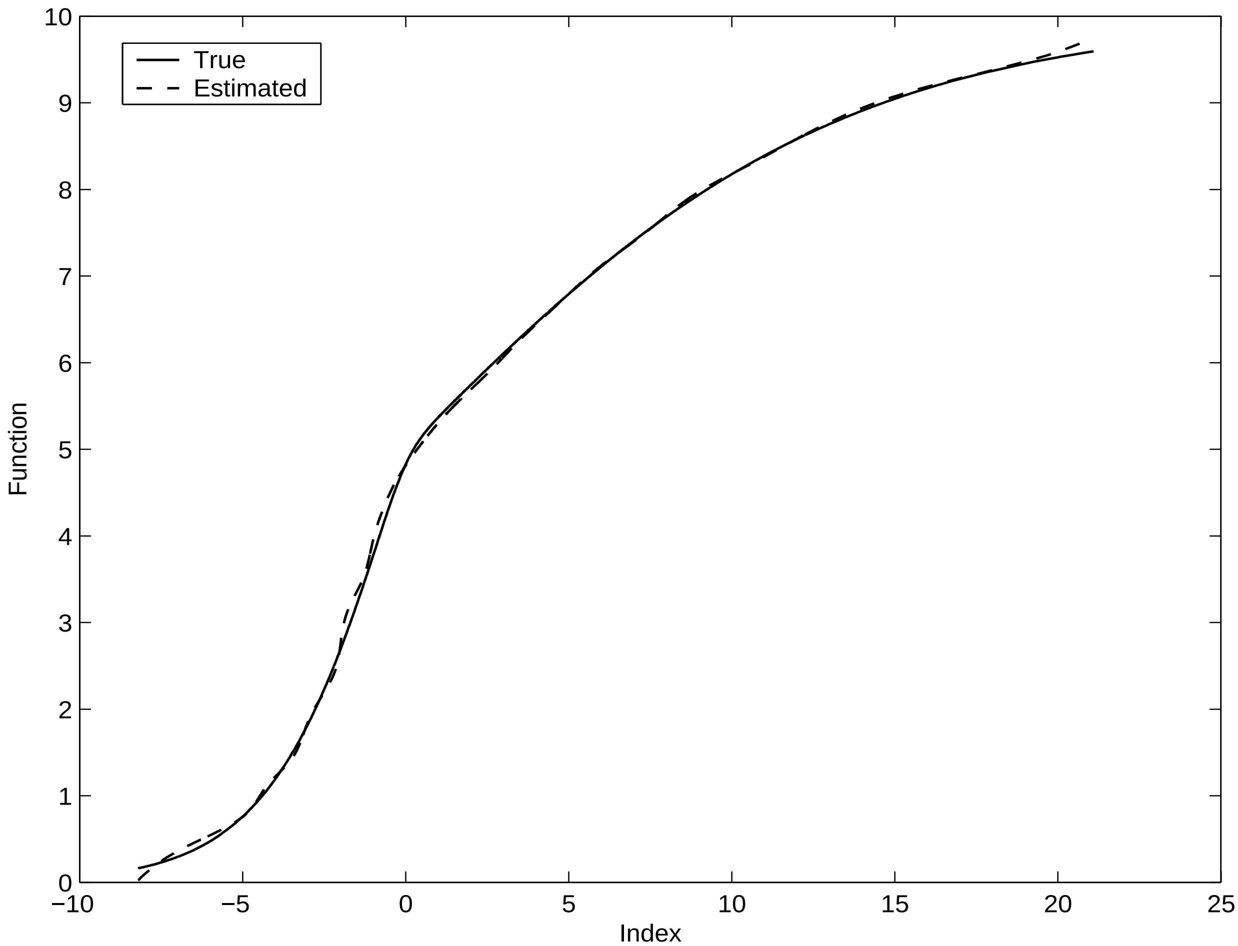
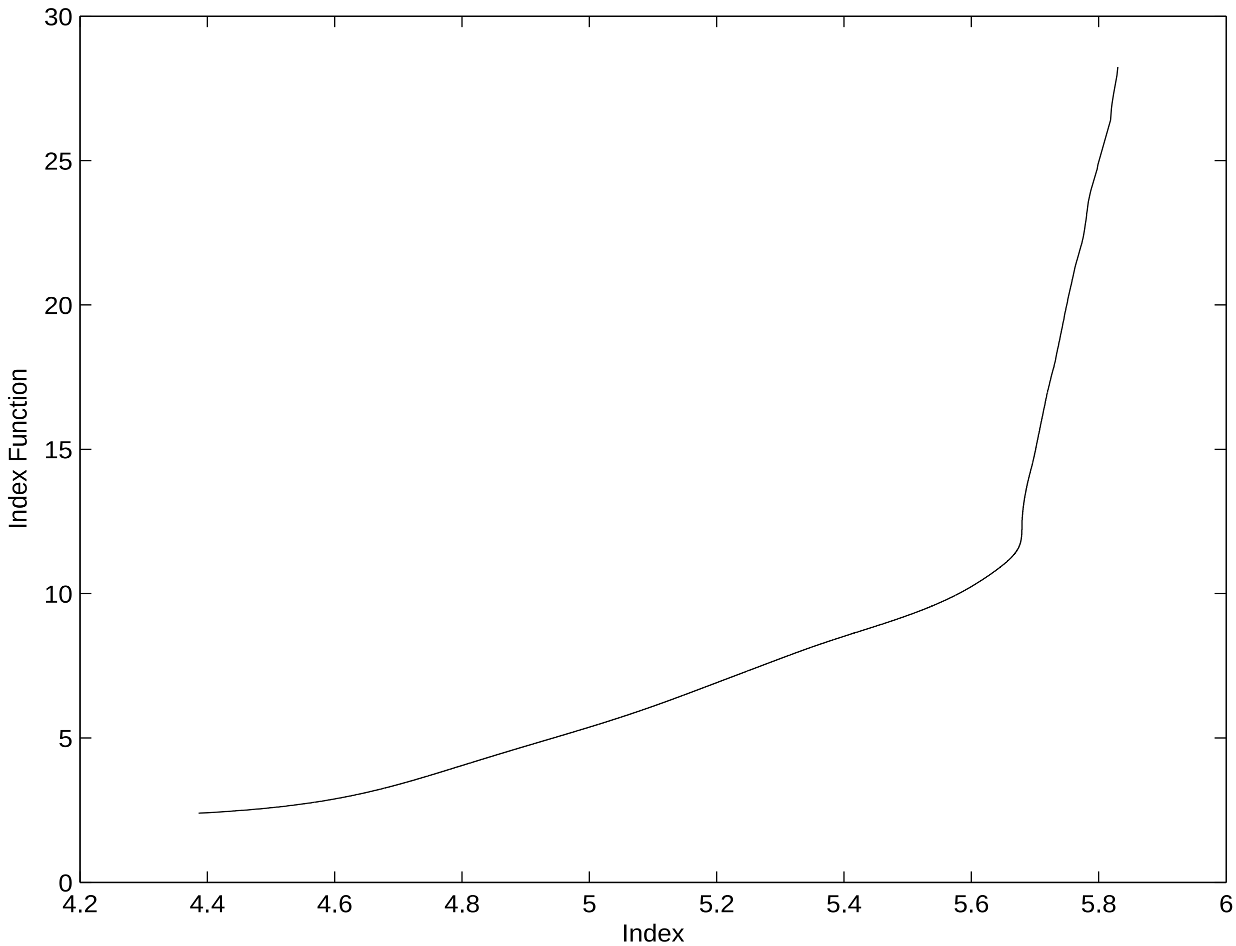


Figure 8: Estimate of index function from wage regression



4.0 NUMERICAL SOLUTIONS OF ASYMMETRIC FIRST PRICE INDEPENDENT PRIVATE VALUES AUCTIONS (WITH JEAN-FRANCOIS RICHARD)

4.1 INTRODUCTION

In this paper, we propose a powerful numerical algorithm to solve first price single object auctions where bidders draw Independent and Private Values (hereafter IPV) from heterogeneous distributions, allowing for subsets of bidders to collude and for a set reserve price. We also provide operational univariate quadratures to evaluate probabilities of winning as well as expected revenues for the bidders and the auctioneer. The latter is used to compute optimal reserves under asymmetric environments. This also enables us to provide insights as to whether collusion among subsets of bidders are sustainable.

We first review some of the relevant literature. Much of the earlier auction literature assumed that bidders draw their signals from a common underlying distribution. Pioneering contributions include Riley and Samuelson [1981], Milgrom and Weber [1982] , Mathews [1983] and Maskin and Riley [1984]. Important theoretical results such as revenue equivalence theorems obtain under symmetry. However, the assumption of symmetry is often far too restrictive for many empirical applications.

Relaxing the symmetry assumption prevents analytical derivation of (first price) bid functions and, therefore, considerably complicates revenue comparisons. Nevertheless, important results have been derived under asymmetry. For example, existence and unicity results under asymmetry can be found in Lebrun [1996, 1999, 2005] or Maskin and Riley [2000a,b]. Furthermore, under stochastic dominance Maskin and Riley [2000a] show that the high bid auction dominates the open bid auction in terms of seller revenue and that the strong bidder (with the stochastically dominant distribution) shades his bid more than the weak bidder. They also provide examples of situations where the seller revenue is higher in open auctions than in high auctions. From a numerical viewpoint, a pioneering contribution which lead to the present paper is found in Marshall et al. [1994] (hereafter MMRS) who proposed a numerical algorithm to compute first price equilibrium bid functions in a two (subgroups of) players asymmetric environment under uniform distributions. Actually, MMRS framework also implicitly assumes stochastic dominance. Marshall and Schulenberg [1998] modified MMRS to accommodate reserve prices set by the auctioneer.

In the present paper we generalize MMRS algorithm to a much broader class of first price asymmetric IPV auction and procurement problems allowing for arbitrary numbers of (subgroups of) players independently drawing their valuations from arbitrary distributions. Common distributions (Exponential, Weibull, Beta, Normal, Lognormal,...) are offered as options in the program. Additional distributions can easily be added by users in the form of a subroutine. Our program takes care of constructing Taylor Series Expansions for these distributions. The only (standard) restriction is that these distributions have common support. Stochastic dominance is not required. This will enable us to investigate whether existing results generalize when stochastic dominance no longer holds. As in MMRS we are actually computing numerical solutions to a system of Ordinary Differential Equations (ODEs) characterizing the first order conditions for a Nash equilibrium. The solution belongs to a class of two-points boundary value problems and is evaluated by recursive application of (low order) Taylor series expansions. Singularity of the system at the origin requires backward extrapolation from an iterated end-point.

For ease of implementation our algorithm currently relies upon equal spacing subdivisions of the support of the component distributions. While this has proved to be numerically stable for most distributions, occasional pathologies (specifically excessive local curvature or densities which are not bounded away from zero on their supports) would require smarter adaptative selection of step size. Such robustification goes beyond the objective of the present paper and are currently addressed by increasing the number of points in the grid as much as needed.

The key advantage offered by our algorithm relative to MMRS lies in its capability to accommodate a wide range of arbitrary distributions, providing us with a powerful tool to investigate whether classic results (revenue equivalence, a.s.o.) extend to situations where symmetry and/or stochastic dominance are no longer assumed. This feature also provides broad flexibility for the analysis of (sub)coalitions.

The paper is organized as follows. The baseline model and the solution algorithm are described in Section 2; Expected revenue calculations are provided in Section 3 for first price auctions and in Section 4 for second price auctions; Numerical examples are presented in Section 5 and Section 6 concludes.

4.2 THE ALGORITHM

4.2.1 Baseline Model

We are considering here a single object IPV first price auction. Risk neutral bidders submit sealed bids. The highest bidder wins and pays his bid. There are N potential bidders. Only those with private valuations above the reserve price R set by the auctioneer submit competitive bids. Bidders are ex-ante heterogeneous. Each bidder belongs to one of n types. Each type is characterized by a distribution function F_i on a common support $[\underline{v},\ \overline{v}]$. There are k_i bidders in group i for a total of $N = \sum_{i=1}^{n} k_i$ (potential) bidders.

Bid functions are denoted by the Greek $\varphi_i, i = 1, \cdots n$. Bidders are assumed to be risk neutral with utility from winning the auction with a bid b given a type v defined as $U_i(v - b) = v - b$. The generalization to constant risk aversion is fairly trivial and will not be discussed here. Clearly, utility from winning the auction is increasing in the individual's signal. Under these assumptions, Proposition 5 of Maskin and Riley [2000b] establishes the existence of a monotonic pure-strategy equilibrium in the standard first price auction. Indeed, Lebrun [1996] has shown that these bid functions are strictly monotone and increasing, therefore, invertible. Inverse bid functions are denoted by the Greek letter $\lambda_i, i = 1, \cdots, n$. Uniqueness of such equilibrium is well established in the case with two types [Lebrun, 1996]. However in the general N player game, equilibrium may not be unique in that we may end up with "non-essential" equilibria [Briesmer and Shubik, 1967].

Here we assume further that F_i is twice continuously differentiable with a density f_i bounded away from zero on $[\underline{v}, \overline{v}]$. Under these assumptions, Lebrun [1999] proves in the general N bidder case that the equilibrium is unique, and that the inverse bid functions have a common support $[R, t_*]$, where t_* is the bid associated with the valuation $\bar{v}$, and R is the reserve price set by the auctioneer. We show in this paper that this equilibrium is amenable to numerical analysis, and presents itself as a natural extension to the methods proposed in MMRS. As such the (numerical) determination of t_* is a critical component of the problem to be solved.

4.2.2 The differential equations

Let $t = \varphi_i(v)$ denote the equilibrium bid submitted by bidder i with private signal $v \in [\underline{v}, \bar{v}]$. For the ease of notation, bidders with signals lower than the reserve R are assumed to bid their signal, whence $\varphi_i(v) = v$ for $v \leq R$. Let $v = \lambda_i(t)$ denote inverse bid functions. Following Lebrun [1999], the λ_i's share a common support $[\underline{v}, t_*]$. For the ease of presentation, we momentarily assume that t_* is known. Bidder i with signal $v \in [R, \bar{v}]$ submits a bid t which is solution of the optimization problem

$$t = \arg\max_{u \in (R, \bar{v})} (v - u) \cdot [F_i(\lambda_i(u))]^{k_i - 1} \Pi_{j \neq i} [F_j(\lambda_j(u))]^{k_j}. \tag{1}$$

The Ordinary Differential Equations (ODEs) associated by the First Order Conditions (FOCs) are given by

$$\Pi_{s=1}^{n} F_s(\lambda_s(t)) = (\lambda_i(t) - t) \cdot [\sum_{j=1}^{n} k_{i,j}^* f_j(\lambda_j(t)) \lambda_j(t) \Pi_{s \neq j} F_s \lambda_s(t))], \tag{2}$$

where $k_{i,i}^* = k_i - 1$ and $k_{i,j}^* = k_j$ for $j \neq i, i : 1 \rightarrow n$. Let $\ell_i(t) = F_i(\lambda_i(t))$. Equation (2) is rewritten as

$$1 = [F_i^{-1}(\ell_i(t)) - t] \cdot \left[\sum_{j=1}^{n} k_{i,j}^* \frac{\ell_j'(t)}{\ell_j(t)} \right], \; i = 1 \rightarrow n \tag{3}$$

The boundary conditions for λ_i and ℓ_i are given by

$$\lambda_i(R) = R, \quad \lambda_i(t_*) = \bar{v}, \quad i : 1 \rightarrow n \tag{4}$$

$$\ell_i(R) = F_i(R), \quad \ell_i(t_*) = 1, \quad i : 1 \rightarrow n \tag{5}$$

respectively. As noted earlier by MMRS under uniform F_is, the system (3) of ODEs is ill behaved at the lower boundary. If, for example, $R = \underline{v}$ then a recursive application of l'Hospital rule for $t \to \underline{v}^+$ produces the result that the right derivative of ℓ_i at $\underline{v}$ is given by

$$\ell_i^\circ = f_i(\underline{v}) \cdot \frac{N}{N-1}, \quad i : 1 \to n \tag{6}$$

and, most importantly, that all higher order right derivatives at $\underline{v}$ are zero. If instead $R > \underline{v}$, then

$$\lim_{t \to R^+} \ell_i'(t) = +\infty, \quad i : 1 \to n \tag{7}$$

as in Marshall and Schulenberg [1998]. Whence, following MMRS, we shall solve the ODEs (3) backward starting from the right boundary $\ell_i(t_*) = 1$, assuming momentarily that t_* is known.

4.2.3 The baseline algorithm

Our algorithm amounts to constructing piecewise polynomial approximations to the ℓ_is from which (as discussed in Section 2.4 below) approximations for the λ_is and φ_is immediately follow. Assuming we just computed $x_{i,0} = \ell_i(t_0)$ where $t_0 \in (R, t_*)$, we describe next how to construct Taylor series expansions for the ℓ_is at t_0 which are then used to compute $x_{i,1} = \ell_i(t_1)$ at the next point $t_1 = t_0 - \Delta t$ where Δt denotes the selected step size. The relevant expansions are denoted as follows:

$$\ell_i(t) = \sum_{j=0}^{\infty} a_{i,j} \cdot (t - t_0)^j, \tag{8}$$

$$\ell_i'(t)/\ell_i(t) = \sum_{j=0}^{\infty} b_{i,j} \cdot (t - t_0)^j, \tag{9}$$

$$F_i^{-1}(\ell_i(t)) - t = \sum_{j=0}^{\infty} p_{i,j}(t - t_0)^j, \tag{10}$$

$$F_i^{-1}(x) = \sum_{j=0}^{\infty} d_{i,j}(x - x_0)^j. \tag{11}$$

Our baseline algorithm relies upon three recursive relationships among the above expansions to construct the $a_{i,j}$s from the $d_{i,j}$s. The relationships between these coefficients and those of other functions of interest such as the F_i's (input) and the φ_i's (output) are discussed in Section 2.4 below. Let J_M denote the selected order of approximations. Step $J (J : 0 \to J_M)$ consists of three parts:

- **The computation of $a_{i,J}$ given $\{(a_{i,j}, b_{i,j}); j < J\}$.** The corresponding recurrence relationship obtains from the identities $\ell_i'(t) = \ell_i(t).\ \ell_i'(t)/\ell_i(t)$ which together with formulae (8) and (9) imply the identities

$$\sum_{j=1}^{\infty} j a_{i,j}(t-t_0)^{j-1} = \left[\sum_{r=0}^{\infty} a_{i,r}(t-t_0)^r\right] \cdot \left[\sum_{s=0}^{\infty} b_{i,s}(t-t_0)^s\right]. \tag{12}$$

Equating the coefficients of order $J-1$ produces the following relationship

$$a_{i,J} = \frac{1}{J}\sum_{r=0}^{J-1} a_{i,r} b_{i,J-r-1}, \quad (i : 1 \to n; J : 1 \to J_M) \tag{13}$$

with initial conditions

$$a_{i,0} = \ell_i(t_0), \quad b_{i,0} = \ell_i'(t_0)/\ell_i(t_0), \quad i : 1 \to n. \tag{14}$$

- **The computation of $p_{i,J}$ given $\{(a_{i,j} d_{i,j}); j \leq J\}$.** The corresponding relationship obtains by application of Lemma 1 in Appendix A to the composition of F_i^{-1} (input) and ℓ_i (output from (13)), accounting for the additional factor $t = [t_0 + (t-t_0)]$. Whence we have

$$p_{i,J} = \sum_{r=1}^{J} d_{i,r}\theta_{i,r,J} - z_J, \quad (i : 1 \to n; J : 1 \to J_M) \tag{15}$$

$$\theta_{i,r,J} = \sum_{s=1}^{J-r+1} a_{i,s}\theta_{i,r-1,J-s}, \quad (r : 1 \to J) \tag{16}$$

with $z_0 = z_1 = 1, z_J = 0$ for $J > 1$, and initial conditions

$$p_{i,0} = F_i^{-1}(x_{i,0}), \quad \theta_{i,0,0} = 1 \quad (i : 1 \to n) \tag{17}$$

- **The computation of $b_{i,j}$ given $\{p_{i,j}; j \leq J); (b_{i,j}; j < J)\}$.** The corresponding relationships originate from the ODEs themselves. Substituting expansions (9) and (10) into equation (3) produces the identities

$$1 = \left[\sum_{r=0}^{\infty} p_{i,r}(t-t_0)^r\right]\left[\sum_{\ell=1}^{n} k_{i,\ell}^{*} \sum_{s=0}^{\infty} b_{\ell,s}(t-t_0)^s\right], \tag{18}$$

for $i : 1 \to n$ and $\ell : 1 \to n$ Equation (18) can be rewritten as

$$1 = \sum_{j=0}^{\infty}\left[\sum_{\ell=1}^{n} k_{i,\ell}^{*}\left(\sum_{r=0}^{j} p_{i,r} b_{\ell,j-r}\right)\right](t-t_0)^j. \tag{19}$$

Equating the coefficients of $(t-t_0)^J$ to 0 for $J > 1$ and rearranging the corresponding identities into matrix form produces the following vectorial recurrence relationship

$$P_0(I_n - i_n k')b_J = c_J, \tag{20}$$

where $P_0 = \text{diag}(p_{1,0}, \cdots, p_{n,0})$, I_n is the identity matrix of order $n, i'_n = (1 \cdots 1)$, $k' = (k_1, \cdots k_n), b'_J = (b_{1,J}, \cdots, b_{n,J}), c_0 = -i_n$ and

$$c_J = \begin{pmatrix} \vdots \\ \sum_{\ell=1}^{n} k_{i,\ell}^{*}\left(\sum_{r=1}^{J} b_{\ell,J-r}\right) \\ \vdots \end{pmatrix}, J > 0 \tag{21}$$

Standard formulae for partitioned matrices produce the following expressions for the determinant and inverse of $(I_n - i_n k')$:

$$|I_n - i_n k'| = 1 - N \quad (I_n - i_n k')^{-1} = I_n - \frac{i_n k'}{N-1}. \tag{22}$$

Formulae (12) to (22) for $J : 0 \to J_M$ define our baseline recurrences algorithm for the evaluation of Taylor Series expansions at an arbitrary base point $t_0 \in (R, t_*)$, from which function values at a new point $t_1 = t_0 - \Delta t$ are approximated.

