WORKING PAPER #435
INDUSTRIAL RELATIONS SECTION
PRINCETON UNIVERSITY
MARCH, 2000
http://www.irs.princeton.edu/pubs/working_papers.html

Estimating a Censored Dynamic Panel Data Model with an Application to Earnings Dynamics

Luojia Hu
Department of Economics
Princeton University

January, 2000

Estimating a Censored Dynamic Panel Data Model with an Application to Earnings Dynamics*

Luojia Hu

Department of Economics

Princeton University

luojiahu@princeton.edu

January 3, 2000

Abstract

This paper proposes a method for estimating a censored panel data model with a lagged *latent* dependent variable and individual-specific fixed effects. The main insight is to trim observations in such a way that a certain symmetry, which was destroyed by censoring, is restored. Based on the restored symmetry, orthogonality conditions are constructed and GMM estimation is implemented. The estimation method is used to study earnings dynamics, using matched data from the Current Population Survey and Social Security Administration (CPS-SSA) Earnings Record for a sample of men who were born in 1930-39 and living in the South during the period of 1957-73. The SSA earnings are top-coded at the maximum social security taxable level. Although linear GMM estimation yields no difference in earnings dynamics by race, the earnings process for white men appears to be more persistent than that for black men (conditional on individual heterogeneity) after censoring is taken into account.

KEYWORDS: Panel data, censored regression, earnings dynamics.

*I am indebted to Bo Honoré for his tireless advice and encouragement. I have also benefited from comments and suggestions from Vivi Alatas, Orley Ashenfelter, Marianne Bertrand, Anne Case, Ken Chay, Gregory Chow, Angus Deaton, Hank Farber, Han Hong, Jeff Kling, Kristin Mammen, Doug Miller, Ceci Rouse, Chris Sims, Elie Tamer, Mark Watson, and participants at the Princeton Econometrics seminar. All errors are my own responsibility. Financial support from the Woodrow Wilson Society of Fellows is gratefully acknowledged. Mailing Address: Department of Economics, Princeton University, Princeton, NJ 08544-1021.

1 Introduction

Dynamic panel data models provide a broad framework for analyzing economic relationships. By allowing for individual heterogeneity and dynamic feedback, such a model can help us better understand the complex relationship between economic outcomes. In addition, the model incorporates information from both time series and cross-sectional dimensions and therefore enriches the identification possibilities. For example, we can combine the notions of stationarity, predeterminedness, and uncorrelated disturbances from a time series model, and random sampling and exogenous instrumental variables techniques from a cross-sectional model to achieve identification in models that are not otherwise identified. However, the richness of the model also leads to some difficulties in econometric analysis. To illustrate the issues, first consider a static panel data model

$$y_{it} = x_{it}\beta + \alpha_i + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T$$

where x is strictly exogenous in the sense that it is uncorrelated with past, present, and future values of ε , β is the parameter of interest common to all individuals, and α_i is the individual specific fixed effect. $\{\alpha_i\}$ is a sequence of nuisance parameters that control for individual heterogeneity. This model can be estimated using OLS regression (or normal maximum likelihood estimation (MLE)) by including N intercept dummy variables. In this case, the OLS estimator (or normal MLE) yields a consistent estimate of the slope parameter β despite the fact that the number of parameters α_i grows with sample size. This is because the Frisch-Waugh Theorem implies that OLS with N intercept dummy variables is numerically equivalent to OLS on deviations from means (i.e. differencing out the α_i 's).

Unfortunately, this method does not generally work once we move out of the static linear regression framework because, in general, for fixed T , the maximum likelihood estimator will not consistently estimate the slope parameters as N goes to infinity, because the number of parameters α_i grows with the number of observations and the classical incidental parameter problem is encountered (see Neyman and Scott (1948)). Specifically, consider a dynamic panel data model where the lagged dependent variable appears on the right hand side as an explanatory variable:

$$y_{it} = x_{it}\beta + \gamma y_{it-1} + \alpha_i + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 2, \dots, T$$

where x is assumed to be strictly exogenous. To see why the standard fixed effects (or within-group) estimator will be an inconsistent estimator of β and γ , consider the following deviations-from-means

regression:

$$y_{it} - y_{i.} = (x_{it} - x_{i.})\beta + \gamma(y_{it-1} - y_{i.-1}) + \varepsilon_{it} - \varepsilon_{i.}$$

where for any variable z_{it} , $z_{i.} = \frac{1}{T} \sum_{t=1}^T z_{it}$ and $z_{i.-1} = \frac{1}{T} \sum_{t=0}^{T-1} z_{it}$. The inconsistency comes from the correlation between $(y_{it-1} - y_{i.-1})$ and the error term $\varepsilon_{i.}$. Monte Carlo evidence from Nerlove (1967, 1971) and Maddala (1971) indicates that there are large asymptotic biases associated with this approach when T is small. Under stationarity assumptions, Nickell (1981) derived an analytical expression for the bias, which is on the order of $O(1/T)$.

Alternatively, we can estimate the model using OLS on first differences:

$$y_{it} - y_{it-1} = (x_{it} - x_{it-1})\beta + \gamma(y_{it-1} - y_{it-2}) + \varepsilon_{it} - \varepsilon_{it-1}.$$

This approach also suffers the asymptotic bias because y_{it-1} and ε_{it-1} are correlated. However, we can appeal to the exogeneity of the variable x (or the predeterminedness of the lagged variable y) and apply instrumental variable estimation using x (or y_{it-j} , for $2 \leq j \leq T-1$), as the instrument. This gives us a consistent and asymptotically normal estimator (see, for example, Anderson and Hsiao (1982)).

Despite its simplicity and theoretical attractiveness, the linear dynamic panel data model cannot always completely satisfy our empirical needs, since empirical analysis often requires us to go beyond the linear framework.

In this paper I estimate a dynamic panel data model for the earnings process. Knowledge of the dynamics of the labor earnings (or income) process is important for our understanding of many economic relationships. For instance, at the aggregate macroeconomic level, the permanent income hypothesis of the relationship between consumption and income often relies on information about the time series features of the underlying income process (see Deaton (1992) and references therein). At the individual microeconomic level, knowing the dynamic features of the earnings process also provides additional information other than just the levels. For example, Farber and Gibbons (1996) studied wage dynamics to test a dynamic model of learning about worker ability in a competitive labor market. Ashenfelter (1978) explored the information in the dynamic earnings process to estimate the effect of training programs on earnings in a changing economy. One good data source for studying earnings dynamics is the matched data from the 1973 and 1978 March

Current Population Survey and Social Security Administration (CPS-SSA) Earnings Record.¹ The sample is nationally representative, covers relatively long histories of individuals' earnings and the earnings information is less subject to measurement error than self-reported data. But the SSA earnings data also have a shortcoming: many records are censored at the social security maximum taxable earnings level. Specifically, the model I am interested in is

$$y_{it}^* = x_{it}\beta + \gamma y_{it-1}^* + \alpha_i + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 2, \dots, T$$

where y_{it}^* represents true earnings, but we only observe

$$y_{it} = \min(y_{it}^*, c_t)$$

where c_t is the tax ceiling in period t .

In this example, censoring destroys the linearity of the model and we cannot directly apply the “tricks” from before: it is not obvious how to difference away the individual fixed effects and how to use instrumental variable type techniques.

Our knowledge of general nonlinear dynamic panel data models is very limited, although much progress has been made in the linear panel data model (see Arellano and Honoré (1999) for review). In most nonlinear models, strict exogeneity of the explanatory variables is still the key assumption. Apart from the difficulties in estimation, interpretation of the “parameters of interest” in those nonlinear panel data models should also be dealt with cautiously. For example, in a discrete choice model, knowing the parameter β only tells us about the relative effects of explanatory variables on the probability distribution of the dependent variable, but does not allow us to “recover” the effect of a single variable on the choice probability. This is not an issue here as the model I consider is linear in the true earnings data, and the nonlinearity is induced only by the construction of the data (due to top-coding). Therefore, the estimated parameters in this model can be interpreted in the usual way as in a linear model.

The difficulties associated with the fixed effects approach naturally prompt us to ask whether a random effects approach provides a fruitful alternative for the nonlinear model. Unfortunately, the random effects approach has its own limitations. First, we need to make assumptions on

¹These data were used in Chay (1996) to evaluate the effects of the Civil Rights Act of 1964 on the black-white wage differentials.

the distribution of individual effects, which is typically unknown to econometricians. Second, the random effects approach has the initial conditions problem: unless we have data from the start of the process, we have to specify the relationship between the first observation in our sample and the individual specific effects. But this relationship depends on the model parameters and the distribution of the explanatory variables in periods prior to the start of the sample. Again, the distribution is seldom known and misspecification may yield biased estimates.²

Although, in general, it is very difficult to estimate the dynamic nonlinear panel data model, some progress has been made in applying the fixed effects approach to estimate the dynamic censored panel data model. Honoré (1993b) proposed some solutions for estimating censored regression panel data models with individual fixed effects and lagged *censored* dependent variables. Honoré and Hu (1999) provided identification results for this approach under certain conditions. In addition, they studied the estimation for panel data models with general predetermined explanatory variables. Unfortunately, the model studied in those papers is not applicable to the empirical analysis considered here. Specifically, while Honoré (1993b) and Honoré and Hu (1999) studied the model where the lagged *censored* variable enters as the explanatory variable, I consider a model where the lagged *latent* variable enters as the explanatory variable.³ Although the former may be sensible for certain economic applications (for example, the dynamic labor supply model where the number of hours an individual works in this period depends on his or her past participation), the latter makes more sense in the empirical example of estimating earnings dynamics since the true earnings process is linear and the nonlinearity is the result of the data construction.⁴

The main insight used in this paper is to estimate the censored dynamic panel data model by artificially truncating the observations to restore a certain symmetry of the latent variables, which was destroyed by the censoring. Then, based on the restored symmetry, orthogonality conditions can be constructed, and the generalized methods of moment (GMM) estimation can be applied. Asymptotic properties of GMM estimators and test-statistics are well-established if we assume

²In practice, one often deals with this problem by assuming some flexible functional form for the distribution of the first observation (see Heckman (1981) for example).

³Honoré (1993b) briefly mentioned the possible extension to the latter case, though he did not pursue this issue.

⁴Arellano and Bover (1997) and Arellano, Bover, and Labeaga (1997) adopted a random effects approach to estimate the model with lagged latent dependent variables.

the parameters are identified. For example, the properly scaled minimized criterion function provides a convenient test for the overidentifying restrictions. Unfortunately, in the model considered here, identification is problematic and, as a result, numerical convergence sometimes fails in the straightforward GMM estimation in the empirical application. I therefore adopt several alternative approaches to carry out the inferences and construct valid asymptotic confidence intervals for the autoregressive parameter γ .

I estimate the dynamic model separately for black and white men in the sample. The main empirical finding is that the linear GMM estimator yields similar estimates of the AR(1) parameter for black and white men, but when censoring is taken into account, a higher persistency (conditional on individual heterogeneity) is found in white men's earnings dynamics. This finding is robust to model specifications with and without cohort effects.

The rest of the paper is organized as follows. Section 2 presents the econometric model and illustrates the estimation method. Section 3 describes the data. Section 4 presents the empirical model to be used in the application. Section 5 presents the linear GMM estimation results and discusses the nonlinear GMM estimation. Section 6 discusses several alternative ways to construct confidence intervals for the AR(1) parameter in the nonlinear GMM model. Section 7 presents some further results from an unbalanced panel and with a more complete model specification. Finally section 8 concludes. The mathematical details are in the appendix.

2 Model and Estimation

In this section, I propose a method for estimating a censored dynamic panel data model. Specifically, I show how censoring destroys a certain symmetry between the latent variables, and I demonstrate how to artificially truncate observations to restore the symmetry.⁵ I then present orthogonality conditions based on the restored symmetry, which must be satisfied at the true parameter values.

The underlying model is a linear panel data model

$$y_{it}^* = \gamma_0 y_{it-1}^* + x_{it} \beta_0 + \alpha_i + \varepsilon_{it} \quad \text{for } t = 1, 2, \dots, T, \quad T \geq 2 \quad (1)$$

where y_{it}^* is the latent variable of interest, x_{it} is a vector of explanatory variables, α_i is an individual specific fixed effect, and $\{\varepsilon_{it}\}$ is a sequence of i.i.d. random variables conditional on

⁵This approach is in the same spirit as in Powell (1986).

$(y_{i0}^*, \{x_{it}\}_{t=1}^T, \alpha)$. Therefore, the dynamic feature of the model is the presence of the lagged dependent variable, and not serial correlation in the error.⁶ In this model, x_{it} is strictly exogenous in the sense that it is uncorrelated with past, present, and future values of the disturbance ε . The individual fixed effect $\{\alpha_i\}$ is a sequence of nuisance parameters. The distribution of α_i and its relationship with x are left unspecified. The parameters of interest are β_0 and γ_0 . The autoregressive parameter γ_0 is one of the key determinants of the persistence of the process conditional on individual heterogeneity, and will be the primary object of our empirical application.

Because of censoring, we only observe

$$y_{it} = \max(0, y_{it}^*).$$

Here, without loss of generality, I assume that the data are censored from below by zero. If the “floor” is nonzero and possibly time varying, such that $y_{it} = \max(c_t, y_{it}^*)$, then one can incorporate c_t in the x_{it} . The case for censoring from above can be dealt with in the same fashion by changing sign and switching *min* to *max*.

The comparison and trimming of observations below are between 2 consecutive periods. For ease of exposition, consider the case $T = 2$. The model then becomes

$$\begin{aligned} y_{i1}^* &= \gamma_0 y_{i0}^* + x_{i1} \beta_0 + \alpha_i + \varepsilon_{i1} \\ y_{i2}^* &= \gamma_0 y_{i1}^* + x_{i2} \beta_0 + \alpha_i + \varepsilon_{i2}. \end{aligned}$$

If ε_{i1} and ε_{i2} are independent, and identically distributed over time conditional on $\{y_{i0}^*, x_{i1}, x_{i2}, \alpha_i\}$, then $(\varepsilon_{i1}, \varepsilon_{i2})'$ is symmetrically distributed around zero conditional on the same variables. It then follows immediately that $(y_{i1}^*, y_{i2}^* - \gamma y_{i1}^*)$ is symmetrically distributed around the 45° line through $(\gamma_0 y_{i0}^* + x_{i1} \beta_0 + \alpha_i, x_{i2} \beta_0 + \alpha_i)$. Hence conditional on $\{y_{i0}^*, x_{i1}, x_{i2}\}$, $(y_{i1}^*, y_{i2}^* - \gamma y_{i1}^*)$ is symmetrically distributed around the 45° line through $(\Delta x_i \beta_0, 0)$ where $\Delta x_i \beta_0 = (x_{i1} - x_{i2}) \beta_0 + \gamma_0 y_{i0}^*$.

This symmetry is the key feature to be exploited in the construction of orthogonality conditions. As will be seen shortly, censoring destroys the symmetry, and the main insight of this paper is to

⁶These two types of models are often referred to as the “state dependence model” and the “serial correlation model,” respectively, in the literature. Alternatively, they are sometimes classified as models with “system dynamics” and models with “error dynamics,” respectively (see, for example, Maddala (1987)). These are different models and have different implications. In this paper, I focus on the former.

restore the symmetry by artificial truncation (or “trimming”). The idea of symmetric trimming has been adopted in previous literature to estimate censored and truncated regression models in both cross sectional and panel data settings. Examples include Powell (1986), Honoré (1992, 1993b), and Honoré and Powell (1994).⁷

Due to censoring, we only observe

$$\begin{aligned} y_{i1} &= \max(0, y_{i1}^*) = \max(0, \gamma_0 y_{i0}^* + x_{i1} \beta_0 + \alpha_i + \varepsilon_{i1}) \\ y_{i2} &= \max(0, y_{i2}^*) = \max(0, \gamma_0 y_{i1}^* + x_{i2} \beta_0 + \alpha_i + \varepsilon_{i2}). \end{aligned} \quad (2)$$

It should be pointed out that, in the current formulation, it is necessary to observe y_{i0}^* and y_{i1}^* (or equivalently $y_{i0} > 0$ and $y_{i1} > 0$). This is because both of them are explanatory variables in the model, and observations of them are needed in the construction of orthogonality conditions. Therefore, the following analysis is all conditioned on $y_{i0} = y_{i0}^* > 0$ and $y_{i1} = y_{i1}^* > 0$.