4.2.4 Additional details

A number of additional details need be addressed next to complete an operational implementation of our baseline algorithm.

4.2.4.1 Numerical Search for t_* With very few exceptions, one of which is found in Appendix A of MMRS, t_* cannot be found analytically. Instead, we shall rely upon the unicity result in Lebrun [1999] together with the initial conditions (5) to define t_* as

$$t_* = \arg \min_{t_f \in (R,\bar{v})} \sum_{i=0}^{n} \left[\ell_i(R|t_f) - F_i(R)\right]^2 \tag{23}$$

where $\ell_i(\cdot|t_f)$ denotes the solutions to the ODEs in (3) under a tentative terminal condition $\ell_i(t_f) = 1$. Note that since

$$\lim_{t \to R^+} [F_i^{-1}(\ell_i(t)) - t] = 0 \tag{24}$$

the coefficients $(p_{i,0}; i: 1 \to n)$ should be zero for $t_0 = R$. This prevents us from solving the system (20) at $t_0 = R$ but we do not need to do so. Instead we compute $\ell_i(R|\cdot)$ from the Taylor series expansions at $t_0 = R + \Delta t$. Substituting these approximate values in the objective function (23) suffices to produce very accurate estimates of t_* for Δt small enough. Alternatively, once we have an estimate of $x_{i,0} = \ell_i(t_0)$, we can also compute $p_{i,0} = F_i^{-1}(x_{i,0})$ and use as objective function $\sum_{i=1}^{n} p_{i,0}^2$. As for the actual minimization, we rely upon the simplex subroutine AMOEBA which is numerically very efficient for our problem.

4.2.4.2 Additional Taylor Series Expansions As described in Section 2.4.4 below, our algorithm constructs Taylor Series expansions of F_i^{-1} to compute those of $\ell_i, i: 1 \to n$. Very little work is required to reformulate it in terms of the primitives of the problem, the distribution F_i and the bid functions φ_i. First, note that the inverse bid functions λ_i are given by

$$\lambda_i(t) = t + \sum_{j=0}^{\infty} p_{i,j}(t - t_0)^j \tag{25}$$

Next, we can rely upon Lemma 2 in Appendix A to transform Taylor Series expansions of F_i and λ_i into those of F_i^{-1} and φ_i, respectively.

4.2.4.3 Support Conditions As expected from the formulation of equation (3), our algorithm can become numerically unstable if the ℓ_is get too close to zero. This can occur on regions of very low probability. Our current program implementation requires that tail areas of (very) low probability be truncated away. Note that such truncations are commonly imposed in empirical applications since most estimation techniques for auction models critically rely upon the invertibility of bid functions and lack robustness relative to tail area behavior of the latter. See e.g., Donald and Paarsch [1996], Laffont et al. [1995] or Florens et al. [2004]. See also Marshall et al. [2005] for an empirical application where truncation of an assumed Weibull distribution had to be imposed for estimation purposes. Note, however, that distributions of interest for the tractability of their order statistics (e.g. exponential, Weibull or extreme value distribution) have unbounded support. In practice any such distribution F_i with unbounded support will be replaced by a truncated vesion thereof

$$F_i^*(v) = \frac{F_i(v) - F_i(\underline{v})}{F_i(v) - F_i(\underline{v})}, \quad \underline{v} < v < \overline{v} \tag{26}$$

Transforming the Taylor Series expansion of F_i^{-1} into that of F_i^{*-1} follows by application of Lemma 1 in Appendix to the following composite function

$$F_i^{*-1}(u) = F_i^{-1}(F_i(\underline{v}) + u[F_i(\overline{v}) - F_i(\underline{v})]), \quad 0 \leq u \leq 1 \tag{27}$$

Such transformations are automated in our computer programs.

4.2.4.4 Automated Taylor Series Expansions Analytical Taylor Series expansions for inverse cdf's are available for a number of standard distributions such as the extreme value distributions which are commonly assumed in empirical applications. However, there are situations where this is not the case. One such important situation is discussed in Section 2.4.5 below where we analyze non-inclusive coalitions. Other important examples would be applications where empirical and/or non-parametric cdf's have been numerically evaluated.

In order to accommodate such situations our program includes a fully automated numerical procedure for the computation of (piecewise) Taylor Series expansions for the inverses of arbitrary cdf's. This procedure incorporates the following steps:

1. We construct an equally spaced grid $\{u_j; j : 1 \to J\}$ for the interval $[0,1]$;

2. Using a standard root finder we compute the corresponding (unequally spaced) grid for the inverse cdf $F^{-1}, \{v_j; v_j = F^{-1}(u_j); j : 1 \rightarrow J\}$;
3. Next, we construct a B-spline interpolator for F^{-1}. Specifically, we invoke the IMSL subroutines DBSNAK (to construct a knot sequence) and DBSINT (to compute B-spline coefficients), see e.g., de Boor (1978) for numerical details.
4. Finally, we invoke the IMSL subroutine BSCPP to convert the B-spline interpolator into a piecewise polynomial approximation, which provides the Taylor Series expansion needed for our algorithm.

4.2.4.5 Non-inclusive Coalitions The object of our paper is not that of providing a theoretical investigation of the stability of non-inclusive coalitions within a first price asymmetric framework (which in many cases would likely required repeated games concepts). Nevertheless, we can use our algorithm to numerically investigate whether such non-inclusive coalitions could potentially be incentive compatible and also whether a strategic auctioneer could reduce the profitability of collusions. Pioneering examples of such computations under (ex-ante symmetric) uniform distributions can be found in MMRS and Marshall and Schulenberg [1998]. See also Marshall and Marx [2005] for an in-depth discussion of incentive compatible mechanisms for non-inclusive cartels as well as an extensive list of related references.

Short of such theoretical analysis our algorithm can be used to numerically evaluate bid functions and expected revenues in the presence of non-inclusive cartels, as long as one treats such a cartel as a single representative bidder. At minimum, such computations can provide useful insight on potential incentives to defect and on the auctioneer's capability to reduce cartels' profitability. For example, MMRS had already illustrated the fact that within an ex-ante uniform symmetric framework outsiders benefit more than insiders (on a per capita basis) from the presence of a non-inclusive cartel. One would not expect such findings to generalize to asymmetric scenarios. In particular, there exist numerous real-life illustrations of the viability of non-inclusive cartels consisting, for example, of better informed players. One such situation was recently highlighted by the conviction of seven leading stamp dealers and auctioneers who, for several years, had agreed not to compete against one another at estate auctions of stamp collections.

Specifically, in the context of our program, an arbitrary cartel consisting of $u = \sum_{i=1}^{n} u_i$ players, were u_i denotes the number of players of type i is treated as a single player drawing her signal from the corresponding highest order statistics cdf.

$$F^*(v) = \Pi_{j=1}^{n} \left[\frac{F_j(v) - F_j(\underline{v})}{F_j(\overline{v}) - F_j(\underline{v})}\right]^{u_j} \tag{28}$$

Taylor Series expansions for the inverse of F^* are automatically produced by application of the numerical procedure described in Section 2.4.4 above. It is also trivial to verify that all probability and expected revenue calculations described below remain valid under such scenarios with the only modification that the revenue computed represents the cartel's total expected revenue. As discussed above, we do not discuss allocation rules among cartel's members and only provide per capita comparisons between insiders and outsiders.

4.3 PROBABILITIES OF WINNING, EXPECTED REVENUES AND OPTIMAL RESERVE PRICE

In this section we demonstrate that expected revenues and probabilities of winning when the auctioneer sets a reserve price R can all be expressed as simple univariate integrals (quadratures) of products of the functions evaluated by our algorithm over the interval (R, t_*), where t_* itself is an implicit function of R.

The following conditions have to be met for a bidder from group i to win

$$R < v_i < \overline{v} \quad \text{and} \quad v_j < \lambda_j(\lambda_i^{-1}(v_i)) \quad \text{for} \quad j \neq i \tag{29}$$

Whence the probability that group i wins is given by

$$P_i(R) = k_i \int_R^{\overline{v}} f_i(v) \Pi_{j=1}^{n} [F_j(\lambda_j(\lambda_i^{-1}(v)))]^{k_{i,j}^*} dv \tag{30}$$

where $k_{i,i}^* = k_i - 1$ and $k_{i,j}^* = k_j$ for $j \neq i$, as in Section 2 above. Introducing the change of variable $t = \lambda_i^{-1}(v)$ and rearranging terms yields the following operational expression

$$P_i(R) = k_i \int_R^{t_*} \frac{\ell_i'(t)}{\ell_i(t)} \cdot \Pi_{j=1}^{n} \, [\ell_j(t)]^{k_j} dt \tag{31}$$

Note that

$$\sum_{i=1}^{n} P_i(R) = \int_{i=1}^{t_*} k_i \ell_i'(t) \Pi_{j=1}^{n} [\ell_j(t)]^{k^{*i,j}}$$

$$= \int_{R}^{t_*} \left[\Pi_{j=1}^{n} \left[\ell_j(t) \right]^{k_j} \right]' dt = 1 - \Pi_{j=1}^{n} \left[F_j(R) \right]^{k_j} \tag{32}$$

confirming the obvious result that the probability that the auctioneer retains the item is given by

$$P_0(R) = \Pi_{j=1}^{n} \left[F_j(R) \right]^{k_j} \tag{33}$$

Group i's expected revenue is given by

$$V_i(R) = k_i \int_{R}^{\bar{v}} [v - \varphi_i(v)] \cdot f_i(v) \cdot \Pi_{j=1}^{n} \left[F_j(\lambda_j(\lambda_i^{-1}(v))) \right]^{k_{i,j}^*} dv \tag{34}$$

which can be rewritten as

$$V_i(R) = k_i \int_{R}^{t_*} \left[F_i^{-1}(\ell_i(t)) - t \right] \cdot \frac{\ell_i'(t)}{\ell_i(t)} \cdot \Pi_{j=1}^{n} \left[\ell_j(t) \right]^{k_j} \tag{35}$$

Per capita expected revenue within group i's accounting for subcoalitions $(u_i \geq 1)$ is then given by $V_i(R)/(k_i \cdot u_i)$. Finally, assuming that the auctioneer receives a fixed percentage of all winning bids, her revenue is proportional to

$$V_a(R) = \sum_{i=1}^{n} k_i \int_{R}^{\bar{v}} \varphi_i(v) f_i(v) \Pi_{j=1}^{n} \left[F_j(\lambda_j(\lambda_i^{-1}(v))) \right]^{k_{i,j}^*} dv \tag{36}$$

$$= \int_{R}^{t_*} t \cdot \left[\Pi_{j=1}^{n} \left[\ell_j(t) \right]^{k_j} \right]' dt \tag{37}$$

Integration by parts produces the following expression

$$V_a(R) = t_* - R \Pi_{j=1}^{n} \left[F_j(R) \right]^{k_j} - \int_{R}^{t_*} \Pi_{j=1}^{n} \left[\ell_j(t) \right]^{k_j} dt \tag{38}$$

Note that formulae (31), (35) and (38) all depend upon univariate integrals of products of the functions which are being evaluated by our algorithm over a fine grid of values of t in (R, t_*). Therefore, these integrals can be evaluated by univariate quadrature as immediate byproducts of our algorithm. As we typically use grids with anywhere from $N = 500$ to $N = 10,000$ equally spaced points, we can rely upon the extended Simpson's rule - see Press et al. [1986] or Abramowitz and Segun [1968][formula 2.5,4.5] (formula 4.1.13) - with remainder proportional to N^{-4} to compute numerically highly accurate estimates of all relevant probabilities and expected revenues.

Moreover, the use of a fixed number of equally spaced grid points implies that these numerical integrals will be continuous functions of a R. Whence numerical simplex maximization of $V_a(R)$ w.r.t R will itself be numerically very accurate. Note that t_* in formulae (31) to (38) is an implicit function of R so that our algorithm has to be rerun for each value of R selected by AMOEBA.

4.4 ASYMMETRIC SECOND PRICE AUCTIONS

One of the immediate intended use of our new algorithm is that of running comparisons between first and second price auctions under a variety of asymmetric environments. In order to do so we need to derive operational expressions for expected revenues under second price auctions. While Vickrey's logic still applies whereby bidders bid their private values, expected revenue calculations are more complex than under first price due to a wider range of scenarios for the price paid by the winner.

Several pricing scenarios need to be considered. Focusing our attention on group i, let $v_1 > v_2$ denote the two highest order statistics in group i (implicitly assuming that $k_i > 1$, but one verifies that the formulae derived below also apply for $k_i = 1$) and let w_j denote the highest order statistic in group j $(j \neq i)$. The following pricing scenarios are relevant:

$$E_{i,R}: \text{ price is } R; \text{ i.e., } v_1 > R,\ v_2 < R,\ w_j < R, \text{ for } j \neq i$$

$$E_{i,i}: \text{ price is } v_2; \text{ i.e., } v_2 > R,\ v_2 > w_j, \text{ for } j \neq i$$

$$E_{i,j}: \text{ price is } w_j; \text{i.e., } w_j > R,\ w_j > v_2,\ w_j > w_\ell, \text{ for } \ell \neq j, i.$$

Probabilities and expected revenues are indexed conformally. The relevant densities are

$$k_i(v_1, v_2) = k_i(k_i - 1)\, f_i(v_1)\, f_i(v_2)\, [F_i(v_2)]^{k_i - 2}, v_1 > v_2 \tag{39}$$

$$k_j(w) = k_j\, f_j(w)\, \left[F_j(w)\right]^{k_j - 1} \tag{40}$$

Note that by relying upon the $k^*_{i,j}$ notation introduced in formula (2), a common treatment applies to scenario $E_{i,i}$ and $E_{i,j}(j \neq i)$. The probability that group i wins and pays either v_2 or w_j is given by

$$\sum_{j=1}^{n} P_{i,j}(R) = k_i \left\{ \sum_{j=1}^{n} k^*_{i,j} \int_R^{\overline{v}} f_i(v_1) \cdot \left\{ \int_R^{v_1} f_j(v) \cdot \left[F_j(v)\right]^{k^*_{i,j}-1} \right.\right.$$
$$\left.\left. \Pi_{\ell \neq j} [F_\ell(v)]^{k^*_{i,\ell}} dv \right\} dv_1 \right\} \tag{41}$$

where v denotes v_2 for $j = i$ and w_j for $j \neq i$. As in Section 3 above, we first apply integration by parts to the outer integral and regroup terms obtaining the following expression

$$\sum_{j=1}^{n} P_{i,j}(R) = k_i \cdot \int_R^{\overline{v}} [1 - F_i(v)] \cdot \left[\Pi_{j=1}^{n} [F_j(v)]^{k^*_{i,j}}\right]' dv \tag{42}$$

A second integration by part produces the result

$$\sum_{j=1}^{n} P_{i,j}(R) = k_i \int_R^{\overline{v}} f_i(v) \Pi_{j=1}^{n} \left[F_j(v)\right]^{k^*_{i,j}} dv$$
$$-k_i \left[1 - F_i(R)\right] \Pi_{j=1}^{n} \left[F_j(R)\right]^{k^*_{i,j}} \tag{43}$$

Note that the second term in the right hand side of formula (43) represents $P_{i,R}(R)$. Whence the probability that group i wins is given by

$$P_i(R) = P_{i,R}(R) + \sum_{j=1}^{n} P_{i,j}(R) = k_i \int_R^{\overline{v}} f_i(v) \Pi_{j=1}^{n} \left[F_j(v)\right]^{k^*_{i,j}} dv \tag{44}$$

Note that

$$\sum_{i=1}^{n} P_i(R) = \int_R^{\overline{v}} \left(\Pi_{j=1}^{n} \left[F_j(v)\right]^{k_j}\right)' dv = 1 - \Pi_{j=1}^{n} \left[F_j(R)\right]^{k_j} = 1 - P_0(R) \tag{45}$$

Next, we derive the auctioneer expected revenue which is given by

$$V_a(R) = \sum_{i=1}^{n} \left\{ P_{i,R}(R) + k_i \left\{ \sum_{j=1}^{n} k^*_{i,j} \int_R^{\overline{v}} f_i(v)_1 \left[\int_R^{v_1} v f_j(v) \right.\right.\right.$$
$$\left.\left.\left. \left[F_j(v)\right]^{k^*_{i,j}-1} \Pi_{\ell \neq j} \left[F_\ell(v)\right]^{k^*_{i,\ell}} dv \right] dv_1 \right\}\right\} \tag{46}$$

The same integration by parts sequence as for the probability produces the following expression paralleling formula (44)

$$V_a(R) = -\sum_{i=1}^{n} k_i \int_R^{\overline{v}} (v[1 - F_i(v)])' \Pi_{j=1}^{n} [F_j(v)]^{k^*_{i,j}} dv \tag{47}$$

$$= \int_R^{\overline{v}} v \left(\Pi_{j=1}^n \left[F_j(v) \right]^{k_j} \right)' dv - \sum_{i=1}^n k_i \int_R^{\overline{v}} [1 - F_i(v)] \, \Pi_{j=1}^n \left[F_j(v) \right]^{k_{i,j}^*} dv \tag{48}$$

or, equivalently after a third integration by parts

$$V_a(R) = \overline{v} - RP_0(R) - \int_R^{\overline{v}} \Pi_{j=1}^n \left[F_j(v) \right]^{k_j} dv$$

$$- \sum_{i=1}^n k_i \int_R^{\overline{v}} [1 - F_i(v)] \, \Pi_{j=1}^n \left[F_j(v) \right]^{k_{i,j}^*} dv \tag{49}$$

The expected revenue for group i is derived in the same way. We first have

$$V_i(R) = k_i \left\{ \left[\int_R^{\overline{v}} (v - R) \, f_i(v) dv \right] \Pi_{j=1}^n \left[F_j(R) \right]^{k_{i,j}^*} \right.$$

$$\left. + \sum_{j=1}^n k_{i,j}^* \int_R^{\overline{v}} f_i(v_1) \left[\int_R^{v_1} (v_1 - v) \, f_j(v) \left[F_j(v) \right]^{k_{i,j}^* - 1} \Pi_{\ell \neq j} \left[F_\ell(v) \right]^{k_{i,\ell}^*} dv \right] dv_1 \right\} \tag{50}$$

Integration by parts of the first integral in v and of the outer integral in v_1 produces the simpler expression

$$V_i(R) = k_i \left\{ \left[(\overline{v} - R) - \int_R^{\overline{v}} F_i(v) dv \right] \Pi_{j=1}^n \left[F_j(R) \right]^{k_{i,j}^*} \right\}$$

$$+ \int_R^{\overline{v}} (\overline{v} - v) \left(\Pi_{j=1}^n \left[F_j(v) \right]^{k_{i,j}^*} \right)' dv + \int_R^{\overline{v}} F_i(v) \left[\Pi_j^n [F_j(v)]^{k_{i,j}^*} - \Pi_j^n [F_j(R)]^{k_{i,j}^*} \right] dv \Bigg\} \tag{51}$$

Integration by parts of the second factor in the right hand side of formula (51) and cancellations produce the following operational expression for V_i

$$V_i(R) = k_i \int_R^{\overline{v}} [1 - F_i(v)] \Pi_{j=1}^n [F_j(v)]^{k_{i,j}^*} dv \tag{52}$$

As above, per capita expected revenue in group i is given by $V_i(R)/(k_i \cdot u_i)$.

As was the case for the first price auction, formulae (44), (49) and (52) are numerically evaluated by quadrature. All probabilities and expected revenues calculations for first price and second price auctions have been incorporated in our algorithm allowing for automated comparisons between first and second price auctions under a wide variety of asymmetric scenarios. Examples are provided below.

4.5 EXAMPLES

In this section we present three numerical illustrations of the capabilities of our program. The parameters and, in particular, the truncation range $(\underline{v}, \overline{v})$ were selected to produce graphically well separated bid functions. For the first two examples, all type distributions are truncated Weibull of the form given in formula (27) together with

$$F_i(v) = 1 - \exp\left[-(v/a)^{b_i}\right] \tag{53}$$

4.5.1 Example 1 (3 individual bidders)

We first consider 3 individual bidders (low, high, median types) and compute their first price asymmetric bid functions without reserve as well as with optimal reserve. We also compute bidder's expected revenues (per capita), bidder's probabilities of winning, auctioneer's expected surplus and probability of retaining the item (under a reserve). The same statistics are also computed for second price auctions. Graphs of the first price asymmetric bid functions with and without reserve are provided in Figure 1. Relevant statistics are regrouped in Table 21. We note that the reserve price impacts the bidders differently, the larger impact being obviously felt by the high-type bidder. We also note that in the absence of reserve first price is more profitable for the auctioneer (by about 5%) but that the ordering is reversed under optimal reserve. The high-type bidder always prefers second price, especially obviously in the absence of a reserve.

4.5.2 Example 2 (2 individiual bidders)

This example illustrates the fact that asymmetric bid functions can cross one another once stochastic dominance no longer applies. Hazard functions are monotone for Weibull distributions. Our choice of shape parameters for this example implies that the hazard function of bidder 1 is increasing $(b_1 = 1.5)$ and that of bidder 2 is decreasing $(b_2 = 0.5)$. With means close to one another it implies that the distribution functions cross one another at $v = 1.45$. It also implies that as illustrated by Figure 2, the two bid functions cross one another at $v = 1.7$. Expected revenues and surpluses together with probabilities of winning with and without reserve prices are regrouped in Table 22. We note that first and second price auctions are virtually revenue equivalent even though, in the absence of reserve, second price favors bidder 1. A systematic numerical investigation of whether revenue equivalence holds under particular asymmetric scenario goes beyond the objectives of the present paper but belongs to our research agenda.