To see why censoring creates asymmetry, I construct⁸

$$\tilde{y}(\gamma) = y_2 - \gamma y_1.$$

At $\gamma = \gamma_0$,

$$\begin{aligned} \tilde{y}(\gamma_0) &= \max(0, \gamma_0 y_1 + x_2 \beta_0 + \alpha + \varepsilon_2) - \gamma_0 y_1 \\ &= \max(-\gamma_0 y_1, x_2 \beta_0 + \alpha + \varepsilon_2) \\ &= \max(-\gamma_0 y_1, y_2^* - \gamma_0 y_1^*). \end{aligned} \quad (3)$$

Comparing equation (2) and (3), we see that the pair $(y_1, \tilde{y}(\gamma_0))$ is similar to $(y_1^*, y_2^* - \gamma_0 y_1^*)$ except that the first element y_1 is y_1^* censored from below by zero, while the latter $\tilde{y}(\gamma_0)$ is $y_2^* - \gamma_0 y_1^*$ censored from below by $-\gamma_0 y_1$ (conditional on $y_1 > 0$). Hence censoring destroys the original symmetry between y_1^* and $y_2^* - \gamma_0 y_1^*$ because of the two different censoring points.

The key insight in the papers cited above is to restore the symmetry by “trimming” the observations in a way that depends on the exact model under consideration. For the present model, it is perhaps easiest to illustrate the idea graphically. Assume for the moment that the true parameter

⁷Honoré, Kyriazidou, and Udry (1997) considered a Type 3 Tobit model using the same technique.

⁸For ease of notation, I drop the subscript i in the rest of this section.

values β_0 and γ_0 are known and that $\gamma_0 \geq 0$ (the latter assumption is maintained throughout). Depending on the sign of $\Delta x\beta_0$, I consider two separate cases in Figure 1 and Figure 2 below.

Figure 1 illustrates the case $\Delta x\beta_0 \geq 0$. The horizontal axis represents the variable y_1 and the vertical axis represents the constructed variable $\tilde{y}(\gamma_0)$. The vector of latent variables $(y_1^*, y_2^* - \gamma_0 y_1^*)$ is symmetrically distributed around line (1), given by $\tilde{y}(\gamma_0) = y_1 - \Delta x\beta_0$, which is the 45° line through $(\Delta x\beta_0, 0)$. However, instead of observing y_1^* and $y_2^* - \gamma_0 y_1^*$, we observe $y_1 = \max(0, y_1^*)$ and $\tilde{y}(\gamma_0) = \max(-\gamma_0 y_1, y_2^* - \gamma_0 y_1^*)$. Therefore, observations are censored from the left at the vertical axis and, for any $y_1 > 0$, the observations are also censored from below at line (2), given by $\tilde{y}(\gamma_0) = -\gamma_0 y_1$. The difference between this model and Honoré (1993b) is that observations on the vertical axis cannot be used to construct orthogonality conditions. This is because one needs to observe the latent variable y_1^* , which is not possible for the observations on the vertical axis. So we have to truncate the observations at the vertical axis. This amounts to throwing away the observations on the line. We can then restore the symmetry by *symmetric truncating*.

Figure 1: $\gamma_0 \geq 0$ and $\Delta x\beta_0 \geq 0$

The trick is to find the mirror images of the two borderlines – the vertical axis and line (2) – around the center line (1), and then symmetrically truncate observations by those two new lines as well. The mirror image of the vertical axis around line (1) is represented by line (3) and given by

$\tilde{y}(\gamma_0) = -\Delta x \beta_0$. The mirror image of line (2) around line (1) is represented by line (4) and given by $\tilde{y}(\gamma_0) = -\frac{1}{\gamma_0} (y_1 - \Delta x \beta_0) - \Delta x \beta_0$. So one should also truncate observations at line (3) from below and at line (4) from the left. The final observations are those falling in the interior of region $A \cup B$ bordered by the 4 bold line segments. Note that

$$\begin{aligned} A \cup B &\equiv \{(y_1, \tilde{y}(\gamma_0)) : y_1 > 0, \tilde{y}(\gamma_0) > -\gamma_0 y_1, y_1 > \Delta x \beta_0 - \gamma_0 (\tilde{y}(\gamma_0) + \Delta x \beta_0), \tilde{y}(\gamma_0) > -\Delta x \beta_0\} \\ &= \{(y_1, \tilde{y}(\gamma_0)) : y_1 > 0, y_2 > 0, y_1 > \Delta x \beta_0 - \gamma_0 (\tilde{y}(\gamma_0) + \Delta x \beta_0), \tilde{y}(\gamma_0) > -\Delta x \beta_0\}. \end{aligned}$$

After the truncation, $(y_1, \tilde{y}(\gamma_0))$ is symmetrically distributed around the 45° line through $(\Delta x \beta_0, 0)$. This symmetry has many implications which can be used to construct orthogonality conditions.

One of the immediate implications of the symmetry is that, conditional on being in region $A \cup B$, the probability of an observation falling in region A equals the probability of it falling in region B . That is

$$E[1\{(y_1, \tilde{y}(\gamma_0)) \in A \cup B\} \cdot [1\{y_1 - \tilde{y}(\gamma_0) - \Delta x \beta_0 > 0\} - 1\{y_1 - \tilde{y}(\gamma_0) - \Delta x \beta_0 < 0\}]] = 0. \quad (4)$$

Another implication is that the expected vertical distance from $(y_1, \tilde{y}(\gamma_0))$ in region A to the line (1) equals the expected horizontal distance from $(y_1, \tilde{y}(\gamma_0))$ in region B to that line. For $(y_1, \tilde{y}(\gamma_0))$ in region A , the vertical distance to line (1) can be written as $-(y_1 - \tilde{y}(\gamma_0) - \Delta x \beta_0)$ whereas for $(y_1, \tilde{y}(\gamma_0))$ in region B , the horizontal distance to line (1) can be written as $y_1 - \tilde{y}(\gamma_0) - \Delta x \beta_0$. Therefore

$$E[1\{(y_1, \tilde{y}(\gamma_0)) \in A \cup B\} \cdot (y_1 - \tilde{y}(\gamma_0) - \Delta x \beta_0)] = 0. \quad (5)$$

More generally, for any function $\xi(\cdot, \cdot)$ satisfying $\xi(e_1, e_2) = -\xi(e_2, e_1)$ for all (e_1, e_2) , we have that $\xi(y_1 - \Delta x \beta_0, \tilde{y}(\gamma_0))$ is symmetrically distributed around zero conditional on y_0^*, x_1, x_2 , and $(y_1, \tilde{y}(\gamma_0)) \in A \cup B$. This gives orthogonality condition

$$E[1\{(y_1, \tilde{y}(\gamma_0)) \in A \cup B\} \cdot \xi(y_1 - \Delta x \beta_0, \tilde{y}(\gamma_0)) \cdot h(y_0, x_1, x_2)] = 0$$

for any function h . If one chooses h to be a constant, the case $\xi(e_1, e_2) = \text{sgn}(e_1 - e_2)$ corresponds to equation (4) and the case $\xi(e_1, e_2) = e_1 - e_2$ corresponds to equation (5).

It should be noted that this estimation method uses three consecutive uncensored observations ($y_{it} > 0, t = 0, 1, 2$). Two are not sufficient; this is because y_0^* is one of the explanatory variables.

The case $\Delta x\beta_0 < 0$ is presented in Figure 2. I proceed in a similar fashion. As before, observations are truncated at the vertical axis from the left and at line (2) from below. The mirror image of the vertical axis around line (1) is given by line (3). Therefore one should also truncate observations at line (3) from below. In this case, this is sufficient to restore the symmetry. Similarly denote

$$A \cup B \equiv \{(y_1, \tilde{y}(\gamma_0)) : y_1 > 0, \tilde{y}(\gamma_0) > -\Delta x\beta_0\}.$$

Again after truncation, observations in the interior of region $A \cup B$ bordered by the bold lines are symmetrically distributed around the 45° line through $(0, -\Delta x\beta_0)$. Note $\Delta x\beta_0 < 0$, thus $\tilde{y}(\gamma_0) > -\Delta x\beta_0$ implies $y_2 > 0$ as well.