4.5.3 Example 3 (non inclusive cartels)

MMRS offer a numerical investigation of incentive compatibility within subcoalitions when individual bidders all draw their valuations from a common uniform distribution. Within this (single object) framework they find that bidders outside the coalitions benefit more than those inside. Here we consider instead an asymmetric scenario where high type bidders collude together in order to protect their informational advantage over low type bidders.

This example is inspired by a recent court case where a group of prominent stamp auctioneers and dealers were found guilty of collusion at estate auctions. While their cartel operated for several years, our example illustrates the fact that such noninclusive cartels could be incentive compatible even within a single object framework (ignoring, however, proxy defections as analyzed by **?**. We consider two bidders of high type (H) against four (non collusive) bidders of low type (L). Signals are lognormally distributed with a common standard deviation 0.35 and means 1.35 and 0.75, respectively. The common support for signals is the interval [1.5, 6.0]. Results for the non collusive benchmark scenario are reported in Table 23. Graphs of the corresponding bid functions with and without optimal reserves are provided in figure 3. Results for subcoalitions $\{$H,H$\}$ and $\{$H,H,L$\}$ are reported in Table 24, and figures 4 and 5.

The impact of the collusion among high types is greatest under second price auction. Under first price, low types also benefit from the presence of the cartel (even more than high types percentage wise). Reserve is most effective under second price (and would be even more effective if items kept by the auctioneer had a resale value). In the future, we plan to investigate whether the effectiveness of optimal reserve requires precise knowledge of the cartel composition by the auctioneer.

4.6 NUMERICAL ACCURACY AND COMPUTATIONAL TIME

Accuracy of the numerical approximation of the equilibrium bid functions depend primarily on two variables. The first is how fine a grid is chosen on which to evaluate the component distributions. A finer grid leads to higher accuracy of the numerical approximations. The second variable is the order of the Taylor Series approximations chosen approximate these distributions. A higher order Taylor series expansion does not necessarily lead to higher accuracy. Indeed, an order of approximation that is too high can lead to significant numerical pathologies.

A reliable method for evaluating accuracy consists of computing pointwise best response for each individual bidder and comparing them to the NE strategies. Given bidder i's signal, his best response depends on the distribution functions and (inverse) bid functions of his competitors. His best response function does not depend on his own distribution function. This is seen clearly in equation (3). Thus, given his competitors equilibrium strategies and distribution functions, we can use equation (3) to compute pointwise the best response of bidder i. His best response function can then be matched against the equilibrium function computed by the algorithm and difference between these two functions provides a measure of the accuracy of the algorithm. A reasonable metric, and the one we use in this paper is the root of the mean squared deviation (RMSE) between the equilibrium bid function and the best response function.

An important illustration of the usefulness of such comparisons is provided by example 3. Figure 5 reveals a curious "blip" in the bid function of the coalition. The bid function dips down between private values of 2.0 and 2.5. This gives rise to the question of whether this is the result of a numerical error, or if the blip is a rational response by the collusion to the strategies of the outsiders. This question can be answered by the method of verification we just described. Figure 6 reproduces the equilibrium bid function of the coalition and also plots the best response of the coalition computed as described in section 5. The reaction function (bold dotted line) coincides exactly with the computed bid function (solid line). This confirms that the blip is indeed an equilibrium reaction by the coalition to the strategies of its competitors.

Higher accuracy of the numerical approximations to the equilibrium bid functions comes at the cost of increased computational time. For a small number of types of bidders, one can be liberal with the size of the grid and the order of the Taylor Series expansions. However, for models with a large number of types of bidders, the computational time increases potentially significantly. Though this is not a significant obstacle in our current applications, computational time can quickly become a problem in others. An important example is that of empirical applications where the algorithm would be instrumental in the estimation of the underlying private values distributions. In this case, the model would have to be solved for each trial value of the vector of parameters of the private values distributions. and one might have to be conservative with the size of the chosen grid. In this section we present a small study of the trade off between accuracy and speed as controlled by these two variables.

Table 25 reports the computational time and the RMSE between the equilibrium bids and reaction functions of two bidders. The first panel fixes the order of Taylor series expansions to 5, and evaluates these statistics for the number of grid points being 500, 1000, 1500, and 2000. The table reveals that the computational time increases linearly with the number of grids. The computational time increases by 0.11 seconds with a one point increase in the number of grid points. The RMSE for each bidder is decreasing and concave in the number of grid points. However, the decrease in the RMSE is very small for large increases in the number of grid points. This suggests that there is not a lot to gain in terms of accuracy by increasing the number of grid points. A relative small number of grid points like 500 provides almost the same numerical accuracy as a larger number of grid points like 2000.

The second panel of Table 25 fixes the number of grid points to 500 and increases order of Taylor series expansions incrementally from 2 to 5. The computational time increases linearly by approximately 7 seconds with each increase in the order of the Taylor series expansions. Interestingly, the numerical accuracy of the bid functions are invariant to the order of Taylor series expansion. The third panel of Table 25 bolsters this conclusion.

The conclusion of this exercise is that the investigator looses very little in terms of numerical accuracy by using a relatively small number of grid points and order of Taylor series expansions. The time saving is however significant.

4.7 PROCUREMENTS

The proposed algorithm is modified to the procurement problem under the same environment as that of the auctions problem. We provide all the options that are provided in the auctions problem, including asymmetry, collusion, the calculation of optimal reserve price, expected revenue to the auctioneer, expected surplus to the bidders, probabilities of winning, and the reaction functions. The necessary modifications are minor and are described briefly in this section. For the procurements problem bidder i with signal $v \in [\underline{v}, R]$ submits a bid t which is solution of the optimization problem

$$t = \arg\max_{u \in (\underline{v}, R)} (u - v) \cdot [H_i(\lambda_i(u))]^{k_i - 1} \Pi_{j \neq i} [H_j(\lambda_j(u))]^{k_j}, \tag{54}$$

where $H_i(x) = 1 - F_i(x)$. The Ordinary Differential Equations (ODEs) associated by the First Order Conditions (FOCs) are given by

$$-1 = [H_i^{-1}(\ell_i(t)) - t] \cdot \left[\sum_{j=1}^{n} k_{i,j}^{*} \frac{\ell_j'(t)}{\ell_j(t)} \right], \ i = 1 \to n \tag{55}$$

where $\ell_i(t) = H_i(\lambda_i(t))$. The boundary conditions for λ_i and ℓ_i are given by

$$\lambda_i(R) = R, \quad \lambda_i(t_*) = \underline{v}, \quad i : 1 \to n \tag{56}$$

$$\ell_i(R) = H_i(R), \quad \ell_i(t_*) = 1, \quad i : 1 \to n \tag{57}$$

respectively. Under this setup, the algorithm to compute equilibrium bids here mimics exactly the one derived in section 2 to compute equilibrium bids in the auctions environment, except for two changes. The first is that the RHS of equation (21) is now i_n instead of $-i_n$. The second is that the recursion on the grid of t is a forward iteration instead of a backward iteration. The probabilities of winning and expected revenues in the first price procurements environment are given as follows:

$$P_i(R) = -k_i \int_{t_*}^{R} \frac{\ell_i'(t)}{\ell_i(t)} \cdot \Pi_{j=1}^{n} [\ell_j(t)]^{k_j} dt, \tag{58}$$

$$P_0(R) = \Pi_{j=1}^{n} \left[H_j(R)\right]^{k_j} \tag{59}$$

$$V_i(R) = -k_i \int_{t_*}^{R} \left[H_i^{-1}(\ell_i(t)) - t\right] \cdot \frac{\ell_i'(t)}{\ell_i(t)} \cdot \Pi_{j=1}^{n} \left[\ell_j(t)\right]^{k_j}, \tag{60}$$

$$V_a(R) = t_* - R\Pi_{j=1}^{n} \left[H_j(R)\right]^{k_j} + \int_{t_*}^{R} \Pi_{j=1}^{n} \left[\ell_j(t)\right]^{k_j} dt. \tag{61}$$

The corresponding probabilities of winning and expected revenues in the second price procurements environment are given as follows:

$$P_i(R) = -k_i \int_{t_*}^{R} \frac{\ell_i'(t)}{\ell_i(t)} \cdot \Pi_{j=1}^{n} \left[\ell_j(t)\right]^{k_j} dt, \tag{62}$$

$$P_0(R) = \Pi_{j=1}^{n} \left[H_j(R)\right]^{k_j} \tag{63}$$

$$V_a(R) = \underline{v} - RP_0(R) + \int_{\underline{v}}^{R} \Pi_{j=1}^{n} \left[H_j(v)\right]^{k_j} dv$$
$$+ \sum_{i=1}^{n} k_i \int_{\underline{v}}^{R} \left[1 - H_i(v)\right] \Pi_{j=1}^{n} \left[H_j(v)\right]^{k_{i,j}^*} dv \tag{64}$$

$$V_i(R) = k_i \int_{\underline{v}}^{R} [1 - H_i(v)] \Pi_{j=1}^{n} [H_j(v)]^{k_{i,j}^*} dv \tag{65}$$

4.8 DISCUSSION OF THE ALGORITHM

With all the described ingredients put together, we have a program that is fully automated and very flexible. The program includes several candidate private values distributions, namely the two parameter Weibull, the Beta, the Normal and the Lognormal distributions. These distributions can be combined to produce hybrid distributions, and other distributions can be trivially added. The necessary Taylor series expansions of the inverse distributions are also fully automated. Furthermore, the program allows for the analysis of a wide variety of collusive arrangements. As an illustration of how user friendly the program is, we present verbatim below the input sequence from example 3 where the two high types and one low type collude to compete against the remaining three low types.

Enter 1 if you want auctions, 2 if you want procurements: 1

Enter number of types: 2

Enter order of Taylor series expansion (5 recommended): 5

Enter the number of subintervals of (t0,t*) to consider: 2000

Enter number of coalitions of each type, separating by

space For example, if you entered 2 for number of types and you

want 3 coalitions in the first and 2 in the second, then enter 3 2: 1 3

Enter lower bound of the support of the distribution of private values: 1.5

Enter upper bound of the support of the distribution of private values: 6.0

Enter reserve price: 1.5

Here is a menu of cdfs to choose from

1 - two parameter Weibull

2 - Beta

3 - Normal

4 - Lognormal

Please enter the number of cdfs to be used: 2

Enter the index of the cdfs you choose to use: 4 4

Enter scale and shape parameter of Lognormal distribution: 1.35 0.35

Enter scale and shape parameter of Lognormal distribution: 0.75 0.35

For type 1 Enter sequence of zeros and ones corresponding to the use of the cdfs: 2 1

For type 2 Enter sequence of zeros and ones corresponding to the use of the cdfs: 0 1

TYPE	MEAN	STD.DEV
1	4.3793	0.8563
2	2.4353	0.7241

Enter 1 if you wish to compute the optimal reserve

Enter 2 if you wish to keep your reserve price: 1

Enter output file name: illustration.txt

Time taken: 290.000s

Writing grid points and bids to file: illustration.txt

Enter 1 for expected revenue and bidder surplus

Enter 2 if not: 1

Enter output file: rillustration.txt

Enter 1 to compute the best response function

Enter 2 if not: 1

Enter best response file name: brillustration.txt

The input sequence is quite self-explanatory, except possibly for the input sequences 5 (number of coalitions), 11 and 12 (sequences of zeros and ones). At the fifth input point the program asks the user to provide the number coalitions for each type. This corresponds to the parameters $k_i, i = 1,2$ in section 2. In this example, we specify that there is 1 coalition making up type one group, and there are 3 coalitions making up the type two group. Input points 11 and 12 are where the user provides the structure of the coalitions. The numbers entered at these points correspond to u_i in section 2.4.5 of this paper. Input sequence 11 specifies that the first coalition consists of the two high types, and one low type player. Input sequence 12 specifies that the other three coalitions are simply the rest of the low type bidders competing individually. The format of the program therefore allows for the construction of a wide variety of hypothetical collusive environments. The program used in this paper, “bidfunc.exe” is available upon request from the first author.

Table 21: 3 bidders (median, low, high).

	Type			Auctioneer
	1	2	3	
k_i	1	1	1	
u_i	1	1	1	
a_i	2.0	1.0	3.39	
b_i	1.0	1.0	2.20	
mean	1.55	0.966	2.71	
std. dev.	1.25	0.911	1.15	
First price, no reserve				
E(revenue)	0.344	0.111	0.912	1.65
Prob 'win'	0.29	0.13	0.58	—-
First price, optimal reserve=2.016				
E(revenue)	0.225	0.061	0.622	1.851
Prob 'win'	0.22	0.08	0.51	0.18
Second price, no reserve				
E(revenue)	0.246	0.069	1.16	1.57
Prob 'win'	0.22	0.08	0.70	—-
Second price, optimal reserve=2.016				
E(revenue)	0.181	0.045	0.692	1.858
Prob 'win'	0.18	0.06	0.58	0.18
$F_i(v) = 1 - e^{-(\frac{v}{a_i})^{b_i}}$, truncated on $[0,5]$.				

Table 22: 2 bidders.

	Type		Auctioneer
	1	2	
hazard	increasing	decreasing	
k_i	1	1	
u_i	1	1	
a_i	1.11	1.50	
b_i	1.50	0.50	
mean	1.00	0.84	
std. dev.	0.67	1.01	
First price, no reserve			
E(revenue)	0.481	0.463	0.440
Prob 'win'	0.58	0.42	—-
First price, optimal reserve=0.98			
E(revenue)	0.211	0.297	0.656
Prob 'win'	0.33	0.28	0.39
Second price, no reserve			
E(revenue)	0.55	0.40	0.44
Prob 'win'	0.64	0.36	—-
Second price, optimal reserve=0.93			
E(revenue)	0.230	0.303	0.660
Prob 'win'	0.37	0.27	0.36
$F_i(v) = 1 - e^{-(\frac{v}{a_i})^{b_i}}$, truncated on $[0,4]$.			

Table 23: No collusion, 2 high types (H) and 4 Low types (L).

			First Price			Second Price		
	Mean	Std. dev.	Prob.	Rev.	Res.	Prob.	Rev.	Res.
H	3.756	1.030	0.393	0.385		0.415	0.413	
L	2.435	0.724	0.053	0.031		0.042	0.025	
Auc.			0.000	3.557		0.000	3.536	
H	3.756	1.030	0.394	0.386		0.415	0.411	
L	2.435	0.724	0.053	0.031		0.042	0.024	
Auc.			0.000	3.558	2.170	0.001	3.537	2.395

$v_i^H \sim LN(1.35, 0.35), v_i^L \sim LN(0.75, 0.35)$ truncated on $[1.5, 6]$.

Table 24: Collusion exercise between 2 high types (H) and 4 Low types (L).

			First Price			Second Price		
	Mean	Std. dev.	Prob.	Rev.	Res.	Prob.	Rev.	Res.
HH	4.346	0.880	0.668	0.906		0.832	1.227	
L	2.435	0.724	0.083	0.050		0.042	0.025	
Auc.			0.000	3.287		0.000	3.135	
HH	4.346	0.880	0.675	0.857		0.801	0.998	
L	2.435	0.724	0.073	0.044		0.038	0.021	
Auc.			0.026	3.297	2.972	0.048	3.237	3.134
HHL	4.379	0.856	0.706	1.019		0.874	1.398	
L	2.435	0.724	0.098	0.060		0.042	0.025	
Auc.			0.000	3.181		0.000	2.989	
HHL	4.379	0.856	0.709	0.902		0.815	0.977	
L	2.435	0.724	0.077	0.045		0.035	0.020	
Auc.			0.048	3.225	3.134	0.079	3.185	3.300

$v_i^H \sim LN(1.35, 0.35), v_i^L \sim LN(0.75, 0.35)$ truncated on $[1.5, 6]$.

Table 25: Study of the trade off between numerical accuracy and computational.

Order of Expansion, J=5			
Grid	Time (sec.)	RMSE 1	RMSE 2
500	37.1880	0.4000	0.0882
1000	95.3590	0.3988	0.0868
1500	146.1250	0.3984	0.0864
2000	202.2500	0.3982	0.0862
Number of grid points = 500			
J	Time (sec.)	RMSE 1	RMSE 2
2	18.1880	0.4000	0.0882
3	32.2810	0.4000	0.0882
4	38.7030	0.4000	0.0882
5	46.4690	0.4000	0.0882
Number of grid points = 2000			
J	Time (sec.)	RMSE 1	RMSE 2
2	93.4530	0.3982	0.0862
3	139.4530	0.3982	0.0862
4	173.8120	0.3982	0.0862
5	202.2500	0.3982	0.0862
$F_1(v) = 1 - e^{-v}, F_2(v) = 1 - e^{-(\frac{v}{3.39})^{2.2}}$, truncated on $[0,5]$.			

4.9 BIBLIOGRAPHY

M. Abramowitz and I.A. Segun. *Handbook of Mathematical Functions*. Dover Publications, New York, 1968.

V. Aguirregabiria and P. Mira. Swapping the nested fixed point algorithm: A class of estimators for distrete markov decision models. *Econometrica*, 70(4):1519–1543, Jul 2002.

Joseph G. Altonji and Rebecca M. Blank. *Race and Gender in the Labor Market*. Elsevier Science Publishers, 1999.

Sumru Altug and Robert A. Miller. Household choices in equilibrium. *Econometrica*, 58(3): 543–570, 1990.

Sumru Altug and Robert A. Miller. Effect of work experience of female wages and labour supply. *The Review of Economic Studies*, 65(1):45–85, 1998.

E. Anderson. Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society*, Series B, 32:283–301, 1970.

M. Arellano and B. Honoré. *Panel Data Models: More Recent Developments*. Elsevier Science Publishers B.V., 2001.

M. Bacolod and J. Hotz. Cohort changes in the transition from school to work: What changed and what consequences did it have for wages? *Unpublished Manuscript:* $http://www.econ.ucla.edu/hotz/working_{p}apers/cohort.pdf$, 2005.

Badi H. Baltagi. *Econometric Analysis of Panel Data*. Wiley, 2001.

Yoram Ben-Porath. The production of human capital and the life cycle of earnings. *The Journal of Political Economy*, 75(4):352–365, 1967.

Alan S. Blinder and Yoram Weiss. Human capital and labor supply: A synthesis. *The Journal of Policitical Economy*, 84(3):449–472, 1976.