Figure 2: $\gamma_0 \geq 0$ and $\Delta x\beta_0 < 0$

Therefore, conditional on $\{y_{it} > 0, t = 0, 1, 2; \tilde{y}(\gamma_0) > -\Delta x\beta_0\}$, y_1 and $\tilde{y}(\gamma_0)$ are symmetrically distributed around the 45° line through $(\Delta x\beta_0, 0)$. Again, it follows that for any function $\xi(.,.)$ satisfying $\xi(e_1, e_2) = -\xi(e_2, e_1)$ for all (e_1, e_2) , $\xi(y_1 - \Delta x\beta_0, \tilde{y}(\gamma_0))$ is symmetrically distributed around zero conditional on y_0^* , x_1 and x_2 . This gives the orthogonality condition

$$E[1\{(y_1, \tilde{y}(\gamma_0)) \in A \cup B\} \cdot \xi(y_1 - \Delta x\beta_0, \tilde{y}(\gamma_0)) \cdot h(y_0, x_1, x_2)] = 0$$

for any function h .

The following proposition summarizes the main results in this section.

Proposition 1 *Assume that y_1 and y_2 are generated from model (1) with y_0^* observed (i.e. $y_0 > 0$) and that ε_1 and ε_2 are i.i.d. conditional on $(y_0^*, x_1, x_2, \alpha)$. Assume $\gamma_0 \geq 0$. For observations falling in the interior of the region $A \cup B$, the following results hold at the true parameter values β_0 and γ_0 :*

(1) $(y_1, \tilde{y}(\gamma_0))$ is symmetrically distributed around the 45° line through $(\Delta x \beta_0, 0)$ conditional on (y_0^*, x_1, x_2) .

(2) For any function $\xi(.,.)$ satisfying $\xi(e_1, e_2) = -\xi(e_2, e_1)$ for all (e_1, e_2) , $\xi(y_1 - \Delta x \beta_0, \tilde{y}(\gamma_0))$ is symmetrically distributed around zero conditional on y_0^*, x_1 , and x_2 .

The proposition suggests estimating the parameter (β_0, γ_0) based on the orthogonality conditions

$$E[1\{(y_1, \tilde{y}(\gamma_0)) \in A \cup B\} \cdot \xi(y_1 - \Delta x \beta_0, \tilde{y}(\gamma_0)) \cdot h(y_0, x_1, x_2)] = 0$$

where

$$\begin{aligned} 1\{(y_1, \tilde{y}(\gamma_0)) \in A \cup B\} &\equiv 1\{y_0 > 0, y_1 > 0, y_2 > 0\} \\ &\cdot [1\{\Delta x \beta_0 \geq 0\} \cdot 1\{y_1 > \Delta x \beta_0 - \gamma_0(\tilde{y}(\gamma_0) + \Delta x \beta_0)\} \cdot 1\{\tilde{y}(\gamma_0) > -\Delta x \beta_0\} \\ &+ 1\{\Delta x \beta_0 < 0\} \cdot 1\{\tilde{y}(\gamma_0) > -\Delta x \beta_0\}] \end{aligned}$$

Two natural choices of the ξ function are $\xi(e_1, e_2) = e_1 - e_2$ and $\xi(e_1, e_2) = \text{sgn}(e_1 - e_2)$.

Remark 1 *This approach uses consecutive uncensored observations and the minimum number of uncensored observations per individual is three.*

If the panel has at least 4 time periods, one can also consider any pair of observations y_t and y_s with $y_{t-1} > 0$, $y_t > 0$, $y_{s-1} > 0$, and $y_s > 0$. Define

$$\begin{aligned} W_{ts}(\beta_0, \gamma_0) &= \max\{0, (x_t - x_s)\beta_0, y_t - \gamma_0 y_{t-1}\} - x_t \beta_0 \\ &= \max\{-x_t \beta_0, -x_s \beta_0, \alpha + \varepsilon_t\} \end{aligned}$$

$$\begin{aligned} W_{st}(\beta_0, \gamma_0) &= \max\{0, -(x_t - x_s)\beta_0, y_s - \gamma_0 y_{s-1}\} - x_s \beta_0 \\ &= \max\{-x_s \beta_0, -x_t \beta_0, \alpha + \varepsilon_s\}. \end{aligned}$$

It follows that if ε_t and ε_s are i.i.d. conditional on (x_t, x_s, α) , then $(W_{ts}(\beta_0, \gamma_0), W_{st}(\beta_0, \gamma_0))$ is distributed symmetrically around the 45° line conditional on (x_t, x_s) . Therefore for any function $\xi(\cdot, \cdot)$ satisfying $\xi(e_1, e_2) = -\xi(e_2, e_1)$ for all (e_1, e_2) , $\xi(W_{ts}(\beta_0, \gamma_0), W_{st}(\beta_0, \gamma_0))$ is symmetrically distributed around zero conditional on (x_t, x_s) . This suggests estimating the parameter (β_0, γ_0) based on the orthogonality condition

$$E[1\{y_{t-1} > 0, y_t > 0, y_{s-1} > 0, y_s > 0\} \cdot \xi(W_{ts}(\beta_0, \gamma_0), W_{st}(\beta_0, \gamma_0)) \cdot h(x_t, x_s)] = 0$$

for any function h .

Remark 2 *This approach uses non-consecutive uncensored observations and the minimum number of uncensored observations per individual is four.*

The reason we need at least four uncensored observations is that we need to know the true (latent) variables for the two pairs (y_t^*, y_{t-1}^*) and (y_s^*, y_{s-1}^*) to construct the orthogonality conditions. This approach requires weaker assumptions than the one based on three consecutive uncensored observations since the conditioning variables do not involve the initial variable y_0 . However, this approach also leads to more severe trimming. It is also worth noting that, in this approach, the i.i.d. assumption of ε (conditional on x and α) is stronger than necessary for the symmetry. One would obtain the same results if ε_t and ε_s are exchangeable (conditional on x and α). However, stationarity is not sufficient for the general odd function ξ other than $\xi(e_1, e_2) = g(e_1) - g(e_2)$.

3 Background and Data

Understanding the earnings process is an important task for the economist. The dynamic feature of labor earnings is an important element for any reasonable characterization of the underlying earnings process. It can provide richer and fuller information about the entire process than levels alone. Although there is a large literature on the black and white wage differential and its convergence since the late 1960s, little is known about the difference in earnings dynamics between blacks and whites. The difference in earnings dynamics could be interesting and important in its own right. If white men had jobs that had better security and were less vulnerable to economic fluctuations than black men, then a higher persistence in the earnings process would be expected from the data.

For example, Hoynes (1999) found that for a variety of labor market outcomes (including both employment and earnings), non-whites experience greater cyclical fluctuations than whites.⁹

I apply the proposed estimation method to study labor earnings dynamics using the matched data linking the 1973 and 1978 March Current Population Survey (CPS) to employer-reported longitudinal social security payroll tax records. In a joint project of the Census Bureau and the Social Security Administration (SSA), respondents of the March CPS were matched, by their social security numbers, to the histories of earnings for which they paid social security taxes. This data source has been exploited in the labor economics literature to study different issues. For example, Bound and Krueger (1991) combined the matched data in 1978 with the 1977 CPS to examine the properties and prevalence of measurement error in longitudinal earnings data. Chay (1996) used both 1973 and 1978 matched data to evaluate the effect of civil rights policies on the changes in black-white wage differentials during the 1960s.

The matched data provide a good source for studying labor earnings dynamics. First, the administrative earnings records are accurate and less subject to measurement error than self-reported earnings data found in most household surveys. Second, this data set covers a relatively long history of earnings information and is particularly suitable for analyzing the dynamic aspect of the earnings process.¹⁰ Finally, the sample is relatively large and nationally representative. However, there is also a shortcoming: a high percentage of the data are top-coded because of the maximum social security taxable earnings level, and that tax ceiling was changing over time (for any given year, the fraction of top-coding ranges from 40% to 60%). In addition, there are also some records that were censored at zero.¹¹ This may occur if the individual was earning income in an untaxed sector (e.g. federal, state, or local government) or if he or she was out of labor force or experiencing

⁹Another possible explanation is that whites are on average better educated and it is possible that for higher education groups, wage setting is more likely to take place in the context of an internal labor market than a spot labor market. One feature of internal labor markets is the stability of the wage process. Therefore, whites are more likely to have higher persistence in their earnings process. (Using different methods, both Bertrand (1997) and Hirano (1999) found a higher autoregressive coefficient in the earnings equation for the higher education group and gave a similar interpretation).

¹⁰The 1973 matched file covers earnings from 1951 to 1975 and the 1978 matched file covers earnings from 1951 to 1978. In this study, I focus on the period from 1957 to 1973.

¹¹For any given year, about 15% of records were censored at zero.

a year-long spell of unemployment.¹² Although the dynamics in employment and labor supply are interesting and important in their own right,¹³ in this paper, I analyze earnings dynamics conditional on employment in the covered sector. I treat the observations with zero earnings as randomly missing. For the first part of the analysis, I use only individuals who had positive SSA earnings in each year during the period of interest (1957-73).¹⁴ In the second part, I use an unbalanced panel of all observations with positive SSA earnings in three consecutive time periods. The latter provides a larger sample for our analysis of a more complicated model specification.

In addition to earnings, the data have information on the respondent's gender, race, education, age, and region of residence in the survey year. I use a sample of black and white men living in the South who were born in 1930-1939. This cohort, with the youngest man aged 18 in 1957 and the oldest man 43 in 1973, was relatively young and therefore might have richer information for the studying of earnings dynamics. In addition, racial discrimination in the labor market was more of concern in the South during that period. If discrimination plays a role in the difference between black and white earnings process, we are more likely to find it in the South.¹⁵

In the SSA earnings data, the real tax ceiling was changing dramatically during the period of interest. It increased from about \$15,000 in late 1950s to \$24,000 in 1973 (inflated to 1982-84 constant dollars). In addition, the percentage of records that were censored differs by race. Figure 3 plots the fraction of the sample that was top-coded over time by race. A significantly higher percentage of records were censored in the white men's sample than in the black men's, and the difference was changing as well. For the black men's sample, the percentage of censoring rose from about 3% in late 1950s to about 30% in early 1970s, while it went from about 20% to 60% for

¹²Since about 90% of wage and salary and self-employed workers were covered by Social Security in early 1970's (Card and Sullivan (1988)), and about 85% of the sample has positive earnings, the sample selection due to the last two factors might not be too important.

¹³The employment dynamics can be examined more carefully by applying some of the existing methods for estimating sample selection panel data models (for example, Kyriaziou (1997, 1999)).

¹⁴This reduces the sample size by about 40%.

¹⁵The number of observations with positive earnings in all 17 years during 1957-73 is 226 for blacks and 1,883 for whites. In this sample, white men are better educated on average than black men. The mean years of schooling is 11.91 (s.d.=3.01) for white men and 10.42 (s.d.=2.91) for black men.

the whites.¹⁶ So in order to consistently estimate the dynamics by race, one has to deal with the censoring.

Figure 3

4 Empirical Model

The dynamic earnings model I estimate is an autoregressive model with individual heterogeneity. The previous literature on earnings dynamics has found that such a model fits the data reasonably well; in particular, it is consistent with the empirical findings that there are large cross-individual variations and persistent autocorrelation over time (see Abowd and Card (1989) for example). Ashenfelter (1978) discussed some additional justifications for the use of such autoregressive models for the earnings process. First, at the most rudimentary level, any theory of the determination of earnings will imply that current earnings are the results of a variety of historical factors, such

¹⁶As a very crude way to investigate the stationarity of the log earnings process, in Appendix Figure 1, I plot the difference in the log-earnings between the 30th and 15th percentile (since those earnings are uncensored). The difference appears pretty constant over time for the white sample, although it has greater fluctuations over time for the black sample.

as education, experience and other factors that influence earnings capacity. Second, such models can characterize the known empirical facts about the structure of earnings. For example, earnings increase with age but at a decreasing rate, and the autoregression can accommodate this fact.¹⁷ Finally, this model can also be rationalized as the result of an optimal investment program in human capital by individual workers (for details, see Ashenfelter (1978)).

Specifically, the model I study is

$$y_{it}^* = \gamma y_{it-1}^* + z_{it}\phi + \eta_i + v_{it} \quad \text{for } t = 1, 2, \dots, T, T \geq 2$$

where y_{it}^* is the logarithm of real annual earnings, z_{it} is a vector of time-varying explanatory variables and η_i is the individual specific fixed effect which does not change over time for a certain individual.¹⁸ Because records are censored from above at the social security maximum taxable earnings level, we only observe

$$y_{it}^0 = \min(c_t, y_{it}^*)$$

where c_t is the tax ceiling at time t .¹⁹

To fit in the framework established in section 2, I translate this right censoring (or top coding) problem into a left censoring (or bottom coding) one. Define

$$y_{it} = c_t - y_{it}^0,$$

then, instead of censoring y_{it}^* from above by c_t , we observe $c_t - y_{it}^*$ censored from below by zero. That is

$$c_t - y_{it}^0 = \max(0, c_t - y_{it}^*).$$

Note that

$$\begin{aligned} c_t - y_{it}^* &= c_t - \gamma y_{it-1}^* - z_{it}\phi - \eta_i - v_{it} \\ &= \gamma (c_{t-1} - y_{it-1}^*) + c_t - \gamma c_{t-1} - z_{it}\phi - \eta_i - v_{it}. \end{aligned}$$

¹⁷As an illustration, consider the AR(1) model $y_t = \alpha + \gamma y_{t-1}$. The difference equation has the solution $y_t = (y_0 - \frac{\alpha}{1-\gamma})\gamma^t + \frac{\alpha}{1-\gamma}$. For $0 < \gamma < 1$ and $y_0 < \frac{\alpha}{1-\gamma}$, earnings approaches the asymptote $\frac{\alpha}{1-\gamma}$ gradually from below in a way that is often observed for the empirical earnings-age profile.

¹⁸From this section on, I use γ and β to denote the true parameter values.

¹⁹As mentioned earlier, I ignore the censoring-from-below problem and treat the observations with zero earnings in the covered sector as randomly missing data.

After re-defining the variables and parameters as follows

$$\begin{aligned}\tilde{y}_{it}^* &= c_t - y_{it}^* \\ x_{it} &= (-z_{it}, c_t, c_{t-1}) \\ \beta &= (\phi, 1, -\gamma)' \\ \alpha_i &= -\eta_i \\ \varepsilon_{it} &= -v_{it},\end{aligned}$$

we have the model discussed in the previous section

$$\begin{aligned}\tilde{y}_{it}^* &= \gamma \tilde{y}_{it-1}^* + x_{it} \beta + \alpha_i + \varepsilon_{it} \\ y_{it} &= \max(0, \tilde{y}_{it}^*).\end{aligned}$$

The data have information on y_{it}^0 (logarithm of annual real earnings) and c_t (logarithm of the real maximum taxable earnings), so we can construct the new series $y_{it} = c_t - y_{it}^0$. Given the sample of men living in the South by race, the only time invariant explanatory variable is education, which cannot be separately identified from the fixed effects α_i if we rule out the time-varying parameters associated with it.²⁰ The only time-varying explanatory variables available are age, the tax ceiling, and the time dummies.²¹ In practice, the time-varying tax ceilings c_t and c_{t-1} are absorbed into the time dummies. If the age variable enters the model linearly (which may be reasonable given the narrowly defined birth cohort), then it is also absorbed by the individual effect α_i and time dummies.²² In this case, the reduced form coefficient on the time dummy is

$$\beta_t = \phi_t + c_t - \gamma c_{t-1}.$$

For the first part of the empirical analysis, I focus on this simple specification where the only exogenous explanatory variables are time dummies. I will relax this restriction of a linear earnings-

²⁰This specification might be unrealistic if the returns to education were changing over time. However, Chay and Honoré (1998) used the same data set to estimate the fixed effects censored regression panel data model (but without lagged dependent variables) and found that the estimates on returns to education were very stable during the period of interest.

²¹We only have information on the demographics at one point in time (1973 or 1978).

²²Note that age_{it} can be decomposed into age_{i0} and t , where the former is absorbed by individual effects α_i and the latter is absorbed by the time dummies β_t .

age profile in the second part of the analysis and explicitly include age dummies to allow for a more general earnings-age profile.

5 Estimation

In this section, I first present results from the linear GMM estimation, which ignores the censoring problem and treats all the earnings as their true values. Next, I discuss the estimation strategy for the nonlinear GMM which is based on the orthogonality conditions constructed in previous sections.