Figure 9: Three Bidders

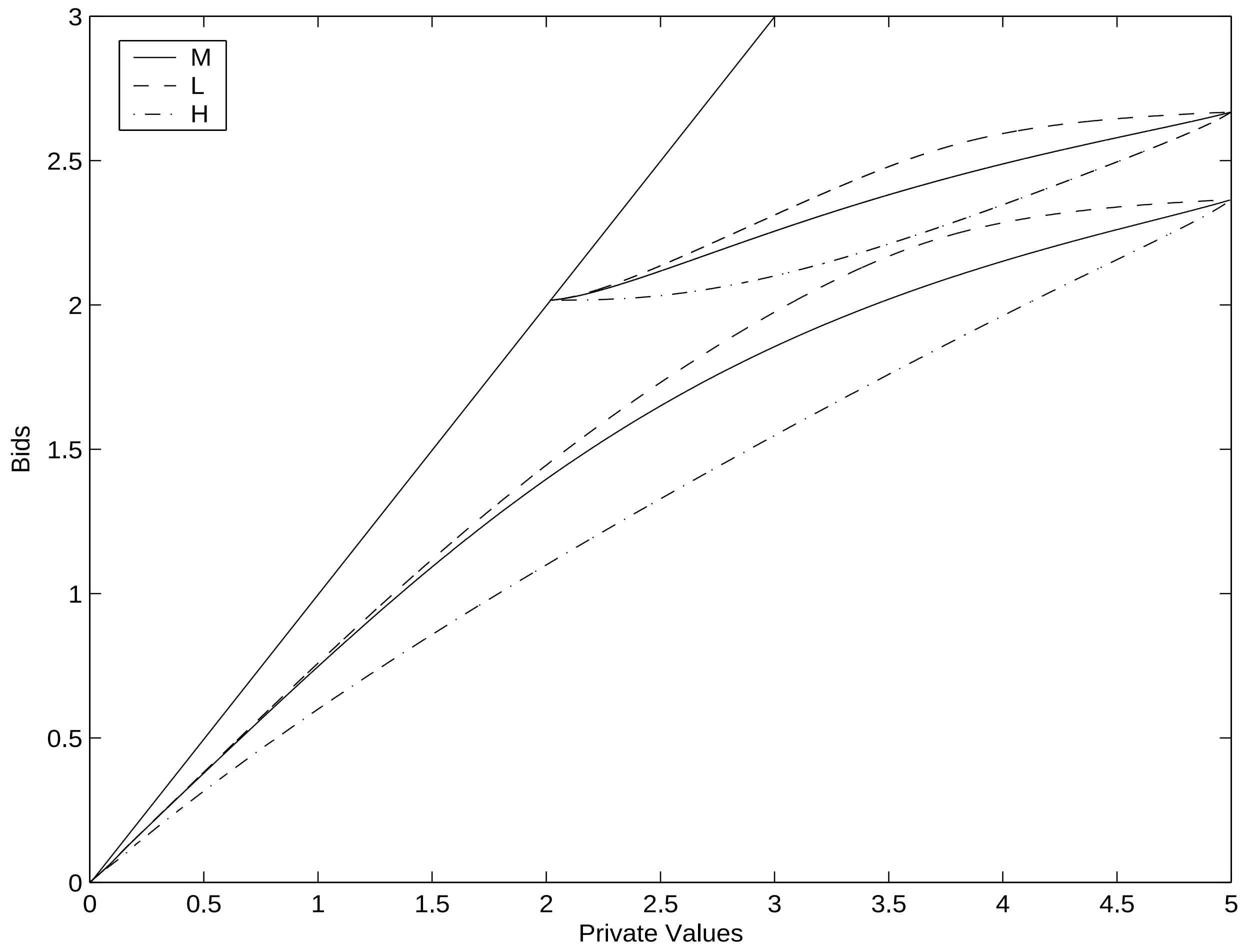


Figure 10: Two Bid Functions Crossing

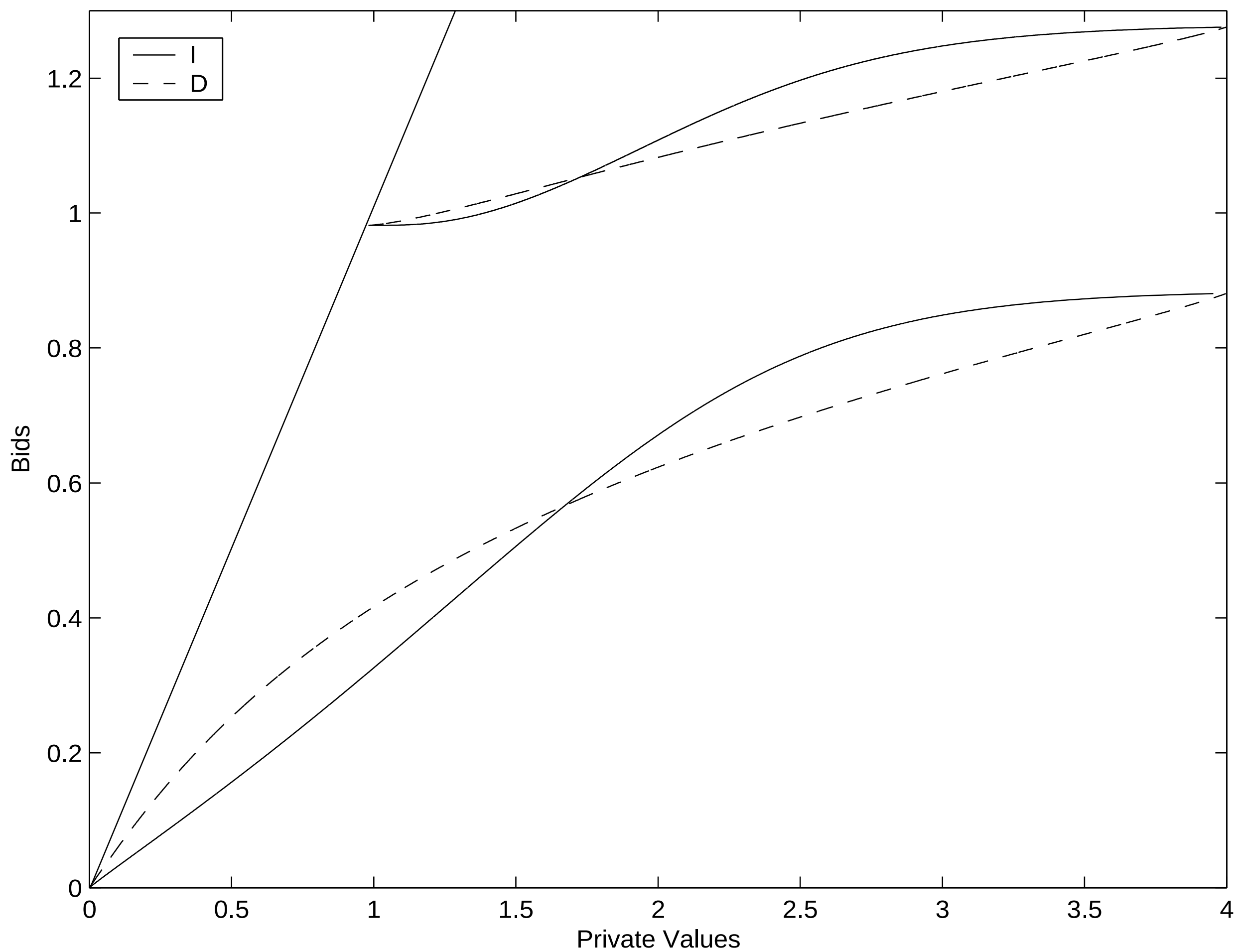


Figure 11: No collusion

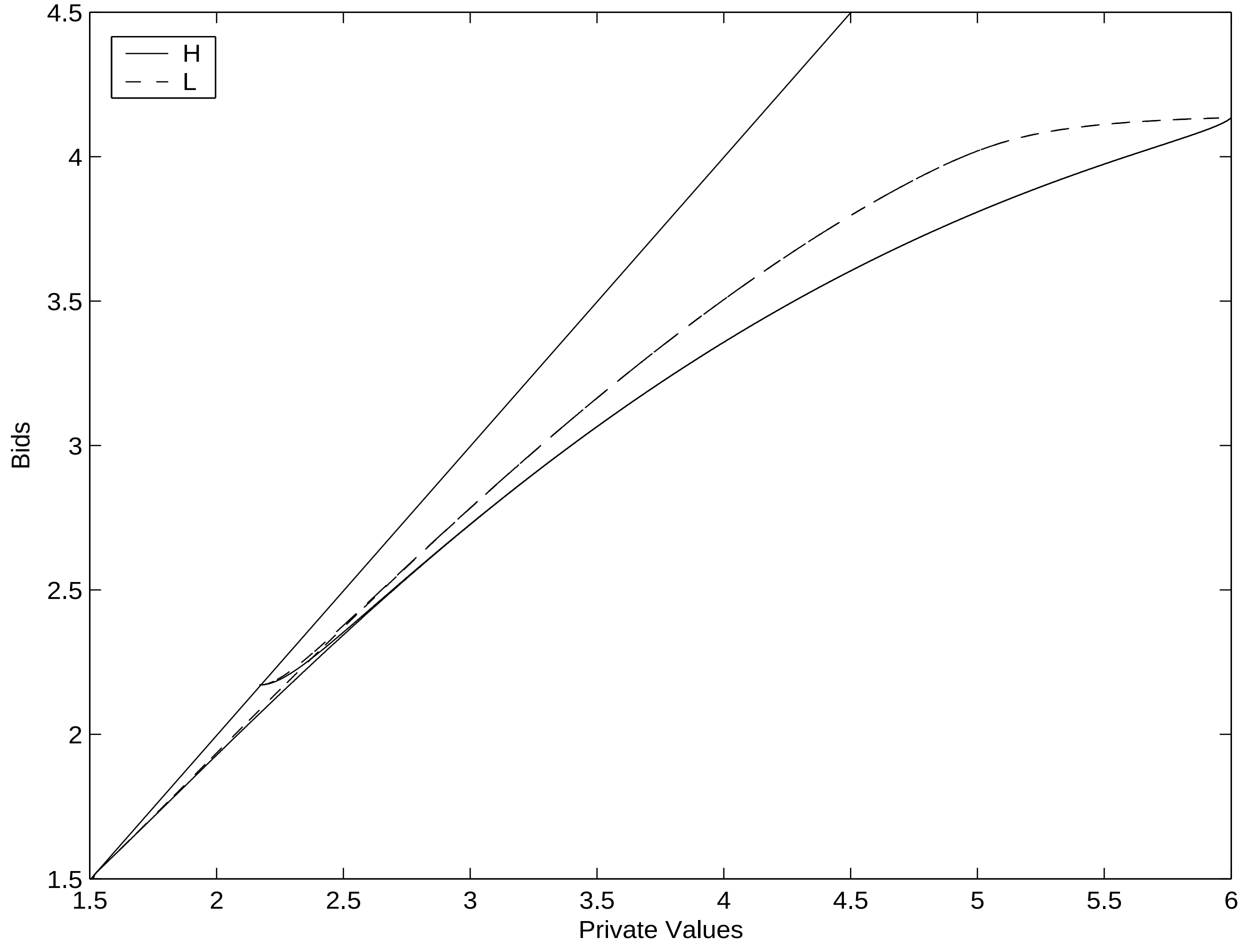


Figure 12: Two high types colluding

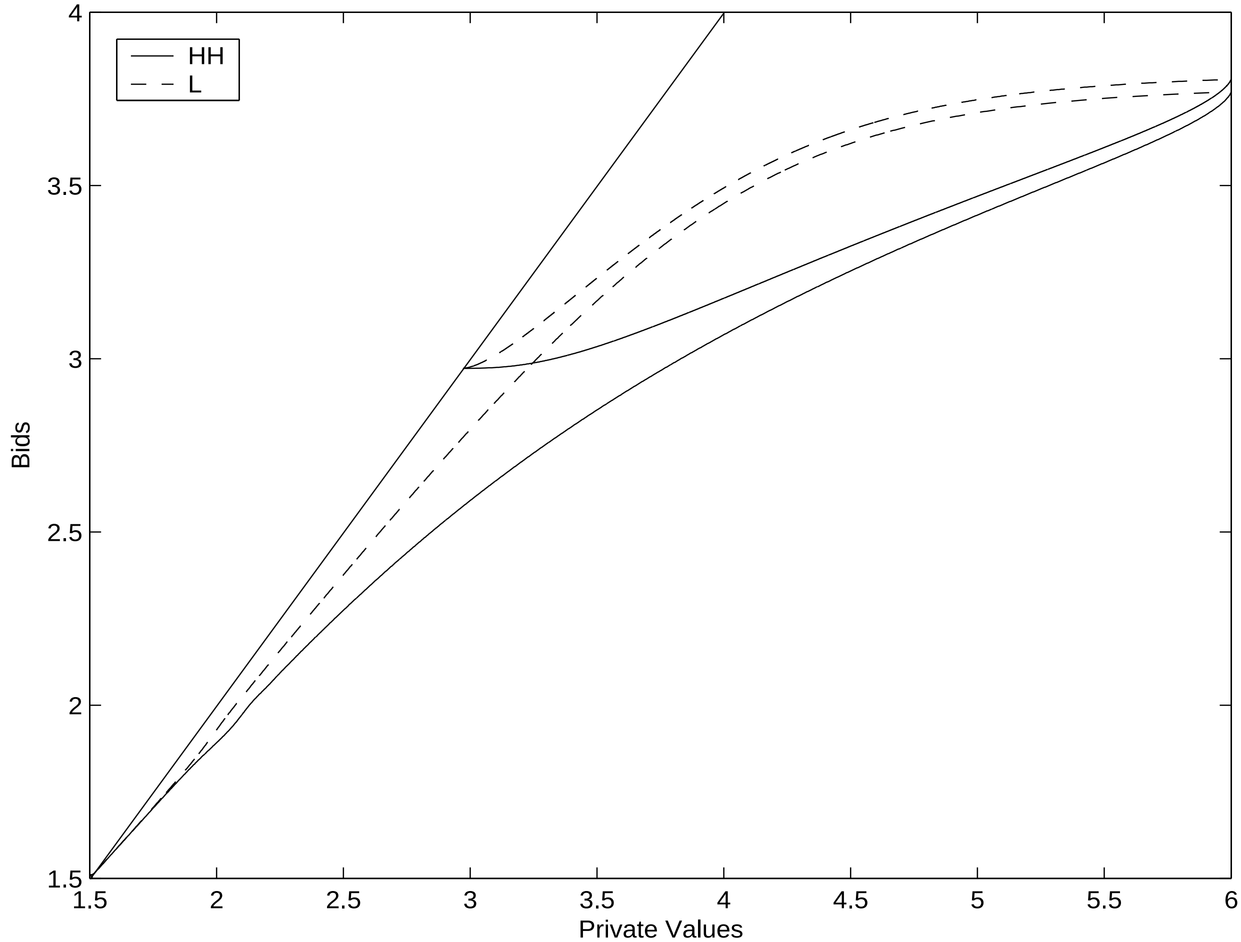


Figure 13: Two high types and one low type colluding

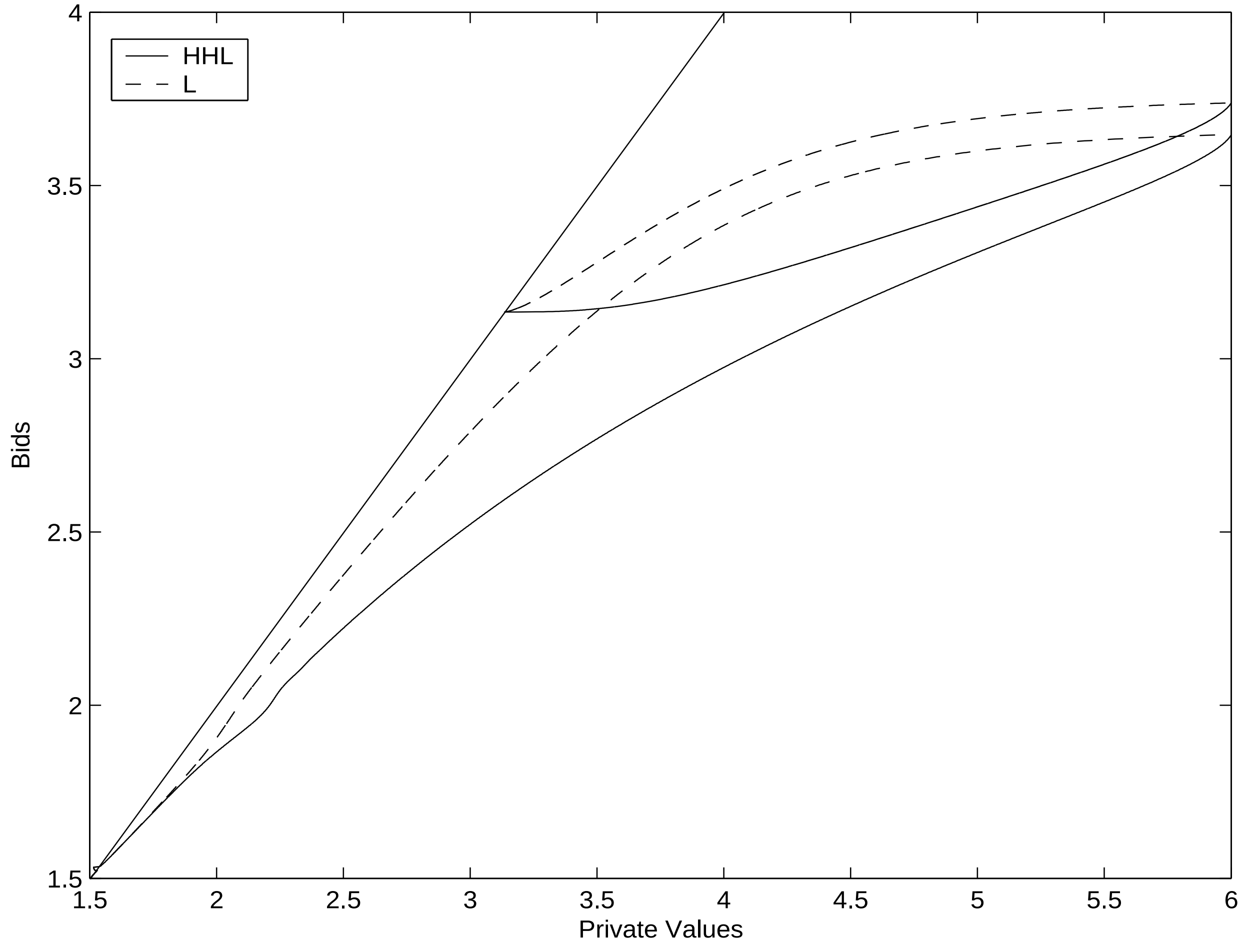
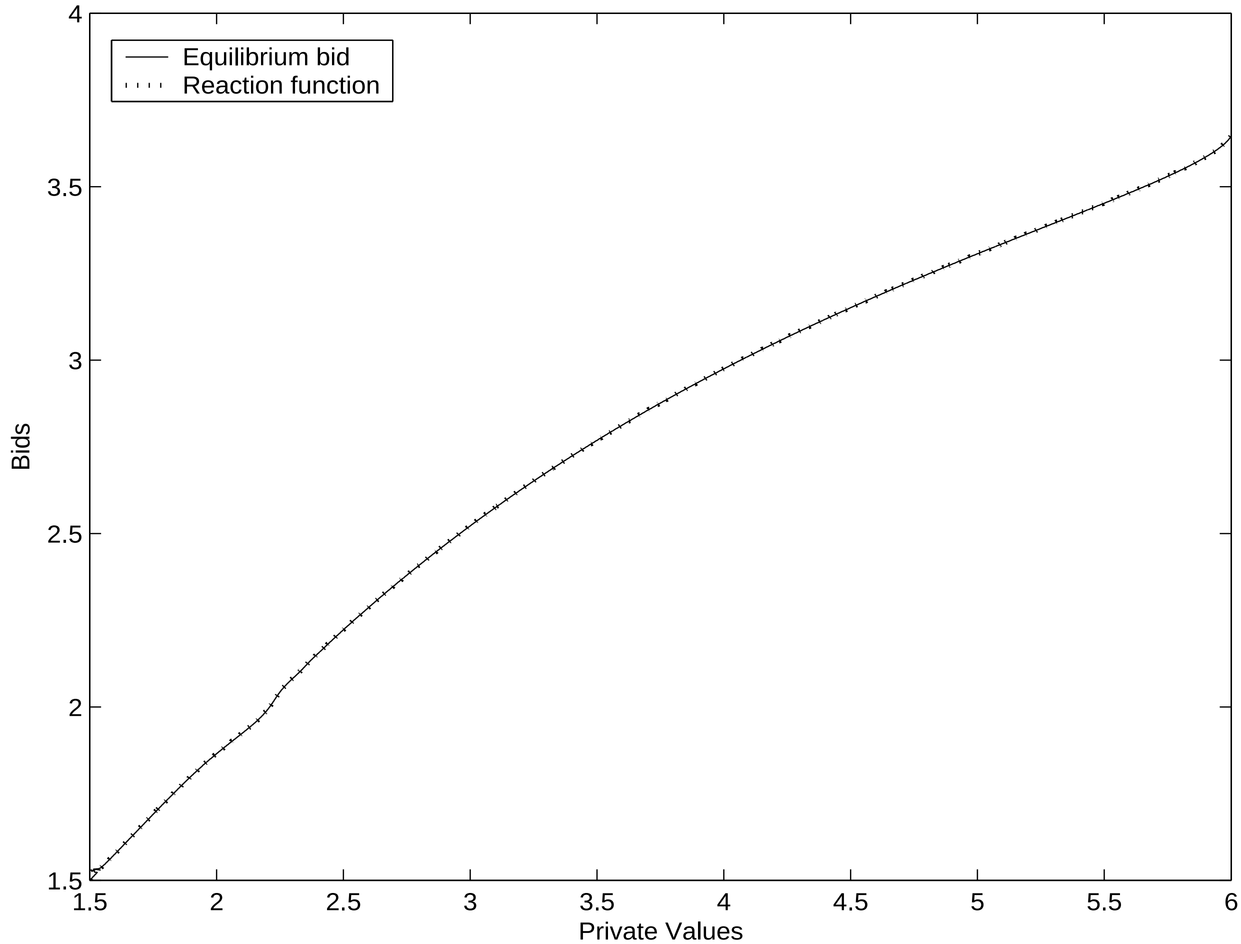


Figure 14: Comparison of equilibrium bid function and reaction function



A-W Bowman, M-C Jones, and I Gijbels. Testing monotonicity of regression. Technical report, 1996.

Levitan R. Briesmer, J.M. and M. Shubik. Towards a study of bidding processes, part four: Games with unknown costs. *Naval Research Quarterly*, 14:415–433, 1967.

A. Buja, T. Hastie, and R. Tibshirani. Linear smoothers and additive models. *The Annals of Statistics*, 17(2):453–555, 1989.

Stephen V. Cameron and James J. Heckman. Life cycle schooling and dynamic seclectino bias: Models and evidence for five cohorts of american males. *NBER Working Paper Series*, (6385), Jan 1998.

Stephen V. Cameron and James J. Heckman. The dynamics of educational attainment for blacks hispanics, and whites. *NBER Working Paper Series*, (7249), Jul 1999.

David Card. *The Causal Effect of Education on Earnings*. Elsevier Science Publishers, 1999.

G. Chamberlain. Analysis of covariance with qualitative data. *Review of Economics Studies*, XLVII:225–238, 1980.

G. Chamberlain. *Panel Data*. Elsevier Science Publishers B.V., 1984.

G. Chamberlain. *Heterogeneity, Omitted Variable Bias, and Duration Dependence*. Cambridge: Cambridge University Press, 1985.

Hung Chen. Asymptotically efficient estimation in semiparametric generalized linear models. *The Annals of Statistics*, 23(4):1102–1129, 1995.

S. Chen. Root-n consistent estimation of a panel data sample selection model. *MIMEO: The Hong Kong University of Science and Technology*, 1998.

Ronald D'Amico. Does employment during high school impair academic progress? *Sociology of Education*, 57(3):152–164, Jul 1984.

C. de Boor. *A Practical Guide to Splines*. Springer-Verlag, New York, 1978.

S. Donald and H. Paarsch. Identification, estimation and testing in parametric empirical models of auctions within the independent private value paradigm. *Econometric Theory*, 12:517–567, 1996.

Z. Eckstein and K. Wolpin. Dynamic labor force participation of married women and endogenous work experience. *The Review of Economic Studies*, 56(3):375–590, 1989.

Zvi Eckstein and Kenneth I. Wolpin. Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica*, 67(6):1295–1339, Nov 1999.

Martin S. Eichenbaum, Hansen Lars Peter, and Kenneth J. Singleton. A time series analysis of representative agent models of consumption and leisure choice under uncertainty. *The Quarterly Journal of Economics*, 103(1):51–78, 1988.

J.P. Florens, C. Protopopescu, and J.F. Richard. Identification and estimation of a class of game theoretic models. *MIMEO, University of Pittsburgh*, 2004.

George-Levi Gayle and Robert A. Miller. Life-cycle fertility and human capital accumulation. *Unpublished Manuscript, Carnegie Mellon University*, 2003.

I. Gibjels. Monotone regression. *Discussion Paper 0334, Université Catholique de Louvian*, 2003. URL http://www.stat.ucl.ac.be.