5.1 Linear GMM: Without Correction for Censoring

First, I consider the estimation based on moment conditions ignoring the censoring problem. If there were no censoring and all the earnings we observe were the true values, then we would have a linear dynamic panel data model with lagged dependent variables and individual fixed effects. The linearity allows us to use the usual “trick” to difference away the fixed effects and construct orthogonality conditions which can be used to estimate the parameters via GMM estimation (see Anderson and Hsiao (1981), for example). More specifically, the model is

$$y_{it}^* = \gamma y_{it-1}^* + \phi_t + \alpha_i + \varepsilon_{it} \quad \text{for } t = 1, 2, \dots, T, \quad T \geq 2 \quad (6)$$

where $\{\varepsilon_{it}\}_{t=1}^T$ is i.i.d. conditional on $\{\alpha_i, y_{i0}\}$ (this is stronger than necessary for the estimation of the linear model). One implication of this assumption is that

$$E \left[\Delta \varepsilon_{it} \mid \{y_{it}^*\}_{t=0}^{t-2} \right] = 0.$$

Then the first difference of equation (6) satisfies

$$E \left[(\Delta y_{it}^* - \gamma \Delta y_{it-1}^* - \Delta \phi_t) \mid y_{it-j}^* \right] = 0 \quad \text{for } 2 \leq j \leq t \leq T \text{ and } 3 \leq t \leq T.$$

Among many others, we can construct $2(T-2)$ orthogonality conditions as follows

$$E \left[(\Delta y_{it}^* - \gamma \Delta y_{it-1}^* - \Delta \phi_t) \right] = 0 \quad \text{for } 3 \leq t \leq T$$

$$E \left[(\Delta y_{it}^* - \gamma \Delta y_{it-1}^* - \Delta \phi_t) \cdot y_{it-2}^* \right] = 0 \quad \text{for } 3 \leq t \leq T.$$

If all y^* 's are observed, then the estimates based on the above equations should be consistent and asymptotically normally distributed. We can then apply the standard GMM estimation to the panel data model with predetermined variables. For ease of notation, the superscript '*' is dropped in the rest of this subsection.

Define a transformation matrix K with dimension $(T-2) \times (T-1)$ and rank $(T-2)$ as

$$K = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}$$

and a block diagonal matrix of instruments Z with dimension $(T-2) \times 2(T-2)$ as

$$Z_i = \begin{pmatrix} 1 & y_{i1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & y_{i2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & y_{i,T-2} \end{pmatrix}.$$

The regressor matrix X_i is $(T-1) \times (T-1)$: $X_i = \begin{pmatrix} y_{i1} & 0 & \cdots & 0 \\ y_{i2} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ y_{iT-1} & 0 & \cdots & 1 \end{pmatrix}$ and the dependent

variable $y_i = \begin{pmatrix} y_{i2} \\ y_{i3} \\ \vdots \\ y_{iT} \end{pmatrix}$. I also denote the combined error term as $u_{it} = \alpha_i + \varepsilon_{it}$, it then follows that $u_i = y_i - X_i\theta$, where $\theta = (\gamma, \phi_3, \dots, \phi_T)'$.²³

Under the standard assumptions on the error term ε conditional on the instruments (e.g. conditional mean equals zero), we can write the moment conditions compactly as

$$E[Z_i' K u_i] = 0.$$

Then the optimal GMM estimator of θ based on these moments above can be computed as

$$\hat{\theta} = (M'_{zx} A M_{zx})^{-1} M'_{zx} A M_{zy} \quad (7)$$

where $M_{zx} = (\sum_{i=1}^n Z_i' K X_i)$, $M_{zy} = (\sum_{i=1}^n Z_i' K y_i)$, and $A = (\sum_{i=1}^n Z_i' K \tilde{u}_i \tilde{u}_i' K' Z_i)^{-1}$ and $\tilde{u}_i = y_i - X_i \tilde{\theta}$ is computed from some preliminary consistent estimate $\tilde{\theta}$ (for example, using $A = I$ in equation (7) to obtain such an estimate). Again standard GMM asymptotics give its variance as

$$Var(\hat{\theta}) = [E(Z_i' K X_i)' [E(Z_i' K \hat{u}_i \hat{u}_i' K' Z_i)]^{-1} E(Z_i' K X_i)]^{-1}$$

²³Note the time dummy for period 2 is dropped to avoid the multicollinearity.

where $\hat{u}_i = y_i - X_i\hat{\theta}$.

The linear GMM estimates are reported in Table 1. The estimates of the AR(1) coefficient γ are approximately 0.4 for both blacks and whites.²⁴

Table 1 Estimates of AR(1) Coefficient from Linear GMM: Dependent Variable is Log Real Annual Earnings (in Thousands)		
	Black	White
Number of Observations	N=226	N=1883
Number of Time Periods	T=17	T=17
Linear GMM Estimate of γ (S.E.)	0.379 (0.030)	0.399 (0.018)

5.2 Nonlinear GMM: With Correction for Censoring²⁵

In this subsection, I consider the nonlinear GMM estimation based on the orthogonality conditions constructed in section 2, which explicitly take into account the censoring problem. Resume the notation y_{it}^* for the latent true earnings and denote the observed value as $y_{it} = \max(0, \tilde{y}_{it}^*)$ where

²⁴As in the standard GMM estimation, we can test the overidentifying restrictions based on the minimized (optimally weighted) criterion function, which gives asymptotically chi-squared distributed test statistics. More specifically, we can calculate

$$S = \left[\sum_{i=1}^n \left(y_i - X_i\hat{\theta} \right)' K' Z_i \right] \left(\sum_{i=1}^n Z_i' K \left(y_i - X_i\hat{\theta} \right) \left(y_i - X_i\hat{\theta} \right)' K' Z_i \right)^{-1} \left[\sum_{i=1}^n Z_i' K \left(y_i - X_i\hat{\theta} \right) \right] \rightarrow \chi^2(k)$$

where $k = 2(T-2) - (T-1) = T-3$. The test-statistics are 16.79 and 22.90 for blacks and whites, respectively. Therefore, the overidentifying restrictions are not rejected at the 95% level for both black and white samples. (The 95% critical value of a $\chi^2(14)$ is 23.69.) It should be noted that failure to reject the model does not imply that censoring should be ignored or that the GMM estimator that ignores censoring will consistently estimate the parameters of the assumed AR(1) model of latent earnings.

²⁵To gauge the importance of individual heterogeneity in the earnings process, I also estimate a dynamic model without individual fixed effects. Specifically, I estimate a cross-sectional censored regression model by year, using some of the existing methods, namely Tobit (MLE) and the Censored Least Absolute Deviations (CLAD) (see Powell (1984)). The results for selected years are reported in Appendix Table 1. Here we see that estimates of the AR(1) coefficient, which vary from 0.7 to 1.1, are much higher than the estimates obtained from the panel data model. This suggests that it is important to allow for individual heterogeneity in any earnings model.

$\tilde{y}_{it}^* = c_t - y_{it}^*$. Furthermore, let δ_t be the coefficient on the reduced form time dummy variables, that is $\delta_t = \Delta\beta_t$. To translate the variables and parameters into the same notation as used in section 2, define

$$\begin{aligned}\tilde{y}_t(\gamma) &\equiv y_t - \gamma y_{t-1} \\ \Delta_t &\equiv \Delta_t(\gamma, \delta_t) \equiv \gamma y_{t-2} - \delta_t\end{aligned}$$

where Δ_t plays the same role as “ $\Delta x\beta$ ” as before. Define an indicator function for trimming, which equals 1 if an observation falls in the interior of region $A \cup B$ (as described in section 2) and zero otherwise:²⁶

$$\begin{aligned}1\{S\} &= 1\{y_t > 0, y_{t-1} > 0, y_{t-2} > 0\} \\ &\cdot [1\{\Delta_t \geq 0\} \cdot 1\{y_{t-1} > \Delta_t - \gamma(\tilde{y}_t(\gamma) + \Delta_t)\} \cdot 1\{\tilde{y}_t(\gamma) > -\Delta_t\} \\ &+ 1\{\Delta_t < 0\} \cdot 1\{\tilde{y}_t(\gamma) > -\Delta_t\}].\end{aligned}$$

Conditional on being in region $A \cup B$, the pair of variables $(y_{i,t-1}, \tilde{y}_t(\gamma))$ is symmetrically distributed around the 45° line through $(\Delta_t(\gamma, \delta), 0)$. Therefore, for a function satisfying $\xi(e_1, e_2) = -\xi(e_2, e_1)$ for all (e_1, e_2) , I can construct $2(T-2)$ orthogonality conditions as

$$E[1\{S\} \cdot \xi(y_{i,t-1} - \Delta_t(\gamma, \delta), \tilde{y}_t(\gamma))] = 0 \quad \text{for } t = 3, \dots, T \quad (8)$$

$$E[1\{S\} \cdot \xi(y_{i,t-1} - \Delta_t(\gamma, \delta), \tilde{y}_t(\gamma)) \cdot y_{i,t-2}] = 0 \quad \text{for } t = 3, \dots, T \quad (9)$$

to estimate the $T-1$ parameters $\theta = (\gamma, \delta_3, \dots, \delta_T)'$. Examples of $\xi(e_1, e_2)$ are $\xi(e_1, e_2) = e_1 - e_2$ and $\xi(e_1, e_2) = \text{sgn}(e_1 - e_2)$.²⁷ In this paper, I focus on the sign function mainly for reason of

²⁶Note that the first piece, $1\{y_t > 0, y_{t-1} > 0, y_{t-2} > 0\}$, requires that there are three consecutive uncensored observations, and the second piece indicates truncation of observations falling outside the region $A \cup B$. In both cases, $\Delta_t \geq 0$ and $\Delta_t < 0$, we need to truncate observations from below at line (3) which is given by $\tilde{y}_t(\gamma) = -\Delta_t$. In addition, in the case of $\Delta_t \geq 0$, we need to further truncate observations from the left at line (4), which is given by $y_{t-1} = \Delta_t - \gamma(\tilde{y}_t(\gamma) + \Delta_t)$.

²⁷When $\xi(e_1, e_2) = e_1 - e_2$, equation (8) corresponds to the condition that the expected vertical distance from a point $(y_{i,t-1}, \tilde{y}_t(\gamma))$ in region A to line (1) equals the expected horizontal distance from a point $(y_{i,t-1}, \tilde{y}_t(\gamma))$ in region B to line (1). When $\xi(e_1, e_2) = \text{sgn}(e_1 - e_2)$, equation (9) corresponds to the condition that the probability an observation $(y_{i,t-1}, \tilde{y}_t(\gamma))$ falls in region A equals the probability that observation falls in region B .

robustness.²⁸

The standard GMM estimator is defined as

$$\hat{\theta}_{GMM} = \arg \min_{\theta} Q_n(\theta) = \arg \min_{\theta} m_n(\theta)' \cdot A_n \cdot m_n(\theta)$$

where $m_n(\theta)$ is the sample analog of the orthogonality conditions

$$m_n(\theta) = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n [1\{S\} \cdot \text{sgn}(y_{i,t-1} - \Delta_t(\gamma, \delta) - \tilde{y}_t(\gamma))] & t = 3, \dots, T \\ \frac{1}{n} \sum_{i=1}^n [1\{S\} \cdot \text{sgn}(y_{i,t-1} - \Delta_t(\gamma, \delta) - \tilde{y}_t(\gamma)) \cdot y_{i,t-2}] & t = 3, \dots, T \end{pmatrix} \quad (10)$$

and A_n is a positive definite weighting matrix.

6 Inference

There are several ways to conduct inference within the GMM framework. One common way to construct asymptotic confidence intervals is based on the asymptotic distribution of the GMM estimator,²⁹ which is generally of the form

$$\sqrt{n}(\hat{\theta}_{GMM} - \theta) \longrightarrow N(0, (\Gamma' A \Gamma)^{-1} [\Gamma' A V A \Gamma] (\Gamma' A \Gamma)^{-1}) \quad (11)$$

where

$$\begin{aligned} \Gamma &= \frac{\partial}{\partial \theta} E[m(\theta_0)] \\ V &= E[m(\theta_0) m(\theta_0)'] \end{aligned}$$

²⁸The sign function is more robust to outliers than the identity function. In addition, there is another efficiency consideration. Calculations in Honoré (1993a) suggest that, in the static censored regression panel data model, the fixed effect estimator based on minimizing an absolute value loss function is more efficient than one based on a quadratic loss function, even when the error is normally distributed. Because the first order conditions of the two loss functions are similar to the sign function and the identity function (both in terms of the difference between the two arguments), respectively, we might expect the estimator based on the sign orthogonality conditions to be more efficient here as well.

²⁹Under standard regularity conditions, the asymptotic distribution of the estimators can be derived, loosely speaking, by taking a Taylor expansion of the first order conditions associated with the minimization of the criterion function, and applying the central limit theorem and the law of large numbers. The discontinuity of the moment conditions can be dealt with in similar fashion as in Pakes and Pollard (1989) (see also Newey and McFadden (1994)).

and $A = \text{plim } A_n$. For the optimal weighting matrix $A = V^{-1}$, the variance-covariance matrix can be simplified to $(\Gamma'V^{-1}\Gamma)^{-1}$. This approach requires consistent estimates of the variance-covariance, and it is sometimes difficult to implement in practice when the moments are not smooth.

As an alternative, we can base our inference on the test-statistics for the overidentifying restrictions. In general, the minimized (optimally weighted) criterion function is distributed chi-squared under the null. If, for example, our main interest is in the autoregressive parameter γ , then we can concentrate out the other parameters and construct the confidence interval for γ based on the constrained minimized (optimally weighted) criterion function. The latter is distributed chi-squared under the null with degrees of freedom equal to the number of moments minus the number of parameters to be estimated (not including γ). One advantage of this approach is that we do not need to estimate the variance-covariance matrix $(\Gamma'V^{-1}\Gamma)^{-1}$.

Unfortunately, both of the approaches to performing inference entail minimization of the quadratic criterion function over a large set of parameters. It turns out that the numerical minimization is often hard to achieve in practice. The main problem is that, although we have seen that the orthogonality conditions hold at the true parameter values, it is not known whether those are the only values that satisfy the conditions. This problem was also present in the model considered by Honoré (1993b). However, using another set of moment conditions, Honoré and Hu (1999) proved that the parameters are indeed identified in that model. That result does not generalize to the present model. In this paper, I will therefore present some alternative ways to conduct inference. I will first illustrate the problem, and then discuss inferences for the key parameter γ based on alternative tests. Specifically, I show how, for a fixed parameter value γ , one can estimate the rest of the parameters, and how to use those to construct asymptotically valid confidence intervals for the autoregressive parameter, γ .

6.1 Difficulties with GMM Estimation

The main difficulty with the GMM estimation in this example is that the orthogonality conditions can be trivially satisfied when the parameter values are arbitrarily large. Intuitively, for a fixed value of γ , when the value of the parameter δ shoots to infinity, the number of observations falling in the (non-truncated) region $A \cup B$ (in Figure 1 and 2) approaches zero and therefore the moment conditions can be trivially satisfied. The problem is illustrated in Figure 4, where the data used

are the white men's sample from 1965-67 and the parameter γ is fixed at 0.5. It is clear from the graph that the moment condition can be “solved” in three ways in this example: either by setting $\hat{\delta} = -0.1$ (approximately), or by letting $\hat{\delta}$ diverge to $+\infty$ or $-\infty$.

Figure 4: Moment Condition as a Function of δ When $T = 3$ and $\gamma = 0.5$

One possible remedy is to use the continuously updating GMM in which the weighting matrix also changes with parameter values at each iteration (instead of using a fixed weighting matrix, say, estimated from the previous iteration).³⁰ Hansen, Heaton, and Yaron (1996) suggested that inferences directly based on the criterion function of the continuously updating GMM yield fairly reliable test statistics and confidence intervals even when identification is weak. In our example, the intuition behind this is that although the moment will approach zero at “wild” parameter values,

³⁰More specifically, for a fixed value of γ_0 , we can estimate δ by

$$\hat{\delta} = \min_{\delta} Q_n(\delta, \gamma_0) = m_n(\delta, \gamma_0)' \cdot A_n \cdot m_n(\delta, \gamma_0)$$

where $A_n = \left[\frac{1}{n} \sum_{i=1}^n (m_n(\delta, \gamma_0)) (m_n(\delta, \gamma_0))' \right]^{-1}$. Denote this minimized function value as $Q_n^\gamma(\hat{\delta}, \gamma_0)$, then $n \cdot Q_n^\gamma(\hat{\delta}, \gamma_0)$ is distributed as chi-squared with k degrees of freedom where k equals the number of moments minus the number of parameters to be estimated. In the current setup, $k = 2(T - 2) - (T - 2) = T - 2$. That is

$$n \cdot Q_n^\gamma(\hat{\delta}, \gamma_0) \longrightarrow \chi^2(T - 2).$$

so will be the variance term, which is used as the weighting matrix. As a result, the quadratic criterion function will not necessarily be minimized at those values, and the continuously updating GMM could potentially be better-behaved. To examine this possibility, for the same white sample, using three time periods from 1965-67, and for a fixed value of $\gamma = 0.5$, I plot the continuous updating GMM quadratic criterion function as a function of the only parameter δ . As can be seen from Figure 5, the function can still be trivially minimized by making δ large in absolute values.³¹

Figure 5: Continuous Updating GMM as a Function of δ When $T = 3$ and $\gamma = 0.5$

6.2 Estimators of the Parameters on the Time Dummy Variables

The difficulty with the usual GMM estimation motivates me to resort to alternative tests other than the commonly used one based on the minimization of the criterion function. The main idea is to use a subset of the moments (an exactly identified system) to estimate the parameters on the time dummy variables, δ'_t s, for a given value of γ . We can then use the rest of the moment conditions to test the validity of the model, and hence, potentially reject the parameter values of γ .