Jinyong Hahn, Jerry Houseman, and Guido Kuersteiner. Bias corrected instrumental variables estimation for dynamic panel models with fixed effects. *Unpublished Manuscript, http://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=145354*, 2001.

L.P. Hansen, J. Heaton, and A. Yaron. Finite-sample properties of some alternative gmm estimators. *Journal of Business and Economics Statistics*, 14(3):262–280, Jul 1996.

James J. Heckman. A life-cycle model of earnings, learning, and consumption. *The Journal of Political Economy*, 84(4):S11–S44, Aug 1976.

B.E. Honoré. Trimmed lad and least squrares estimation of truncated censored regression models with fixed effects. *Econometrica*, 60:533–565, 1992.

B.E Honoré and E. Kyriazidou. Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68(4):839–874, 2000.

J. Horowitz. A smoothed maximum score estimator for the binary response model. *Econometrica*, 60(3):505–531, 1992.

Joseph V. Hotz and Robert A. Miller. Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies*, 60:497–529, 1993.

Joseph V. Hotz, Finn E. Kydland, and Guilherme Sedlacek. Intertemporal preferences and labor supply. *Econometrica*, 56(2):355–360, 1988.

H. Ichimura. Semiparametric least squares (sls) and weighted sls estimation of single index models. *Journal of Econometrics*, 58:71–120, 1993.

Susumu Imai. Intertemporal labor supply and human capital accumulation. *Unpublished Manuscript, Pennsylvania State University*, 2000.

Michael P. Keane. Financial aid, borrowing constraints and college attendance: Evidence from structural estimates. *American Economic Review (Papers and Proceedings)*, 92(2):293–297, 2002.

Michael P Keane and Kenneth I. Wolpin. The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review*, 42:1051–1103, Nov, 2001.

Michael P Keane and Kenneth I. Wolpin. The career decisions of young men. *The Journal of Political Economy*, 105(3):473–522, Jun 1997.

Michael P. Keane and Kenneth I. Wolpin. The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte carlo evidence. *The Review of Economics and Statistics*, 76(4):648–672, Nov 1994.

R. Klein and R. Spady. An efficient semiparamtric estimator for binary response models. *Econometrica*, 61(2):387–421, 1993.

J.J. Laffont, H. Ossard, and Q. Vuong. Econometrics of first-price auctions. *Econometrica*, 63: 953–980, 1995.

B. Lebrun. Uniqueness of the equilibrium in first-price auctions. *Games and Economic Behavior, forthcoming*, 2005.

B. Lebrun. Existence of an equilibrium in first price auctions. *Economic Theory*, 7:421–443, 1996.

B. Lebrun. First price auctions in the asymmetric *n* bidder case. *International Economic Review*, 40:125–142, 1999.

T. Magnac. Panel binary variables and sufficiency: Generalizing conditional logit. *Econometrica*, 72(6):1859–1876, 2004.

T. Magnac. State dependence and heterogeneity in youth employment histories. *Working Paper, INRA and CREST, Paris*, 1997.

E. Mammen, J.S. Marron, B.A. Turlach, and M.P. Wald. A general projection framework for constrained smoothing. *Statistical Science*, 16(3):232–248, 2001.

C. Manski. Semiparametric analysis of random effects linear models from binary panel data. *Econometrica*, 55(2):357–362, 1987.

R.C. Marshall and L.M. Marx. Bidder collusion. *MIMEO, Duke University*, 2005.

R.C. Marshall and S.P. Schulenberg. Numerical analysis of asymmetric acutions with optimal reserve prices. *MIMEO, Duke University*, 1998.

R.C. Marshall, M.J. Meurer, J.F. Richard, and W. Stromquist. Numerical analysis of asymmetric first price auctions. *Games and Economic Behavior*, 7:193–220, 1994.

R.C. Marshall, M.E. Raiff, J.F. Richard, and S.P. Schulenberg. The impact of delivery synergies on bidding in the georgia school milk market. *Forthcoming in the B.E. Journals in Economic Analysis and Policy*, 2005.

E.S. Maskin and J.G. Riley. Asymmetric auctions. *Review of Economic Studies*, 67:413–438, 2000a.

E.S. Maskin and J.G. Riley. Equilibrium in sealed high bid auctions. *Review of Economic Studies*, 67:439–454, 2000b.

E.S. Maskin and J.G. Riley. Optimal auctions with risk averse buyers. *Econometrica*, 52:1473–1518, 1984.

S. Mathews. Selling to risk averse buyers with unobservable tastes. *Journal of Economic Theory*, 30:370–400, 1983.

P.R. Milgrom and R.J. Weber. A theory of acutions and competitive bidding. *Econometrica*, 50: 1089–1122, 1982.

Robert A. Miller and S. Sanders. Human capital development and welfare participation. *Carnegie-Rochester Conference Series on Public Policy*, 46:237–253, 1997.

Derek A. Neal and William R. Johnson. The role of premarket factors i black-white wage differences. *Journal of Political Economics*, 105(5):869–895, 1996.

W. Newey. The asymptotic variance of semiparametric estimators. *Econometrica*, 62(6):1349–1382, 1994.

Whitney K. Newey and Daniel McFadden. *Large Sample Estimation and Hypothesis Testing*. Elsevier Science Publishers, 1994.

Gerald S. Oettinger. Does high school employment affect high school academic performance? *Industrial and Labor Relations Review*, 53(1):136–151, Oct 1999.

W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes*. The Art of Scientific Computing, Cambridge University Press, Cambridge, 1986.

W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in Fortran 90:*

The Art of Parrallel Scientific Computing, volume 2. New York: Cambridge University Press, 1996.

G. Rasch. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Denmark Pædagogiske Institut., 1960.

R.B. Riley and W.F. Samuelson. Optimal auctions. *American Economic Review*, 71:381–392, 1981.

Christopher J. Ruhm. Is high school employment consumption of investment? *Journal of Labor Economics*, 15(4):735–776, Oct 1997.

Kathryn L. Shaw. Life-cycle labor supply with human capital accumulation. *International Economic Review*, 30(2):431–456, 1989.

S. van de Geer. *Empirical Processes in M-Estimation*. Cambridge University Press, 2000.

A.W. van der Vaart. Semiparametric statistics. *Lecture notes*, 1998. URL http://www.math.vu.nl/publications/stochastics/2000/WS2000-1.ps.

A.W. van der Vaart, Arusharka Sen, and Subhashis Ghosal. Testing monotonicity of regression, December 22 1998. URL http://www.cs.vu.nl/~ghosal/ps/monotonicity.ps.

Robert J. Willis. *Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions*. Elsevier Science Publishers, 1986.

A.1 APPENDIX TO CHAPTER 2

A.1.1 Data and Sample Construction

The data is taken from the 1979 youth cohort of the National Longitudinal Survey of Labor Market Experience (NLSY79), a comprehensive panel data set that follows individuals over the period 1979 to 2000, who were 14 to 21 years of age as of January 1, 1979. The data set initially consisted of 12,686 individuals: a representative sample of 6,111 individuals, a supplemental sample of 5,295 Hispanics, non-Hispanic blacks, and economically disadvantaged, non-black, non-Hispanics, and a supplemental sample of 1,280 military youth. Interviews were conducted on an annual basis though 1994, after which they adopted a biennial interview schedule. This study makes use of the first 16 years of interviews, from 1979 to 1994. By 1990, the NLSY79 experienced attrition of 2,250 sample members, of which 1,097 were from the supplemental sample of military youth. I discuss briefly the construction of some of the key variables used in estimation

Employment

The NLSY79 collects detailed work history data for individuals in the sample. The work history data includes beginning and ending dates for all of 5 possible jobs, a maximum of 5 possible gaps in employment with each of the 5 possible jobs, the usual hours worked per day or per week on each job, and the hourly rate of pay on each job. The biggest complication in calculating hours worked is the fact that it must be calculated for the relevant year, which is the school year in this case. Since the actual weeks that comprise the school year vary from state to state, the dates chosen for the school year are somewhat arbitrary. Following Eckstein and Wolpin [1999], the year for those not attending school starts at October 1^{st} in year t and ends September 30^{st} of year t+1. For those attending school the school year instead ends at June 30 of year t+1. Weeks employed is then calculated based on these calendar dates. Hours worked per week or per day and hourly rate of pay is reported retrospectively back to the previous interview date. These variables were also adjusted to the above specified calendar dates. From these, we then construct hours worked for the relevant years, as well as average hourly rate of pay and an employment rate variable, which is the fraction of the relevant year in which the respondent was actively employed.

Education

The NLSY79 also collects information on the respondents' education. In particular, the NLSY79 collects , among others, enrollment status, highest grade level completed, current grade level, and degree held. The primary variables used in the paper are highest grade completed and enrollment status. In 1981, the NLSY collected information on the patterns of school activities of the respondents that are enrolled in school. In particular, the NLSY asked these respondent about the amount of hours they spent in school during the week before the interview date. They asked whether or not the time the reported is typical or not, and if no, to report the typical hours spent in school. The NLSY also asked the respondents to report the number of hours they spent studying outside of school during the week before the interview date. The response to these questions are used in the paper to estimate the study pattern of individuals enrolled in school.

There are a number of missing observations on highest grade completed. Many of these missing observations could be recovered from the information provided by enrollment status and highest grade completed in other years by the respondent. Since the model relies very much on the data on highest grade completed, we decide not to impute those years that are not recoverable with very high confidence.

The model construction and estimation requires data on the cost of schooling for an individual who decides to enroll in school. The yearly in-state tuition and required fees for four-year institutions and two-year institutions are taken from the NCES web site. Also, to identify the the aggregate shocks in wages and consumption, all nominal variables have to normalized to the same base year. To do this, the CPI is taken from the BLS web site, and converted to have a base year of 1981.

Asset holdings

Beginning in 1985, the NLSY79 began collecting comprehensive information on the asset holdings of the respondents. This information was collected annually up to and including 1994, except for the year 1991 where asset data is missing. The best way to deal with these missing observations on asset holdings depends on exactly how the data will be used in estimation. In the case of Keane and Wolpin [2001] and Imai [2000], asset holding itself plays a central role in their model. Their method of imputation was therefore to model and asset holdings as normally distributed, and the estimate the mean and variance, from which they impute the missing years. In my case however, I require savings balance to impute total family consumption. For years in which the data is available, this is simply the difference between the Asset holding from one year to the next. For the years in which the data is missing, I take savings balance to be zero. For the early years of the cohort, net savings is relatively small and centered around zero. This suggests that the bias induced by this imputation is small. Furthermore, in estimating the consumption equation, savings is one the right hand side of the equation. The consistency of parameter estimates in the case where the left hand side variable is measured with a mean zero error is well documented in classical econometric textbooks. Finally, if there were large biases introduced by this imputation, they would show up in the estimated aggregate prices, These is no unusual visible discrete change in estimated aggregate prices for these periods. All these reasons lead me to believe that such imputations results in minimal biases in the parameters of interest.

Consumption

The NLSY79 does not collect data on individual consumption. However, the unique advantage of this data set that it collects detailed information on individual asset holding. To estimate the parameters in the above equation, family consumption is imputed from family income, family savings, four year schooling costs, and two year schooling costs. The way this is done is a follows. Subtracting family savings is taken from family income gives an estimate of the total resources available to the family in that year, net of savings. If the individual goes to high school, then his cost of schooling is assumed to be 0. If he goes to a two-year college, his cost of schooling is the two-year tuition cost, and if he goes to a four-year college, his cost of schooling is the four-year tuition cost. The individual's cost of schooling is subtracted from his individual resources. The yearly averages of the imputed consumption is given in Table 2.

Demographics

Demographic and family background variables collected by the NLSY79 and used in this study include age, race, mother's education, Father's education, family income, and year of experience working. Experience is calculated from the employment history section of the data set, which gives complete employment status for each year. Missing observations in family income are imputed by first using a three year moving average smoothing technique, followed by regressing family income on other covariates, some of which not listed here, and using the predicted income for the cases in which family income is missing. The resulting distribution of imputed family income match the distribution of actual (observed) family remarkably well.

Sample Restriction

As stated above, the data employed in this paper span the years of 1979 though 1994. The model specified in section (3.2) does not include the decision to enter the military, and thus as the first restriction on the data we drop all males who enter the military in 1979. This restriction reduces the sample size to 11406. As stated above, we drop respondents for cases where missing observations in highest grade completed cannot be recovered with very high confidence. This reduces the sample to 7814 respondents. This is clearly are somewhat severe restriction on the data, and it may pay to invest is less restrictive imputation rules. This however is not pursued here. In the literature, female members are treated differently from male sample members. The choice set of a female is generally considered larger than that of a male. The additional decisions usually included in the choice set for women are marriage decisions and fertility decisions. To avoid these additional complications, the data is restricted to include males only. This results in a sample size of 3916 male respondents. The summary statistics and all estimations make use of this sample.

A.1.2 Standard Errors for the Probability of Grade Promotion

Let y_{nt} be in indicator variable equal to 1 if the individual advances a grade level, and 0 otherwise. Define:

$$g(x_4, B_4, B_3) \equiv x_4 \left(y - \frac{e^{x_4' B_4}}{1 + e^{x_4' B_4}} \right) \tag{A.1.1}$$

$$h(x_3, B_3) \equiv x_3(\ln(s) - x_3' B_3) \tag{A.1.2}$$

$$f(x, \theta) \equiv [g(x_4, B_3, B_4)', h(x_3, B_3)']' \tag{A.1.3}$$

where $\theta \equiv (B_4', B_3')'$. Equation (A.1.1) is the score contribution of a single individual from the likelihood function constructed from equation (2.8.2). Equation (A.1.2) is the moment condition derived from the study time equation (2.8.1). I assume that these two moments are uncorrelated, and we have by construction that $\frac{1}{N} \sum_n f(x, \hat{\theta}) = 0$. The proof that $\hat{\theta} \xrightarrow{p} \theta_0$ is straightforward and therefore omitted. Let

$$
\begin{aligned}
G_4 &\equiv E[\Delta_{B_4} g(x_4, B_4, B_3)] = -E\left[x_4 x_4' \frac{e^{x_4' B_4}}{1+e^{x_4' B_4}} \frac{1}{1+e^{x_4' B_4}}\right] && \text{(A.1.4)} \\
G_3 &\equiv E[\Delta_{B_3} g(x_4, B_4, B_3)] = -E\left[x_4 x_3' (sB_{4,1} + 2s^2 B_{4,2}) \frac{e^{x_4' B_4}}{1+e^{x_4' B_4}} \frac{1}{1+e^{x_4' B_4}}\right] && \text{(A.1.5)} \\
H_3 &\equiv E[\Delta_{B_3} h(x_3, B_3)] = -E[x_3 x_3']. && \text{(A.1.6)}
\end{aligned}
$$

Since $f(x,\theta)$ satisfies conditions $(i)-(v)$ of Theorem 3.4 of Newey and McFadden [1994], $\hat{B}_4$ is asymptotically normal and $\sqrt{n}(\hat{B}_4 - B_4) \xrightarrow{d} N(0,V)$, where

$$
V = G_4^{-1} E[g(x_4) g(x_4)'] G_4^{-1\prime} + G_4^{-1} G_3 H_3^{-1} E[h(x_3) h(x_3)'] H_3^{-1\prime} G_3' G_4^{-1\prime} \tag{A.1.7}
$$

Thus the variance can be consistently estimated by replacing the jacobian terms in the equation (A.1.7) with their sample averages.

A.1.3 The estimation method for the CCP's and the conditional state probabilities

Let $K[\delta_N^{-1}(\Psi_{mr}^N - \Psi_{nt}^N)]$ be a kernel, where δ_N is an appropriately chosen bandwidth. Then the nonparametric estimate of p_{ntj} is computed using the kernel estimator

$$
p_{ntj}^N \equiv \frac{\sum_{m=1}^N \sum_{r=1}^T d_{mrj} K[\delta_N^{-1}(\Psi_{mr}^N - \Psi_{nt}^N)]}{\sum_{m=1}^N \sum_{r=1}^T K[\delta_N^{-1}(\Psi_{mr}^N - \Psi_{nt}^N)]}. \tag{A.1.8}
$$

To define the conditional state probabilities we first define the set of possible histories that will become relevant in the model. Accordingly, the $(2\rho+K+1)$-dimensional vectors

$$\begin{aligned}
x_{nt0}^{(i)} \equiv\ & (h_{nt-\rho+i},\cdots,h_{nt-1},0,\cdots,0,s_{nt-\rho+i},\cdots,s_{nt-1},0,\cdots,0,\\
& S_{nt-\rho+i+1},\cdots,S_{nt},S_{nt},\cdots,S_{nt},E_{nt-\rho+i},z_{nt+i}),\\
x_{nt1}^{(i)} \equiv\ & (h_{nt-\rho+i},\cdots,h_{nt-1},h_{nt}^{*},\cdots,0,s_{nt-\rho+i},\cdots,s_{nt-1},0,\cdots,0,\\
& S_{nt-\rho+i+1},\cdots,S_{nt},S_{nt},\cdots,S_{nt},E_{nt-\rho+i},z_{nt+i}),\\
x_{nt2}^{(i)} \equiv\ & (h_{nt-\rho+i},\cdots,h_{nt-1},0,\cdots,0,s_{nt-\rho+i},\cdots,s_{nt-1},s_{nt}^{*},\cdots,0,\\
& S_{nt-\rho+s+1},\cdots,S_{nt},S_{nt}+1,\cdots,S_{nt}+1,E_{nt-\rho+i},z_{nt+s}),\\
x_{nt3}^{(i)} \equiv\ & (h_{nt-\rho+i},\cdots,h_{nt-1},0,\cdots,0,s_{nt-\rho+i},\cdots,s_{nt-1},s_{nt}^{*},\cdots,0,\\
& S_{nt-\rho+i+1},\cdots,S_{nt},S_{nt},\cdots,S_{nt},E_{nt-\rho+i},z_{nt+i}),\\
x_{nt4}^{(i)} \equiv\ & (h_{nt-\rho+i},\cdots,h_{nt-1},h_{nt}^{*},\cdots,0,s_{nt-\rho+i},\cdots,s_{nt-1},s_{nt}^{*},\cdots,0,\\
& S_{nt-\rho+i+1},\cdots,S_{nt},S_{nt}+1,\cdots,S_{nt}+1,E_{nt-\rho+i},z_{nt+i}),\\
x_{nt5}^{(i)} \equiv\ & (h_{nt-\rho+i},\cdots,h_{nt-1},h_{nt}^{*},\cdots,0,s_{nt-\rho+i},\cdots,s_{nt-1},s_{nt}^{*},\cdots,0,\\
& S_{nt-\rho+s+1},\cdots,S_{nt},S_{nt},\cdots,S_{nt},E_{nt-\rho+i},z_{nt+s}),
\end{aligned} \tag{A.1.9}$$

for $i=1,\cdots,\rho$, where h_{nt}^{*} and s_{nt}^{*} is the fraction of time individual n devotes to working and schooling conditional on participating and enrolling. Define the state vectors $\Psi_{ntk}^{(i)} \equiv (x_{ntk}^{(i)},\mu_n\eta_n\omega_{nt+i}\lambda_{t+i})$, $k=0,\cdots,5$, where $\omega_{nt} \equiv \omega_{t1}^{e_{nt1}}\omega_{t2}^{e_{nt2}}$. For example, $\Psi_{nt1}^{(i)}$ is the state of a young man who has accumulated the history

$$(h_{nt-\rho},\cdots,h_{nt-1},s_{nt-\rho},\cdots,s_{nt-1},S_{nt-\rho+1},\cdots,S_{nt},E_{nt-\rho+1})$$

up to period t, chooses not to enroll in school and to work h_{nt}^{*} hours in period t, and not to enroll nor work for $i-1$ periods following t. Similarly, $\Psi_{nt3}^{(i)}$ is the state of a young man who has accumulated the same history up to period t, chooses not to work, to and study s_{nt}^{*} hours in period t, gets promoted a grad at the end of year t, and chooses not to enroll nor work for $i-1$ periods following t.

Define $p_j(\Psi_{ntk}^{(i)})$, $j=0,\cdots,3$, $k=0,\cdots,5$, as the the probability that individual n chooses alternative j in period $t+i$ conditioned on realizing the state vector Ψ_{ntk}^{i} in period $t+i$. The intuition for estimating these future state probabilities is to condition on observationally equivalent men in the current period. To do this, define the indicator variables:

$$d_{ntj}^{(i)} \equiv \begin{cases} d_{nt-i,j} \prod_{r=1}^{i-1} d_{nt-r,0}, & \text{for } j = 0,1, \\ y_{nt-i} d_{nt-i,j} \prod_{r=1}^{i-1} d_{nt-r,0}, & \text{for } j = 2,4, \\ (1-y_{nt-i}) d_{nt-i,j} \prod_{r=1}^{i-1} d_{nt-r,0}, & \text{for } j = 3,5, \end{cases} \tag{A.1.10}$$

where y_{nt} is equal to one if the individual is promoted a grade level at the end of period t, and zero otherwise. Therefore, $d_{ntj}^{(i)}$ allows us to condition of the appropriate history for computing the estimators of the state probabilities $p_k(\Psi_{ntj}^{(i)})$, which are computed as

$$p_k^N(\Psi_{ntj}^{(i)}) \equiv \frac{\sum_{m=1}^{N} \sum_{r=1}^{T} d_{mrk} d_{mrj}^{(i)} K[\delta_N^{-1}(\Psi_{mr}^N - \Psi_{nt}^N)]}{\sum_{m=1}^{N} \sum_{r=1}^{T} d_{mrj}^{(i)} K[\delta_N^{-1}(\Psi_{mr}^N - \Psi_{nt}^N)]}. \tag{A.1.11}$$

Estimation of the parameters characterizing preference also require that the derivatives of the probabilities with respect to h be estimated. The methodology employed to estimate these quantities is found in Altug and Miller [1998].

A.1.4 Derivation of the moment conditions for the final stage estimation

Hotz and Miller [1993] prove the existence of a mapping $q : [0,1] \rightarrow \Re$ such that

$$q(p_k(\Psi_{nt})) \quad = \quad V_j(\Psi_{nt}) - V_k(\Psi_{nt}), \tag{A.1.12}$$

Equations (A.1.12) and (2.3.14) are used to derive the alternative representation of the conditional valuation function V_{ntk} for the finite dependence case. To do so, define

$$u_j(\Psi_{nt}) \equiv \begin{cases} u_1(S_{nt},0) + u_2(x_{nt},0) + u_3(x_{nt},1) + \alpha^{-1}\eta_n\lambda_t c_{nt} & for\, j = 0, \\ u_1(S_{nt},0) + u_2(x_{nt},1) + u_3(x_{nt},1-h_{nt}^*) + \alpha^{-1}\eta_n\lambda_t c_{nt} & for\, j = 1, \\ u_1(S_{nt},1) + u_2(x_{nt},0) + u_3(x_{nt},1-s_{nt}) + \alpha^{-1}\eta_n\lambda_t c_{nt} & for\, j = 2, \\ u_1(S_{nt},1) + u_2(x_{nt},1) + u_3(x_{nt},1-h_{nt}^*-s_{nt}) + \alpha^{-1}\eta_n\lambda_t c_{nt} & for\, j = 3. \end{cases} \tag{A.1.13}$$

Recall that $F_j(\Psi_{nt}^{(i)}|\Psi_{nt})$ is the probability that the state vector of individual n in period $t+i$ is $\Psi_{nt}^{(i)}$, given that his state vector in period t is Ψ_{nt} and he chooses alternative j in period t. Then by recursive application of the law of iterated expectations, the conditional valuation function can be expressed as

$$\begin{aligned} V_j(\Psi_{nt}) &= u_j(\Psi_{nt}) + E_t\Big\{ \textstyle\sum_{i=1}^{\rho} \Big[\beta^i \sum_{\mathcal{A}_{ntj}^{(i)}} \Big[u_0(\Psi_{nt}^{(i)}) + \varphi_0(p_0(\Psi_{nt}^{(i)})) \\ &+ \textstyle\sum_{k=1}^{3} p_k(\Psi_{nt}^{(i)})(q(p_k(\Psi_{nt}^{(i)})) + \varphi_k(p_k(\Psi_{nt}^{(i)})) \\ &- \varphi_0(p_0(\Psi_{nt}^{(i)})))\Big] F_j(\Psi_{nt}^{(i)}|\Psi_{nt}) \\ &+ \beta^{\rho+1} \textstyle\sum_{\mathcal{A}_{ntj}^{(\rho+1)}} \Big[V_0(\Psi_{nt}^{(\rho+1)}) + \varphi_0(p_0(\Psi_{nt}^{(\rho+1)})) \\ &+ \textstyle\sum_{k=1}^{3} p_k(\Psi_{nt}^{(\rho+1)})(q(p_k(\Psi_{nt}^{(\rho+1)})) + \varphi_k(p_k(\Psi_{nt}^{(\rho+1)})) \\ &- \varphi_0(p_0(\Psi_{nt}^{(\rho+1)})))\Big] F_j(\Psi_{nt}^{(\rho+1)}|\Psi_{nt})\Big]\Big\}, \end{aligned} \tag{A.1.14}$$

Notice that the recursive substitution employed to obtain the alternative representation is only valid up to where $p_0(\Psi_{ntj}^{i}) > 0$. In the context of this paper, this condition is true at $i = 2$ for $j = 0, 1$, and $i = 1$ for $j = 2, \cdots, 5$. Equation (A.1.14) gives the following alternative representation of the Euler equations for labor supply and schooling

$$\begin{aligned} 0 = & \frac{\partial u_j(\Psi_{nt})}{\partial g_{nt}} + E_t\Big\{ \textstyle\sum_{i=1}^{\rho} \Big[\sum_{\mathcal{A}_{ntj}^{(i)}} \Big[\frac{\partial[u_0(\Psi_{nt}^{(i)}) + \varphi_0(p_0(\Psi_{nt}^{(i)}))]}{\partial g_{nt}} \\ &+ \textstyle\sum_{k=1}^{3} p_k(\Psi_{nt}^{(i)}) \frac{\partial[(q(p_k(\Psi_{nt}^{(i)})) + \varphi_k(p_k(\Psi_{nt}^{(i)})) - \varphi_0(p_0(\Psi_{nt}^{(i)})))]}{\partial g_{nt}} \\ &+ \textstyle\sum_{k=1}^{3} [(q(p_k(\Psi_{nt}^{(i)})) + \varphi_k(p_k(\Psi_{nt}^{(i)})) \\ &- \varphi_0(p_0(\Psi_{nt}^{(i)})))] \frac{p_k(\Psi_{nt}^{(i)})}{\partial g_{nt}} \Big] F_j(\Psi_{nt}^{(i)}|\Psi_{nt}) \Big] \\ &+ \textstyle\sum_{\mathcal{A}_{ntj}^{(\rho+1)}} \Big[u_0(\Psi_{nt}^{(i)}) + \varphi_0(p_0(\Psi_{nt}^{(i)})) \\ &+ \textstyle\sum_{k=1}^{3} [p_k(\Psi_{nt}^{(i)})(q(p_k(\Psi_{nt}^{(i)})) + \varphi_k(p_k(\Psi_{nt}^{(i)})) \\ &- \varphi_0(p_0(\Psi_{nt}^{(i)})))] \frac{F_j(\Psi_{nt}^{(i)}|\Psi_{nt})}{\partial g_{nt}} \Big] \Big\}, \end{aligned} \tag{A.1.15}$$

where $g_{nt} = \{h_{nt}, s_{nt}\}$. Assume that $\varepsilon_{ont}, \cdots, \varepsilon_{nt3}$ are identically and independently distributed over (n,t) as Type 1 extreme value random variables. This assumption leads to convenient representations for the differences in the conditional valuation functions, and the expected values of the alternative specific unobservables when their corresponding alternative have been chosen. Specifically we have that $q(p_k(\Psi_{nt})) = \ln\left[\frac{p_k(\Psi_{nt})}{p_0(\Psi_{nt})}\right]$, $\varphi_k(p_k(\Psi_{nt})) = \gamma - \ln(p_k(\Psi_{nt}))$, and $\varphi_k(p_k(\Psi_{nt})) - \varphi_0(p_0(\Psi_{nt})) = -\ln\left[\frac{p_k(\Psi_{nt})}{p_0(\Psi_{nt})}\right]$.

Note that the transition matrix is degenerate conditional on the individual choosing not to enroll in school. If he chooses to enroll in school, the probability of advancing a grade level is $F(x_{nt})$. This implies that the transition probabilities for $i = 1, \cdots, \rho$ are given by $F(\Psi_{nt,j}^{(i)}|\Psi_{nt}) = 1$, for $j = 0,1$, $F(\Psi^{(i)}_{nt,j}|\Psi_{nt}) = F(x_{nt})$ for $j = 2,4$, and $F(\Psi^{(i)}_{nt,j}|\Psi_{nt}) = (1-F(x_{nt}))$ for $j = 2,4$. Define $\xi nt \equiv (1-\alpha)^{-1}\ln(\eta_n\lambda_t)$. Then we marginal utility of consumption can be expressed as $\eta_n\lambda_t \equiv \exp((1-\alpha)\xi_{nt})$.

The parametric assumptions on the utility functions and the idiosyncratic taste shifters, and the Euler conditions for work and schooling from equation (A.1.16) are used to form population moment conditions. We can then define

$$\begin{aligned} m_{nt1}(\Theta) \equiv \;& d_{nt1}\Big[\alpha^{-1}\eta_n\lambda_t w_{nt} - z'_{nt}B_5 - 2\delta_0 l^{(1)}_{nt} - \textstyle\sum_{i=1}^{\rho}\delta_i(l_{nt-i}+\beta^i) \\ & - \textstyle\sum_{i=1}^{\rho}\beta^i p_0(\Psi^{(i)}_{nt1})^{-1}\frac{\partial p_0(\Psi^{(i)}_{nt1})}{\partial h_{nt}}\Big] \\ & + d_{nt3}\Big[\alpha^{-1}\eta_n\lambda_t w_{nt} - z'_{nt}B_5 - 2\delta_0 l^{(3)}_{nt} - \textstyle\sum_{i=1}^{\rho}\delta_i(l_{nt-i}+\beta^i) \\ & - \textstyle\sum_{i=1}^{\rho}\beta^i\Big[p_0(\Psi^{(i)}_{nt4})^{-1}\frac{\partial p_0(\Psi^{(i)}_{nt4})}{\partial h_{nt}}F(x_{nt}) + p_0(\Psi^{(i)}_{nt5})^{-1}\frac{\partial p_0(\Psi^{(i)}_{nt5})}{\partial h_{nt}}(1-F(x_{nt})) \\ & + \ln\left(\frac{p_0(\Psi^{(i)}_{nt5})}{p_0(\Psi^{(i)}_{nt4})}\right)\frac{\partial F(x_{nt})}{\partial h_{nt}}\Big]\Big]. \end{aligned}$$

$$\begin{aligned} m_{nt2}(\Theta) \equiv \;& d_{nt2}\Big[-z'_{nt}B_5 - 2\delta_0 l^{(2)}_{nt} - \textstyle\sum_{i=1}^{\rho}\delta_i(l_{nt-i}+\beta^i) \\ & - \textstyle\sum_{i=1}^{\rho}\beta^i\Big[p_0(\Psi^{(i)}_{nt2})^{-1}\frac{\partial p_0(\Psi^{(i)}_{nt2})}{\partial s_{nt}}F(x_{nt}) + p_0(\Psi^{(i)}_{nt3})^{-1}\frac{\partial p_0(\Psi^{(i)}_{nt3})}{\partial s_{nt}}(1-F(x_{nt})) \\ & + \ln\left(\frac{p_0(\Psi^{(i)}_{nt3})}{p_0(\Psi^{(i)}_{nt2})}\right)\frac{\partial F(x_{nt})}{\partial s_{nt}}\Big]\Big] + d_{nt3}\Big[-z'_{nt}B_5 - 2\delta_0 l^{(3)}_{nt} - \textstyle\sum_{i=1}^{\rho}\delta_i(l_{nt-i}+\beta^i) \\ & - \textstyle\sum_{i=1}^{\rho}\beta^i\Big[p_0(\Psi^{(i)}_{nt4})^{-1}\frac{\partial p_0(\Psi^{(i)}_{nt4})}{\partial s_{nt}}F(x_{nt}) + p_0(\Psi^{(i)}_{nt5})^{-1}\frac{\partial p_0(\Psi^{(i)}_{nt5})}{\partial s_{nt}}(1-F(x_{nt})) \\ & + \ln\left(\frac{p_0(\Psi^{(i)}_{nt5})}{p_0(\Psi^{(i)}_{nt4})}\right)\frac{\partial F(x_{nt})}{\partial s_{nt}}\Big]\Big]. \end{aligned}$$

The parametric assumptions on the utility functions, the distribution of the idiosyncratic taste shifters, equation (A.1.12) and equation (A.1.14) are used to obtain the following additional moment conditions[1]

$$
\begin{aligned}
m_{nt3}(\Theta) \equiv \;& d_{nt1}\left[\ln\left(\frac{p_{nt1}}{p_{nt0}}\right) - x_{nt}'B_6 + x_{nt}'B_7(l_{nt}^{(0)} - l_{nt}^{(1)}) + \delta_0(l_{nt}^{(0)2} - l_{nt}^{(1)2})\right. \\
& \left. + \sum_{i=1}^{\rho}\delta_i(l_{nt}^{(0)} - l_{nt}^{(1)})(l_{nt-i} + \beta^i) - \frac{\eta_n\lambda_t}{\alpha}(w_{nt}h_{nt}) - \sum_{i=1}^{\rho}\beta^i \ln\left(\frac{p_0(\Psi_{nt0}^{(s)})}{p_0(\Psi_{nt1}^{(s)})}\right)\right],
\end{aligned}
$$

$$
\begin{aligned}
m_{nt4}(\Theta) \equiv \;& d_{nt2}\left[\ln\left(\frac{p_{nt2}}{p_{nt0}}\right) - x_{nt}'B_5 + x_{nt}'B_7(l_{nt}^{(0)} - l_{nt}^{(2)})\right. \\
& + \delta_0(l_{nt}^{(0)2} - l_{nt}^{(2)2}) + \sum_{i=1}^{\rho}\delta_i(l_{nt}^{(0)} - l_{nt}^{(2)})(l_{nt-i} + \beta^i) + \frac{\eta_n\lambda_t}{\alpha}\pi_{nt} \\
& \left. - \sum_{i=1}^{\rho}\beta^i\left[\ln p_0(\Psi_{nt0}^{(i)}) - \ln p_0(\Psi_{nt2}^{(i)})F(x_{nt}) - \ln p_0(\Psi_{nt3}^{(i)})(1 - F(x_{nt}))\right]\right],
\end{aligned}
$$

$$
\begin{aligned}
m_{nt5}(\Theta) \equiv \;& d_{nt3}\left[\ln\left(\frac{p_{nt3}}{p_{nt0}}\right) - x_{nt}'B_5 - x_{nt}'B_6 + x_{nt}'B_7(l_{nt}^{(0)} - l_{nt}^{(3)})\right. \\
& + \delta_0(l_{nt}^{(0)2} - l_{nt}^{(3)2}) + \sum_{i=1}^{\rho}\delta_i(l_{nt}^{(0)} - l_{nt}^{(3)})(l_{nt-i} + \beta^i) - \frac{\eta_n\lambda_t}{\alpha}(w_{nt}h_{nt} - \pi_{nt}) \\
& \left. - \sum_{i=1}^{\rho}\beta^i\left[\ln p_0(\Psi_{nt0}^{(i)}) - \ln p_0(\Psi_{nt4}^{(i)})F(x_{nt}) - \ln p_0(\Psi_{nt5}^{(i)})(1 - F(x_{nt}))\right]\right].
\end{aligned}
$$

A.1.5 Consistent Asymptotic Variance Estimation

Some preliminary results are in needed. The first is concerned with the estimation of the CCP's themselves. In estimation, a the data was trimmed to ensure that the density is bounded away from zero. This fixed trimming condition defines a compact subset of the support of the density over which the density affects the estimator. Assumptions 8.1 - 8.3, and the assumptions in Lemma 8.10 of Newey and McFadden [1994] ensures the resulting kernel density estimators of the CCP's and their derivatives converge uniformly:

$$
\sqrt{N}\|p^N(\Psi) - p^0(\Psi)\|^2 \xrightarrow{p} 0, \tag{A.1.16}
$$

where the norm is the Sobolev norm. Assume that , θ^N is the unique solution to:

$$
\frac{1}{N}\sum_{n=1}^{N} m(x_n, \theta, \xi_n(B_1^N), s_n(B_3^N)F_n(s_n(B_3^N), B_4^N), p_n^N). \tag{A.1.17}
$$

Assume also that $\theta_0 \in \Theta$, a compact set. Inspection of the equations in (**??**) shows that $m(x, \theta)$ is continuous in each θ. Further inspection along with the fixed trimming condition on the data in

[1] The construction of the moment conditions show that the choice of the normalizing alternative (alternative 0) is not completely arbitrary. This alternative has to sufficiently saturate the state space so that $p_{nt0} > 0$ and $p_0(\Psi_{ntj}^i) > 0$.

estimation implies that $m(z,\theta)$ is uniformly bounded over θ. These conditions ensures that $\theta^N \xrightarrow{p} \theta_0$ as shown in Theorem 2.6 of Newey and McFadden [1994].

Define the following influence functions from equations (**??**) and from the definitions in section A.1.2

$$\begin{aligned} \varphi_1(x_{1n}) &\equiv -E[\Delta x'_{1n} A_n^{-1} \Delta x_{1n}]^{-1} \Delta x'_{1n} A_n^{-1} \Delta v_{1n}, \quad \varphi_3(x_{3n}) \equiv -H_3^{-1} h(x_{3n}), \\ \varphi_4(x_{4n}) &\equiv -H_4^{-1} h(x_{4n}). \end{aligned} \tag{A.1.18}$$

Define the following matrices

$$M_{1nt} \equiv \begin{bmatrix} (d_{nt1}+d_{nt3})(\frac{1-\alpha}{\alpha})\exp((1-\alpha)\xi_{nt}))w_{nt} \\ 0 \\ -d_{nt1}(\frac{1-\alpha}{\alpha})\exp((1-\alpha)\xi_{nt}))w_{nt}h_{nt} \\ d_{nt2}(\frac{1-\alpha}{\alpha})\exp((1-\alpha)\xi_{nt}))\pi_{nt} \\ d_{nt3}(\frac{1-\alpha}{\alpha})\exp((1-\alpha)\xi_{nt}))(w_{nt}h_{nt}-\pi_{nt}) \end{bmatrix} \left[-\frac{1}{N}\sum_n x'_{1nt}\right],$$

$$M_{1n}(x_n) \equiv (M'_{1n1}, \cdots, M'_{1nT})', \text{ and, } \alpha_1(x_n) \equiv E[M_{1n}]\varphi_1(x_{1n}). \tag{A.1.19}$$

$$M_{2nt} \equiv \begin{bmatrix} d_{nt1}\sum_i d^s_{nt-i}\delta_i + \\ d_{nt3}\left[2\delta_0 + \sum_i \left(d^s_{nt-i}\delta_i - \beta^i\left(\left(\frac{1}{p^i_{0nt4}}\frac{\partial p^i_{0nt4}}{\partial h_{nt}} - \frac{1}{p^i_{0nt5}}\frac{\partial p^i_{0nt5}}{\partial h_{nt}}\right)\frac{\partial F(x_{nt})}{\partial s_{nt}} + \ln(\frac{p^i_{0nt5}}{p^i_{0nt4}})\frac{\partial^2 F(x_{nt})}{\partial h_{nt}\partial s_{nt}}\right)\right)\right] \\ d_{nt2}\left[2\delta_0 + \sum_i \left(d^s_{nt-i}\delta_i - \beta^i\left(\left(\frac{1}{p^i_{0nt2}}\frac{\partial p^i_{0nt2}}{\partial s_{nt}} - \frac{1}{p^i_{0nt3}}\frac{\partial p^i_{0nt3}}{\partial s_{nt}}\right)\frac{\partial F(x_{nt})}{\partial s_{nt}} + \ln(\frac{p^i_{0nt3}}{p^i_{0nt2}})\frac{\partial^2 F(x_{nt})}{\partial s^2_{nt}}\right)\right)\right] + \\ d_{nt3}\left[2\delta_0 + \sum_i \left(d^s_{nt-i}\delta_i - \beta^i\left(\left(\frac{1}{p^i_{0nt4}}\frac{\partial p^i_{0nt54}}{\partial s_{nt}} - \frac{1}{p^i_{0nt5}}\frac{\partial p^i_{0nt5}}{\partial s_{nt}}\right)\frac{\partial F(x_{nt})}{\partial s_{nt}} + \ln(\frac{p^i_{0nt5}}{p^i_{0nt4}})\frac{\partial^2 F(x_{nt})}{\partial s^2_{nt}}\right)\right)\right] \\ -d_{nt1}\sum_i \delta_i(l^0_{nt} - l^1_{nt})d^s_{nt-i} \\ d_{nt2}\left[x'_{6nt}B_6 + 2\delta_0 l^2_{nt} + \sum_i \delta_i l_{nt-i}\sum_i \beta^i\left[\ln(\frac{p^i_{0nt2}}{p^i_{0nt3}})\frac{\partial F(x_{nt})}{\partial s_{nt}}\right]\right] \\ d_{nt3}\left[x'_{6nt}B_6 + 2\delta_0 l^3_{nt} + \sum_i \delta_i l_{nt-i}\sum_i \beta^i\left[\ln(\frac{p^i_{0nt4}}{p^i_{0nt5}})\frac{\partial F(x_{nt})}{\partial s_{nt}}\right]\right] \end{bmatrix} \left[s_{nt}x'_{3nt}\right],$$

$$M_{2n}(x_n) \equiv (M'_{2n1}, \cdots, M'_{2nT})', \text{ and, } \alpha_2(x_n) \equiv E[M_{2n}]\varphi_2(x_{3n}). \tag{A.1.20}$$

$$M_{4nt} \equiv \begin{bmatrix} -d_{nt3}\sum_i \beta^i\left[\frac{1}{p^i_{0nt4}}\frac{\partial p^i_{0nt4}}{\partial h_{nt}} - \frac{1}{p^i_{0nt5}}\frac{\partial p^i_{0nt5}}{\partial h_{nt}} + \ln(\frac{p^i_{0nt5}}{p^i_{0nt4}})\frac{\partial^2 F(x_{nt})}{\partial h^2_{nt}}\right] \\ -d_{nt2}\sum_i \beta^i\left[\frac{1}{p^i_{0nt2}}\frac{\partial p^i_{0nt2}}{\partial s_{nt}} - \frac{1}{p^i_{0nt3}}\frac{\partial p^i_{0nt3}}{\partial s_{nt}} + \ln(\frac{p^i_{0nt3}}{p^i_{0nt2}})\frac{\partial^2 F(x_{nt})}{\partial h_{nt}\partial F}\right] - \\ d_{nt3}\sum_i \beta^i\left[\frac{1}{p^i_{0nt4}}\frac{\partial p^i_{0nt4}}{\partial s_{nt}} - \frac{1}{p^i_{0nt5}}\frac{\partial p^i_{0nt5}}{\partial s_{nt}} + \ln(\frac{p^i_{0nt5}}{p^i_{0nt4}})\frac{\partial^2 F(x_{nt})}{\partial h_{nt}\partial F}\right] \\ 0 \\ d_{nt2}\sum_i \beta^i\left[\ln(\frac{p^i_{0nt2}}{p^i_{0nt3}})\right] \\ d_{nt3}\sum_i \beta^i\left[\ln(\frac{p^i_{0nt4}}{p^i_{0nt5}})\right] \end{bmatrix} \left[F(x_{nt})(1-F(x_{nt}))x'_{4nt}\right],$$

$$M_{4n}(x_n) \equiv (M'_{4n1}, \cdots, M'_{4nT})', \text{ and, } \alpha_4(x_n) \equiv E[M_{4n}]\varphi_4(x_{4n}). \tag{A.1.21}$$

$$D_{nt0} \equiv E\left[\frac{\partial m_{nt}}{\partial p_{nt0}}|\Psi_{nt}\right] = -p_{nt0}^{-1}(0,0,p_{nt1},p_{nt2},p_{nt3})'$$
$$D_{n0}(x_n) \equiv (D'_{n10}, \cdots, D'_{nT0})', \text{ and, } \alpha_5(x_n) \equiv D_{n0}[d_{n0} - p_{n0}]. \tag{A.1.22}$$

$$D_{nt1} \equiv E\left[\frac{\partial m_{nt}}{\partial p_{nt1}}|\Psi_{nt}\right] = (0,0,1,0,0)'$$
$$D_{n1}(x_n) \equiv (D'_{n11}, \cdots, D'_{nT1})', \text{ and, } \alpha_6(x_n) \equiv D_{n1}[d_{n1} - p_{n1}]. \tag{A.1.23}$$

$$D_{nt2} \equiv E\left[\frac{\partial m_{nt}}{\partial p_{nt2}}|\Psi_{nt}\right] = (0,0,0,1,0)'$$
$$D_{n2}(x_n) \equiv (D'_{n12}, \cdots, D'_{nT2})', \text{ and, } \alpha_7(x_n) \equiv D_{n2}[d_{n2} - p_{n2}]. \tag{A.1.24}$$

$$D_{nt3} \equiv E\left[\frac{\partial m_{nt}}{\partial p_{nt3}}|\Psi_{nt}\right] = (0,0,0,0,1)'$$
$$D_{n3}(x_n) \equiv (D'_{n13}, \cdots, D'_{nT3})', \text{ and, } \alpha_8(x_n) \equiv D_{n3}[d_{n3} - p_{n3}]. \tag{A.1.25}$$

For $i = 1, \cdots, \rho$ define.

$$D_{nt0i} \equiv E\left[\frac{\partial m_{nt}}{\partial p_{0nt0}^{(i)}}|\Psi_{nt0}^{(i)}\right] = \beta^i\left(0,0,\frac{p_{1nt0}^{(i)}}{p_{0nt0}^{(i)}},\frac{p_{2nt0}^{(i)}}{p_{0nt0}^{(i)}},\frac{p_{3nt0}^{(i)}}{p_{0nt0}^{(i)}}\right)'$$
$$D_{n0i}(x_n) \equiv (D'_{n10i}, \cdots, D'_{nT0i})', \text{ and, } \alpha_{9i}(x_n) \equiv D_{n0i}[d_{n0} - p_{n0}^{(i)}]. \tag{A.1.26}$$

$$D_{nt1i} \equiv E\left[\frac{\partial m_{nt}}{\partial p_{0nt1}^{(i)}}|\Psi_{nt1}^{(i)}\right] = \beta^i\left(\frac{p_{1nt1}^{(i)}}{(p_{0nt1}^{(i)})^2}\nabla_h p_{0nt1}^{(i)},0,\frac{p_{1nt1}^{(i)}}{p_{0nt1}^{(i)}},0,0\right)'$$
$$D_{n1i}(x_n) \equiv (D'_{n11i}, \cdots, D'_{nT1i})', \text{ and, } \alpha_{10i}(x_n) \equiv D_{n1i}[d_{n1} - p_{n1}^{(i)}]. \tag{A.1.27}$$

$$D_{nt2i} \equiv E\left[\frac{\partial m_{nt}}{\partial p_{0nt2}^{(i)}}|\Psi_{nt2}^{(i)}\right] = \beta^i\left(0,\frac{p_{2nt2}^{(i)}}{(p_{0nt2}^{(i)})^2}\nabla_s p_{0nt2}^{(i)}F(x_{nt}) - \frac{p_{2nt2}^{(i)}}{p_{0nt2}^{(i)}}\nabla_s F(x_{nt}),0,\frac{p_{2nt2}^{(i)}}{p_{0nt2}^{(i)}}F(x_{nt}),0\right)'$$
$$D_{n2i}(x_n) \equiv (D'_{n12i}, \cdots, D'_{nT2i})', \text{ and, } \alpha_{11i}(x_n) \equiv D_{n2i}[d_{n0} - p_{0n2}^{(i)}]. \tag{A.1.28}$$

$$D_{nt3i} \equiv E\left[\frac{\partial m_{nt}}{\partial p_{0nt3}^{(i)}}|\Psi_{nt3}^{(i)}\right] = \beta^i\left(0,\frac{p_{2nt3}^{(i)}}{(p_{0nt3}^{(i)})^2}\nabla_s p_{0nt3}^{(i)}(1-F(x_{nt})) + \frac{p_{2nt3}^{(i)}}{p_{0nt3}^{(i)}}\nabla_s F(x_{nt}),0,\frac{p_{2nt3}^{(i)}}{p_{0nt3}^{(i)}}(1-F(x_{nt})),0\right)'$$
$$D_{n3i}(x_n) \equiv (D'_{n13i}, \cdots, D'_{nT3i})', \text{ and, } \alpha_{12i}(x_n) \equiv D_{n3i}[d_{n0} - p_{0n3}^{(i)}]. \tag{A.1.29}$$

$$D_{nt4i} \equiv E\left[\frac{\partial m_{nt}}{\partial p^{(i)}_{0nt4}}|\Psi^{(i)}_{nt4}\right] = \beta^i \begin{bmatrix} \frac{p^{(i)}_{3nt4}}{(p^{(i)}_{0nt4})^2}\nabla_h p^{(i)}_{0nt4}F(x_{nt}) - \frac{p^{(i)}_{3nt4}}{p^{(i)}_{0nt4}}\nabla_h F(x_{nt}) \\ \frac{p^{(i)}_{3nt4}}{(p^{(i)}_{0nt4})^2}\nabla_s p^{(i)}_{0nt4}F(x_{nt}) - \frac{p^{(i)}_{3nt4}}{p^{(i)}_{0nt4}}\nabla_s F(x_{nt}) \\ 0 \\ 0 \\ \frac{p^{(i)}_{3nt4}}{p^{(i)}_{0nt4}}\nabla_s F(x_{nt} \end{bmatrix}$$

$$D_{n4i}(x_n) \equiv (D'_{n14i}, \cdots, D'_{nT4i})', \text{ and, } \alpha_{13i}(x_n) \equiv D_{n4i}[d_{n0} - p^{(i)}_{0n4}]. \quad \text{(A.1.30)}$$

$$D_{nt5i} \equiv E\left[\frac{\partial m_{nt}}{\partial p^{(i)}_{0nt5}}|\Psi^{(i)}_{nt5}\right] = \beta^i \begin{bmatrix} \frac{p^{(i)}_{3nt5}}{(p^{(i)}_{0nt5})^2}\nabla_h p^{(i)}_{0nt5}(1 - F(x_{nt})) + \frac{p^{(i)}_{3nt5}}{p^{(i)}_{0nt5}}\nabla_h F(x_{nt}) \\ \frac{p^{(i)}_{3nt5}}{(p^{(i)}_{0nt5})^2}\nabla_s p^{(i)}_{0nt5}F(x_{nt}) + \frac{p^{(i)}_{3nt5}}{p^{(i)}_{0nt5}}\nabla_s F(x_{nt}) \\ 0 \\ 0 \\ \frac{p^{(i)}_{3nt5}}{p^{(i)}_{0nt5}}\nabla_s F(x_{nt} \end{bmatrix}$$

$$D_{n5i}(x_n) \equiv (D'_{n15i}, \cdots, D'_{nT5i})', \text{ and, } \alpha_{14i}(x_n) \equiv D_{n5i}[d_{n0} - p^{(i)}_{0n5}]. \quad \text{(A.1.31)}$$

Let $f^i_{ntj} \equiv f(\Psi^i_{ntj})$ be the density of Ψ^i_{ntj} $j = 1, \cdots, 5,\ i = 1, \cdots, \rho$. Define also $\vartheta^i_{ntj} \equiv (f(\Psi^i_{ntj}))^{-1}\frac{\partial f(\Psi^i_{ntj})}{\partial h_n}$.

For $i = 1, \cdots, \rho$ let

$$
\begin{aligned}
{}_hM_{nt1i} &\equiv E\left[\frac{\partial m_{nt}}{\partial \nabla_h p^{(i)}_{0nt1}} | \Psi^{(i)}_{nt1}\right] = \beta^i \left(\frac{p^{(i)}_{1nt1}}{p^{(i)}_{0nt1}}, 0,0,0,0\right)' \\
{}_{hh}M_{nt1i} &\equiv E\left[\frac{\partial m_{nt}}{\partial \nabla_h p^{(i)}_{0nt1} \partial h_{nt}} | \Psi^{(i)}_{nt1}\right] = \beta^i \left(\frac{p^{(i)}_{1nt1}}{(p^{(i)}_{0nt1})^2} \nabla_h p^{(i)}_{0nt1}, 0,0,0,0\right)' \\
{}_hD_{nt1i} &\equiv -\left[{}_{hh}M_{nt1i} + 2\, {}_hM_{nt1i} \vartheta^i_{nt1}\right] && \text{(A.1.32)} \\
{}_hD_{n1i}(x_n) &\equiv (\, {}_hD'_{n11i}, \cdots, \, {}_hD'_{nT1i})', \text{ and, } \alpha_{15i}(x_n) \equiv \, {}_hD_{n1i}[d_{n0} - p^{(i)}_{0n1}]. && \text{(A.1.33)}
\end{aligned}
$$

$$
\begin{aligned}
{}_sM_{nt2i} &\equiv E\left[\frac{\partial m_{nt}}{\partial \nabla_s p^{(i)}_{0nt2}} | \Psi^{(i)}_{nt2}\right] = \beta^i \left(0, \frac{p^{(i)}_{2nt2}}{p^{(i)}_{0nt2}} F(x_{nt}), 0,0,0\right)' \\
{}_{ss}M_{nt2i} &\equiv E\left[\frac{\partial m_{nt}}{\partial \nabla_s p^{(i)}_{0nt2} \partial s_{nt}} | \Psi^{(i)}_{nt2}\right] = \beta^i \left(0, \frac{p^{(i)}_{2nt2}}{(p^{(i)}_{0nt2})^2} \nabla_s p^{(i)}_{0nt1} F(x_{nt}) - \frac{p^{(i)}_{2nt2}}{p^{(i)}_{0nt2}} \nabla_s F(x_{nt}), 0,0,0\right)' \\
{}_sD_{nt2i} &\equiv -\left[{}_{ss}M_{nt2i} + 2\, {}_sM_{nt2i} \vartheta^i_{nt2}\right] && \text{(A.1.34)} \\
{}_sD_{n2i}(x_n) &\equiv (\, {}_sD'_{n12i}, \cdots, \, {}_sD'_{nT2i})', \text{ and, } \alpha_{16i}(x_n) \equiv \, {}_sD_{n2i}[d_{n0} - p^{(i)}_{0n2}]. && \text{(A.1.35)}
\end{aligned}
$$

$$
\begin{aligned}
{}_sM_{nt3i} &\equiv E\left[\frac{\partial m_{nt}}{\partial \nabla_s p^{(i)}_{0nt3}} | \Psi^{(i)}_{nt3}\right] = \beta^i \left(0, \frac{p^{(i)}_{3nt3}}{p^{(i)}_{0nt3}} (1 - F(x_{nt})), 0,0,0\right)' \\
{}_{ss}M_{nt3i} &\equiv E\left[\frac{\partial m_{nt}}{\partial \nabla_s p^{(i)}_{0nt3} \partial s_{nt}} | \Psi^{(i)}_{nt3}\right] = \beta^i \left(0, \frac{p^{(i)}_{3nt3}}{(p^{(i)}_{0nt3})^2} \nabla_s p^{(i)}_{0nt1} (1 - F(x_{nt})) + \frac{p^{(i)}_{3nt3}}{p^{(i)}_{0nt3}} \nabla_s F(x_{nt}), 0,0,0\right)' \\
{}_sD_{nt3i} &\equiv -\left[{}_{ss}M_{nt3i} + 2\, {}_sM_{nt3i} \vartheta^i_{nt3}\right] && \text{(A.1.36)} \\
{}_sD_{n3i}(x_n) &\equiv (\, {}_sD'_{n13i}, \cdots, \, {}_sD'_{nT3i})', \text{ and, } \alpha_{17i}(x_n) \equiv \, {}_sD_{n3i}[d_{n0} - p^{(i)}_{0n3}]. && \text{(A.1.37)}
\end{aligned}
$$

The construction of ${}_hD_{nt4i}$ and ${}_sD_{nt4i}$ are the same as ${}_sD_{nt2i}$ with the correct indexes. Likewise, the construction of ${}_hD_{nt5i}$ and ${}_sD_{nt45i}$ are the same as ${}_sD_{nt3i}$ with the correct indexes. This gives additional influence functions $\alpha_{18i}, \cdots, \alpha_{21i}$. Define also $\alpha(x_n) \equiv \sum_{j=1}^{8} \alpha_{nj}(x_n) + \sum_{j=9}^{21} \sum_{i=1}^{\rho} \alpha_{ji}(x_n)$. The fixed trimming condition, the smoothness properties of $m(x, \cdot)$, and condition A.1.16 ensures linearization is possible in the necessary arguments, that the above matrices are well defined (in particular, all expectations are well defined), and that assumptions 5.1-5.6 of Newey [1994] are satisfied. Define

$$
\begin{aligned}
M_\theta &\equiv E\left[\frac{\partial m(x_n, \theta_0)}{\partial \theta}\right] && \text{(A.1.38)} \\
W &\equiv E[\{m(x_n, \theta_0) + \alpha(x_n)\}\{m(x_n, \theta_0) + \alpha(x_n)\}'] && \text{(A.1.39)}
\end{aligned}
$$

Therefore, by lemma 5.3 of Newey [1994], we have that

$$\sqrt{N}(\theta_N - \theta_0) \xrightarrow{p} N(0, V),$$

where

$$V \equiv (M'_\theta \Omega^{-1} M_\theta)^{-1} M'_\theta \Omega^{-1} W \Omega^{-1} M_\theta (M'_\theta \Omega^{-1} M_\theta)^{-1} \quad \text{(A.1.40)}$$

A consistent estimator of jacobians with respect to the finite dimensional parameters are obtained by replacing the parameters (both finite and infinite dimensional) with their respective estimates and taking averages over N. A consistent estimator jacobians with respect to the ccp's and their derivatives are obtained by replacing the parameters with their estimated counterparts and then performing nonparametric regression of these quantities on their appropriate conditioning vectors Ψ^{i}_{nj}. The residuals needed to complete the formation of $\hat{\alpha}(x_n)$ are readily obtained from all the parametric and nonparametric pre-estimates. By similar substitutions and averaging consistent estimates of M_θ $m(x_{nt},\theta)$, and Ω are formed, denoted by M^{n}_{θ}, $m^{N}(x_n)$, and Ω^N, A consistent estimate of W is then obtained by

$$W^N = N^{-1}\sum_{n=1}^{N}\left[m^N(x_n)+\alpha^N(x_n)\right]\left[m^N(x_n)+\alpha^N(x_n)\right]'. \tag{A.1.41}$$

Putting all these estimated quantities together, a consistent estimator for the asymptotic variance is given by

$$V^N \equiv \left(M^{N\prime}_{\theta}(\Omega^N)^{-1}M^{N}_{\theta}\right)^{-1}M^{N\prime}_{\theta}\left(\Omega^N\right)^{-1}W^N\left(\Omega^N\right)^{-1}M^{N}_{\theta}\left(M^{N\prime}_{\theta}\left(\Omega^N\right)^{-1}M^{N}_{\theta}\right)^{-1}. \tag{A.1.42}$$

A.2 APPENDIX TO CHAPTER 3

A.2.1 Proof of Theorem 3.3.2

Proof. Inverting the index functions φ_0 and φ_1 in equations (3.2.3) and (3.3.1) respectively gives:

$$\begin{aligned} \varphi_0^{-1}(x_{it}'\beta_0 + f_0(z_i)) &= \varphi_1^{-1}(x_{it}'\beta_1 + f_1(z_i)) \Leftrightarrow \\ x_{it}'\beta_0 + f_0(z_i) &= \varphi_0(\varphi_1^{-1}(x_{it}'\beta_1 + f_1(z_i))), \end{aligned} \tag{A.2.1}$$

since both sides of the first equality are equal to P_{it0}. Also, since the index function is strictly increasing, it is differentiable almost everywhere. Differentiating equation (A.2.1) with respect to the continuous regressor x_{itk} gives:

$$a := \frac{\beta_{0k}}{\beta_{1k}} = \frac{\varphi_0'(\varphi_1^{-1}(x_{it}'\beta_1 + f_1(z_i)))}{\varphi_1'(\varphi_1^{-1}(x_{it}'\beta_1 + f_1(z_i)))} > 0, \tag{A.2.2}$$

where the positive sign follows trivially from the assumption that the index function is strictly increasing. We have from equation (A.2.2) that $\varphi_0'(P_{it0}) = a\varphi_1'(P_{it0})$ which implies that:

$$\varphi_0(P_{it0}) = a\varphi_1(P_{it0}) + c. \tag{A.2.3}$$

Taking first difference of equations (3.2.3), (3.3.1) and (A.2.3) we have that:

$$\begin{aligned} \Delta[\varphi_0(P_{it0})] &= \Delta x_{it}'\beta_0 \\ \Delta[\varphi_1(P_{it0})] &= \Delta x_{it}'\beta_1 \\ \Delta[\varphi_0(P_{it0})] &= a\Delta[\varphi_1(P_{it0})] \end{aligned} \tag{A.2.4}$$

which implies that

$$\begin{aligned} a\Delta[\varphi_1(P_{it0})] &= \Delta x_{it}'\beta_0 \\ a\Delta[\varphi_1(P_{it0})] &= a\Delta x_{it}'\beta_1. \end{aligned} \tag{A.2.5}$$

Equating the RHS of the equations in (A.2.5), pre-multiplying by $\Delta x_{it}'$ and taking expectations gives:

$$E[\Delta x_{it}\Delta x_{it}']\beta_0 = aE[\Delta x_{it}\Delta x_{it}']\beta_1. \tag{A.2.6}$$

Then by the invertibility of $E[\Delta x_{it}\Delta x_{it}']$ we have

$$\beta_0 = a\beta_1. \tag{A.2.7}$$

The assumption that $||\beta_0|| = ||\beta_1|| = 1$ implies from equation (A.2.7)that $|a| = 1$. But $a > 0$, which implies that $a = 1$. Thus equations (A.2.7) and (A.2.3) imply that:

$$\beta_0 = \beta_1 \tag{A.2.8}$$

$$\varphi_0(P_{it}) = \varphi_1(P_{it}) + c. \tag{A.2.9}$$

From equations (A.2.8) and (A.2.9), (3.2.3) becomes:

$$\begin{aligned} \varphi_1(P_{it0}) + c &= \Delta x_{it}'\beta_1 + f_0(z_i) \\ \Rightarrow \Delta x_{it}'\beta_1 + f_1(z_i) + c &= \Delta x_{it}'\beta_1 + f_0(z_i) \\ \Rightarrow f_1(z_i) + c &= f_0(z_i). \end{aligned} \tag{A.2.10}$$

This completes the first part of the proof. The fact that $c = 0$, follows from assumption (3.3.1.1) and equation (A.2.10) by taking the expectations of both sides of equation (A.2.10). □

A.2.2 Finite Entropy Lemma

For any given $R > 0$, let

$$\mathcal{G} := \{S(x,\beta,\varphi,P_0) \mid \|\beta\|_K \leq 1, \varphi \in \mathcal{S}_{\mathcal{K}}\}.$$

Then we have the following result:

Lemma A.2.1. If assumption (3.6.2) holds, then

1. The class $\mathcal{G}$ is uniformly bounded.
2. For any $\delta > 0$,

$$H_\infty(\delta,\mathcal{G}) \leq C_1 \ln\left(\frac{4C_2+\delta}{\delta}\right) + \frac{C_3}{\delta},$$

   for some $C_1 > 0$, $C_2 > 0$ and $C_3 > 0$, where $H_\infty(\delta,\mathcal{G})$ is the δ-entropy of $\mathcal{G}$ for the supremum norm (See Definition 2.3 of van de Geer [2000]).

Proof. Since η is continuous, it is bounded over $\mathcal{K}$, say with lower and upper bounds R_1 and R_2 respectively. Thus the set $\mathcal{S}_{\mathcal{K}}$ is uniformly bounded. Thus, $S(x,\beta,\varphi,P_0) \leq (\|\Delta X_i\|_2^T.\|\beta\|_2 + 2(T-1)\|\varphi\|)^2 \leq (R_0 + 2(T-1)R)^2$ for some $R > 0$.
Note that the entropy of the class $\mathcal{G}$ is at most that of the cartesian product $B_K(0,1) \times \mathcal{G}$. The ball $B_K(0,1)$ can be covered by $\left(\frac{8+\delta}{\delta}\right)^K$ balls with radius $\frac{\delta}{2}$ (van de Geer, 2000). Since a ball of radius δ can be covered by a K-dimensional cube of length 2δ, the ball $B_K(0,1)$ can be covered by $\left(\frac{8+\delta}{\delta}\right)^K$ cubes of diameter δ. This in turn implies that $H_\infty(\delta, B_K(0,1)) \leq K \ln\left(\frac{8+\delta}{\delta}\right)$. Furthermore, there is a constant C such that the entropy $H_\infty(\delta, \mathcal{C} := \{g : \mathcal{K} \to [R_1,R_2] \mid \int_{\mathcal{K}} |g'(x)| \mathrm{d}x \leq M\})$ has the upper bound $\frac{C}{\delta}$ for some $C > 0$, (van de Geer [2000]). $\mathcal{G}(R)$ is included in $\mathcal{C}$. The entropy bound now results from the fact that the entropy of a cartesian product is the sum of the entropies of the sets in the product. □

Corollary A.2.2. If assumptions (3.6.2) holds, then the following uniform convergence holds:

$$\sup_{(\beta,\varphi)\in\mathcal{G}} |S_N(x,\beta,\varphi,P_0) - S_0(x,\beta,\varphi,P_0)| \to 0, \text{ } Q\text{-almost surely}.$$

Proof. Lemma A.2.1 establishes that $\mathcal{G}$ has finite entropy in the sup-norm. Lemma 2.1 of van de Geer [2000] shows that the entropy in the sup-norm bounds above the entropy with bracketing for the $L_1(Q)$-metric, implying finite entropy with bracketing in the $L_1(Q)$-metric. Then apply Lemma 3.1 of van de Geer [2000] to see that $\mathcal{G}$ satisfies the ULLN. □

A.2.3 Proof of Theorem 3.6.3

Proof. Define

$$\begin{aligned}
m(x_i,\beta,\varphi,P_i) &:= \Delta\varphi(P_i) - \Delta x_i \beta \\
S(x_i,\beta,\varphi,P_i) &:= m(x_i,\beta,\varphi,P_i)' m(x_i,\beta,\varphi,P_i) \\
S_N(\beta,\varphi,P) &:= \frac{1}{N}\sum_{i=1}^{N} S(x_i,\beta,\varphi,P_i) \\
S_0(\beta,\varphi,P) &:= E[S(x,\beta,\varphi,P)].
\end{aligned}$$

Clearly, the pair $(\hat{\beta},\hat{\varphi})$ is defined to be minimizing $S_N(\beta,\varphi,\hat{P})$, and (β_0,φ_0) minimizes $S_0(\beta,\varphi,P_0)$. In fact $S_0(\beta_0,\varphi_0,P_0)=S_N(\beta_0,\varphi_0,P_0)=0$. Thus

$$S_N(\hat{\beta},\hat{\varphi},\hat{P}) \quad \leq \quad S_N(\beta_0,\varphi_0,\hat{P}) \xrightarrow{p} 0, \tag{A.2.11}$$

by the continuous mapping theorem. Thus from equation (A.2.11) we have:

$$\begin{aligned} 0 \leq S_0(\hat{\beta},\hat{\varphi},P_0) &= S_N(\hat{\beta},\hat{\varphi},\hat{P})+S_0(\hat{\beta},\hat{\varphi},P_0)-S_N(\hat{\beta},\hat{\varphi},\hat{P}) \\ &\leq |S_0(\hat{\beta},\hat{\varphi},P_0)-S_N(\hat{\beta},\hat{\varphi},P_0)| \\ &\quad +|S_N(\hat{\beta},\hat{\varphi},P_0)-S_N(\hat{\beta},\hat{\varphi},\hat{P})|+o_p(1) \end{aligned} \tag{A.2.12}$$

The first term of the RHS of the last inequality is an $o_p(1)$ by corollary (A.2.2). To see that the last term is also an $o_p(1)$, we add and subtract $m(x_i,\hat{\beta},\hat{\varphi},\hat{P}_i)'m(x_i,\hat{\beta},\hat{\varphi},P_{i0})$ to get the following:

$$\begin{aligned} & |S_N(\hat{\beta},\hat{\varphi},P_0)-S_N(\hat{\beta},\hat{\varphi},\hat{P})| \\ \leq\ & \frac{1}{N}\sum_{i=1}^{N}\left\{\left[\|m(x_i,\hat{\beta},\hat{\varphi},\hat{P}_i)\|_2^T+\|m(x_i,\hat{\beta},\hat{\varphi},P_{i0})\|_2^T\right]\|m(x_i,\hat{\beta},\hat{\varphi},\hat{P}_i)-m(x_i,\hat{\beta},\hat{\varphi},P_{i0})\|_2^T\right\} \\ \leq\ & C\frac{1}{N}\sum_{i=1}^{N}\left\{\|m(x_i,\hat{\beta},\hat{\varphi},\hat{P}_i)-m(x_i,\hat{\beta},\hat{\varphi},P_{i0})\|_2^T\right\} \\ \leq\ & C\frac{1}{N}\sum_{i=1}^{N}\sup_{\varphi\in S_{\mathcal{K}}}\|\varphi(\hat{P}_i)-\varphi(P_{i0})\|_2^T. \end{aligned} \tag{A.2.13}$$

As discussed in the proof of Lemma A.2.1, the set $S_{\mathcal{K}}$ has finite entropy in the sup-norm, and is thus totally bounded in the sup-norm. Its closure $\overline{S_{\mathcal{K}}}$ is therefore a compact set of continuous functions defined on $\mathcal{K}$. Since $\hat{P}_i \longrightarrow P_{i0}$ in probability, by the continuous mapping theorem, the sequence $(\varphi(\hat{P}_i)-\varphi(P_{i0}))$ converges pointwise (i.e., for each φ) to 0 over $\overline{S_{\mathcal{K}}}$. Note also that $\overline{S_{\mathcal{K}}}$ is equicontinuous by the Arzela-Ascoli theorem. Since $\overline{S_{\mathcal{K}}}$ is compact, the sequence also converges uniformly in $\overline{S_{\mathcal{K}}}$, implying,

$$\sup_{\varphi\in S_{\mathcal{K}}}\|\varphi(\hat{P}_i)-\varphi(P_{i0})\|_2^T \xrightarrow{p} 0$$

as $n\longrightarrow\infty$. This in turn implies that the last term on the RHS of equation (A.2.13) goes to zero in probability. Then from equation (A.2.12) we have that

$$0 \leq S_0(\hat{\beta},\hat{\varphi},P_0) \leq o_p(1). \tag{A.2.14}$$

Since the model is identified, for all $\delta>0$ there exists $\varepsilon>0$ such that

$$d[(\beta,\varphi),(\beta_0,\varphi_0)]>\delta \Rightarrow S_0(\beta,\varphi,P_0)>\varepsilon.$$

So we have that

$$\Pr\{d[(\hat{\beta},\hat{\varphi}),(\beta_0,\varphi_0)]>\delta\} \leq \Pr\{(S_0(\hat{\beta},\hat{\varphi},P_0)>\varepsilon\} \longrightarrow 0,$$

where the convergence comes from equation (A.2.14). This proves that $\hat{\beta} \xrightarrow{p} \beta$ and $\sup_{P\in\mathcal{K}}|\hat{\varphi}(P)-\varphi_0(P)| \xrightarrow{p} 0.$ □

A.2.4 Proof of Theorem 3.6.4

Proof. The proof of asymptotic normality of $\hat{\beta}$ relies heavily on the results in Newey and McFadden [1994]. By consistency results, we have $\hat{\varphi} = \varphi_0 + o_P(1)$. The consistency of $\hat{\beta}$ also implies $\hat{\beta} = \beta_0 + o_P(1)$. Hence, we have

$$\begin{aligned} \Delta x_i \hat{\beta} &= \Delta[\hat{\varphi}(\hat{P}_i)] \\ &= \Delta[\varphi_0(\hat{P}_i)] + (\Delta[\hat{\varphi}(\hat{P}_i)] - \Delta[\varphi_0(\hat{P}_i)]) \\ &= \Delta[\varphi_0(\hat{P}_i)] + o_P(1). \end{aligned} \tag{A.2.15}$$

Identification of the model and the above implies the following equality

$$\begin{aligned} \Delta x_i(\hat{\beta} - \beta_0) &= \Delta[\varphi_0(\hat{P}_i)] - \Delta[\varphi_0(P_{i0})] + o_P(1) \\ \Delta x_i(\hat{\beta} - \beta_0) &= R_i(\hat{P}_i - P_{i0}) + o_P(1), \\ \Delta x_i' \Delta x_i(\hat{\beta} - \beta_0) &= \Delta x_i' R_i(\hat{P}_i - P_{i0}) + o_P(1), \\ \Delta x_i' \Delta x_i(\hat{\beta} - \beta_0) &= \Delta x_i' R_i(\hat{P}_i - P_{i0}) + o_P(1), \\ \sqrt{N}\left(\frac{\sum_{i=1}^N \Delta x_i' \Delta x_i}{N}\right)(\hat{\beta} - \beta_0) &= \frac{1}{\sqrt{N}}\sum_{i=1}^N \Delta x_i' R_i(\hat{P}_i - P_{i0}) + o_P(1/\sqrt{N}) \end{aligned} \tag{A.2.16}$$

The second equality is due to the mean value theorem, where R_i is as in the statement of the theorem, except that the components $\overline{P}_{it}$, are between $\hat{P}_{it}$ and P_{it0}. Linearizing $\hat{P}_{it}$ around P_{it0} and stacking in t gives:

$$\begin{aligned} \sqrt{N}\left(\frac{\sum_{i=1}^N \Delta x_i' \Delta x_i}{N}\right)(\hat{\beta} - \beta_0) &= \frac{1}{\sqrt{N}}\sum_{i=1}^N \Delta x_i' R_i \underline{f}^{-1}(w_i) G_i[\hat{\gamma}(w_i) - \gamma_0(w_i)] \\ &\quad + C\sqrt{N}\|\hat{\gamma}(w_i) - \gamma_0(w_i)\|^2 + o_P(\frac{1}{\sqrt{N}}) \end{aligned} \tag{A.2.17}$$

where

$$\underline{f}^{-1}(w_i) = \begin{pmatrix} f^{-1}(w_{i1}) & 0 & \cdots & 0 \\ 0 & f^{-1}(w_{i2}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & f^{-1}(w_{iT}) \end{pmatrix}$$

$$G_i = \begin{pmatrix} -P_{i10} & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -P_{i20} & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -P_{iT0} & 1 \end{pmatrix}$$

$$[\hat{\gamma}(w_i) - \gamma_0(w_i)] = \begin{pmatrix} \hat{\gamma}_1(w_{i1}) - \gamma_{10}(w_{i1}) \\ \hat{\gamma}_2(w_{i1}) - \gamma_{20}(w_{i1}) \\ \vdots \\ \hat{\gamma}_1(w_{iT}) - \gamma_{10}(w_{iT}) \\ \hat{\gamma}_2(w_{i1}) - \gamma_{20}(w_{i1}) \end{pmatrix}$$

For notational convenience, define the $T \times 2T$ matrix $M_i := \Delta X_i' R_i \underline{f}^{-1}(w_i) G_i$. As discussed in equation (3.6.1), the second to term on the RHS of equation (A.2.17) converges to zero in probability. Then from equation (A.2.17) we have:

$$\begin{aligned}
\sqrt{N}\left(\frac{\sum_{i=1}^{N}\Delta x_i'\Delta x_i}{N}\right)(\hat{\beta}-\beta_0) \quad &= \quad \sqrt{N}\int M(w)[\hat{\gamma}(w)-\gamma_0(w)]\,f(w)\mathrm{d}w \\
&\quad + \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} M(w_i)[\hat{\gamma}(w_i)-\gamma_0(w_i)]\right. \\
&\quad \left. -\sqrt{N}\int M(w)[\hat{\gamma}(w)-\gamma_0(w)]\,f(w)\mathrm{d}w\right] \\
&\quad + o_P(\frac{1}{\sqrt{N}})
\end{aligned} \tag{A.2.18}$$

Equation (3.6.1) along with the triangle inequality results in the term in brackets on the RHS of equation being an $o_P(1/\sqrt{N})$. As for the first term on the RHS:

$$\begin{aligned}
\int M(w)[\hat{\gamma}(w)-\gamma_0(w)]\,f(w)\mathrm{d}w \quad &= \quad N^{-1}\sum_{i=1}^{N}\int M(w)q_i J_\sigma(w-w_i)\,f(w)\mathrm{d}w \qquad \text{(A.2.19)} \\
&\quad - \int M(w)\gamma_0\,f(w)\mathrm{d}w \\
&= \quad N^{-1}\sum_{i=1}^{N}\int [\,f(w)M(w)q_i - E[\,f(w)M(w)q]]\,J_\sigma(w-w_i)\mathrm{d}w.
\end{aligned}$$

As discussed in Newey and McFadden [1994], the conditions in assumption 3.6 ensures that this integral is close to the empirical measure. This with equation (A.2.18) implies that

$$\begin{aligned}
\sqrt{N}\left(\frac{\sum_{i=1}^{N}\Delta x_i'\Delta x_i}{N}\right)(\hat{\beta}-\beta_0) \quad &= \quad \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\,f(w)M(w)q_i - E[\,f(w)M(w)q]] \\
&\quad + o_P(\frac{1}{\sqrt{N}})
\end{aligned} \tag{A.2.20}$$

Substituting for $M(w)$ and observing that by the law of iterated expectations, the term on the RHS in expectations is zero, we have the following:

$$\begin{aligned}
\sqrt{N}\left(\frac{\sum_{i=1}^{N}\Delta x_i'\Delta x_i}{N}\right)(\hat{\beta}-\beta_0) \quad &= \quad \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta x_i' R_i (y_i - P_{i0}) + o_p(\frac{1}{\sqrt{N}}) \\
&= \quad \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta x_i' R_i \varepsilon_i + o_P(\frac{1}{\sqrt{N}})
\end{aligned} \tag{A.2.21}$$

The result then follows immediately from the Slutsky theorem combined with the WLLN and a multivariate version of the Linberg-Levy CLT. □

A.2.5 Proof of Theorem 3.6.5

Proof. The proof of efficiency uses the results developed in Newey [1994]. To proceed, we first set up the environment so that the results are directly applicable.

Notice that the model construction in section 3.2 implies the following equivalent moment condition

$$E_Q[\Delta x_i'(\Delta[\varphi_0(P_{i0})] - \Delta x_i\beta_0)] = E_Q[\tilde{m}(x_i, \beta, \varphi, P_{i0})] = 0.$$

This moment condition can be seen as the first order condition of $S(x_i, \beta, \varphi, P_i)$ with respect to beta. Furthermore, the limit of our estimate $\hat{\varphi}$ maximizes $E_Q[S(x_i, \beta, \varphi, P_i)]$. Thus by proposition 2 of Newey [1994], the estimation of φ can be ignored in calculating the asymptotic variance. So we work only with $\varphi = \varphi_0$.

Let the distribution Q belong to a general family of distributions $\mathcal{Q}$. Define the parametric submodel $\mathcal{Q}(\theta) := \{Q_\theta : Q_\theta \in \mathcal{Q},\ Q_\theta = Q_0 \text{ at } \theta = 0\}$. We assume f_θ to be a probability density relative to a fixed measure μ, the map $\theta \mapsto \sqrt{f_\theta(w)}$ is continuously differentiable in a neighborhood of 0, and $\theta \mapsto \int \left[(\partial f_\theta/\partial\theta)^2/f_\theta\right] \mathrm{d}\mu$ is finite and continuous in this neighborhood. Then by Lemma 1.9 of van der Vaart [1998], $\theta \mapsto Q_\theta$ is a differentiable path. We use this differentiable path to induce parametric submodels for the parameters that $\hat{\beta}$ and $\hat{P}_i$ are estimating. That is, we define $\mu(\theta) = \mu(Q_\theta) := \text{plim}\ \hat{\beta}$ and $P_i(\theta) = P_i(Q_\theta) := \text{plim}\ \hat{P}_i$, where $\mu(Q_\theta)$ satisfies:

$$E_\theta[\tilde{m}(x, \mu, P(\theta))] = 0 \tag{A.2.22}$$

The rest of the proof involves finding the pathwise derivative $d(w)$ satisfying $\frac{\partial\mu(\theta)}{\partial\theta} = E[d(w)g(w)]$, where $g(w) := \frac{\partial}{\partial\theta_{|\theta=0}} \ln f_\theta(w)$ is the corresponding score. Then the variance bound for the estimation of $\mu(\theta)$ is $\text{Var}(d(w))$. Differentiating equation (A.2.22) with respect to θ and solving for $\frac{\partial\mu(\theta)}{\partial\theta}$ gives

$$\frac{\partial\mu(\theta)}{\partial\theta} = -M^{-1}\left\{E\left[\frac{\partial}{\partial P}\tilde{m}(x, \beta_0, P(\theta))\frac{\partial P(w,\theta)}{\partial\theta}\right] + \frac{\partial}{\partial\theta}E_\theta\left[\tilde{m}(x, \beta_0, P_0)\right]\right\}, \tag{A.2.23}$$

where $M := \frac{\partial}{\partial\beta}E[\tilde{m}(x, \beta_0 P_0)] = E[\Delta x'\Delta x]$, which is invertible by assumption (3.3.1.3). From equation (A.2.22), the last term on the RHS of equation (A.2.23) is zero. Defining $\delta(x) := \frac{\partial}{\partial P}\tilde{m}(x, \beta_0, P(\theta))$ and applying the law of iterated expectations to $P(w,\theta) = E[y|w]$ gives

$$\begin{aligned} \frac{\partial\mu(\theta)}{\partial\theta} &= -M^{-1}\left\{\frac{\partial}{\partial\theta}E_\theta[\delta(w)(y - P_0(w))]\right\} \\ &= [-(M^{-1}\delta(w)(y - P_0))S(w)] \end{aligned} \tag{A.2.24}$$

Thus giving $d(w) = -M^{-1}\delta(w)(y - P_0)$. Noting that $\delta(w_i) = \Delta x_i' R_i$, we have that

$$\text{Var}(d(w)) = E[\Delta x'\Delta x]^{-1}E[\Delta x' R]\Omega E[R'\Delta x]E[\Delta x'\Delta x]^{-1} \tag{A.2.25}$$

which is the asymptotic variance of $\hat{\beta}$ derived in theorem 3.6.4. □

A.2.6 Contraction Mapping

Proof. Here we show that equation (3.5.4) indeed defines a contraction mapping. For simplicity we drop the s subscript. Recall that the usual kernel smoother is indeed a projection (see Mammen et al., 2001). We therefore write equation (3.5.4) as:

$$\varphi_{j+1}(P) = Pr_{P_t}\Delta x\hat{\beta} + Pr_{P_t}\varphi_j(\hat{P}_{t-1}). \tag{A.2.26}$$

Taking differences gives

$$\begin{aligned} \varphi_{j+1}(P) - \varphi_j(P) &= Pr_{P_t}\varphi_j(\hat{P}_{t-1}) - Pr_{P_t}\varphi_{j-1}(\hat{P}_{t-1}). && \text{(A.2.27)} \\ &= Pr_{P_t}\left(\varphi_j(\hat{P}_{t-1}) - \varphi_{j-1}(\hat{P}_{t-1})\right). && \text{(A.2.28)} \end{aligned}$$

Computing these projections at $\hat{P}_{t-1}$ and norming both sides of this equation gives the inequality

$$\|\varphi_{j+1}(\hat{P}_{t-1}) - \varphi_j(\hat{P}_{t-1})\| < \|\varphi_j(\hat{P}_{t-1}) - \varphi_{j-1}(\hat{P}_{t-1})\|, \tag{A.2.29}$$

since the projection is a contraction mapping. □

A.3 APPENDIX TO CHAPTER 4

We prove here two lemmas which are used in the paper to evaluate Taylor Series expansions for composite and inverse functions. Lemma 1 is taken from MMRS but is included here for the ease of reference.

Lemma A.3.1. *Let*

$$f(u) = \sum_{j=0}^{\infty} f_j (u-u_0)^j, \quad g(t) = \sum_{j=0}^{\infty} g_j (t-t_0)^j, \tag{A1}$$

together with $u_0 = g(t_0)$. *Then*

$$(f \circ g)(t) = \sum_{j=0}^{\infty} a_j (t-t_0)^j \tag{A2}$$

where $a_0 = f_0$ *and for* $j \geq 1$

$$a_j = \sum_{k=1}^{j} f_k \theta_{k,j} \tag{A3}$$

and where the θ*s are evaluated recursively as follows*

$$\theta_{k,j} = \sum_{s=1}^{j-k+1} g_s \theta_{k-1,j-s}, \quad 1 \leq k \leq j \tag{A4}$$

with $\theta_{0,0} = 1$.

Proof: We have

$$(f \circ g)(t) = \sum_{k=0}^{\infty} f_k \left[\sum_{s=1}^{\infty} g_s (t-t_0)^s \right]^k \tag{A5}$$

Whence a_j is given by formula (A3) where $\theta_{k,j}$ denotes the coefficient of $(t-t_0)^j$ in the k-th power of the factor in brackets. Formula (A4) follows from the identity

$$\sum_{j=k}^{\infty} \theta_{k,j} (t-t_0)^j = \left[\sum_{r=k-1}^{\infty} \theta_{k-1,r} (t-t_0)^r \right] \cdot \left[\sum_{s=1}^{\infty} g_s (t-t_0)^s \right] \tag{A6}$$

Lemma A.3.2. *Let* f^{-1} *denote the inverse of* f

$$f^{-1}(x) = \sum_{j=0}^{\infty} h_j (x-x_0)^j \tag{A7}$$

with $x_0 = f(t_0)$. *Then*

$$h_0 = x_0, \quad h_1 = f_1^{-1} \tag{A8}$$

$$h_j = -f_1^{-j} \cdot \left[\sum_{k=1}^{j-1} h_k \theta_{k,j} \right] \tag{A9}$$

Proof: We apply Lemma 1 together with $g = f^{-1}$, whence

$$(g^{-1} \circ g)(t) = t = t_0 + (t-t_0)$$

This implies that $a_0 = a_1 = 1$ and $a_j = 0$ for $j > 1$ in Formula (A3). The proof follows from formulae (A3) and (A4), with the latter implying that $\theta_{j,j} = f_1^g$.