³¹In fact, large non-sensible estimates were indeed obtained in the GMM estimation of δ'_t s. For example, estimates of δ were in the magnitude of 10 to 20 for some time periods. Since δ measures the difference in log real earnings between two consecutive time periods, these numbers are too large to be credible.

The rationale for this approach is that, for a fixed value of γ , we can integrate each of the moment functions to “recover” an objective function of δ_t . The first order condition from minimizing this objective function over δ_t should be the same as the moment condition. The advantage of minimizing the objective function over finding the roots of the FOC is that the objective function is piecewise linear, and therefore can be easily minimized at the sensible parameter values of δ_t .

To see this, note that for a fixed value of γ , each of the $(T - 2)$ moment conditions in (8) with $\xi(e_1, e_2) = \text{sgn}(e_1 - e_2)$ is a function of λ_t with $\lambda_t \equiv \Delta(\gamma, \delta_t) = \gamma y_{t-2} - \delta_t$. Define $z_1 = y_{t-1}$ and $z_2 = \tilde{y}_t(\gamma)$. Then the moment function is of the form

$$E[f(z_1, z_2; \lambda_0)] = 0$$

where

$$f(z_1, z_2; \lambda) = 1 \left\{ -z_2 \leq \lambda \leq \frac{z_1 + \gamma z_2}{1 - \gamma} \right\} \cdot \text{sgn}(z_1 - z_2 - \lambda).$$

The indicator function $1 \left\{ -z_2 \leq \lambda \leq \frac{z_1 + \gamma z_2}{1 - \gamma} \right\}$ corresponds to the condition that an observation is in the region $A \cup B$ (in Figure 1 and 2). The function f equals 1 if an observation is in region A and -1 if it is in region B , and zero otherwise. Note that f is not a continuous function because of both the nature of the sign function and the truncation (by $1 \left\{ -z_2 \leq \lambda \leq \frac{z_1 + \gamma z_2}{1 - \gamma} \right\}$). However, instead of defining the estimator of λ as the value that satisfies $\frac{1}{n} \sum_{i=1}^n f(z_{i1}, z_{i2}; \tilde{\lambda}) = 0$, I define an estimator $\hat{\lambda}$ as the minimizer of the objective function $\frac{1}{n} \sum_{i=1}^n F(z_{i1}, z_{i2}; \tilde{\lambda})$, where

$$F(z_1, z_2; \lambda) = \begin{cases} z_1 & \text{if } \lambda < -z_2 \\ |z_1 - z_2 - \lambda| & \text{if } -z_2 \leq \lambda \leq \frac{z_1 + \gamma z_2}{1 - \gamma} \\ \frac{z_2 + \gamma z_1}{1 - \gamma} & \text{if } \frac{z_1 + \gamma z_2}{1 - \gamma} < \lambda. \end{cases}$$

Now that the function F is continuous, piecewise linear, and almost everywhere differentiable, and the FOCs of the minimization of the objective function $\frac{1}{n} \sum_{i=1}^n F(z_{i1}, z_{i2}; \tilde{\lambda})$ are just the moment conditions $\frac{1}{n} \sum_{i=1}^n f(z_{i1}, z_{i2}; \tilde{\lambda}) = 0$.³²

Specifically, for a fixed value of γ_0 , define the estimator of δ_t as

$$\hat{\delta}_t = \arg \min_{\delta_t} Q_n(\delta_t)$$

³²More rigorously, the objective function is left- and right-differentiable in its argument, and the estimator $\tilde{\lambda}$ can be shown to approximately satisfy the first order condition for the minimization of this objective function.

where

$$\begin{aligned}
Q_n(\delta_t) = & \frac{1}{n} \sum_{i=1}^n 1 \{ (y_{it}, y_{it-1}, y_{it-2}) > 0 \} \\
& \cdot \left\{ |y_{it-1} - \tilde{y}_t(\gamma_0) - (\gamma_0 y_{it-2} - \delta_t)| \cdot 1 \left\{ -\tilde{y}_t(\gamma_0) < \gamma_0 y_{it-2} - \delta_t < \frac{\gamma_0 \tilde{y}_t(\gamma_0) + y_{it-1}}{1 - \gamma_0} \right\} \right. \\
& + y_{it-1} \cdot 1 \{ \gamma_0 y_{it-2} - \delta_t < -\tilde{y}_t(\gamma_0) \} \\
& \left. + \left| \frac{\tilde{y}_t(\gamma_0) + \gamma_0 y_{it-1}}{1 - \gamma_0} \right| \cdot 1 \left\{ \gamma_0 y_{it-2} - \delta_t > \frac{\gamma_0 \tilde{y}_t(\gamma_0) + y_{it-1}}{1 - \gamma_0} \right\} \right\}.
\end{aligned}$$

Figure 6 plots the objective function $Q_n(\delta_t)$ as a function of δ using the white men's sample from the period 1965 to 1967 and fixing $\gamma = 0.5$. Comparing Figure 6 to the moment function in Figure 4, we see that the minimization problem has its minimum at the “sensible” solution to the moment conditions.

The moment condition using $\xi(e_1, e_2) = e_1 - e_2$ has the same problem and can be dealt with in a similar fashion. See the appendix for details.

Figure 6: Objective Function $Q_n(\delta)$ as a Function of δ When $T = 3$ and $\gamma = 0.5$

6.3 Alternative Tests

In this section, I construct confidence intervals for the AR(1) parameter γ . The basic idea is to fix $\gamma = \gamma_0$ and then test whether a set of overidentifying restrictions are satisfied. The confidence interval is the set of values of γ for which they are not rejected.

I consider two tests. For the first test, for a fixed value of γ , a subset of $T - 2$ moments are used to estimate the $T - 2$ parameters $\{\delta_t\}_{t=3}^T$, and then I test whether the rest of the $T - 2$ moments are satisfied at those estimates. For the second test, (again for a fixed value of γ), I estimate δ , separately from the two sets of $T - 2$ moment conditions, and then find the optimal linear combination of those two estimators. The resulting estimator is asymptotically equivalent to the optimal GMM estimator. We can therefore evaluate the optimally weighted quadratic criterion function at this estimator, which should be asymptotically distributed as chi-squared with degrees of freedom equal to the number of overidentifying restrictions.

It is not obvious that the resulting confidence intervals are asymptotically justified. The basic problem is identification. Consider the first test that estimates the δ from one set of moments and then test whether a second set is satisfied. If the first set of moments does not identify the δ , then it is possible that a “false” solution will be obtained. If that false solution does not happen to also be a false solution to the second set of moment conditions, then the true γ_0 will be falsely rejected (asymptotically). A similar problem can occur for the second test. The following proposition gives the regularity conditions under which this will not happen and hence the asymptotic confidence intervals are justified. For ease of notation (and added generality), we consider the case with general explanatory variables x .

Proposition 2 *Assume the data are i.i.d and generated from model $y_{it}^* = \gamma_0 y_{it-1}^* + x_{it} \beta_0 + \alpha_i + \varepsilon_{it}$ $t = 1, 2$. Assume*

(1) ε_{i1} and ε_{i2} are independent, and identically and continuously distributed conditional on $(y_{i0}^*, x_{i1}, x_{i2}, \alpha_i)$.

In addition, the (common) density of ε_{i1} and ε_{i2} is strictly log-concave.

(2) $E[x_{i1}] < \infty$ and $E[x_{i2}] < \infty$.

(3) *There is no proper linear subspace of $\mathcal{R}^k$ containing the random variable $1\{y_{i0} > 0, y_{i1} > 0, y_{i2} > 0\} \Delta x$ with probability 1 where $\Delta x = x_{i1} - x_{i2}$.*

Denote $m(\gamma, \beta) = \begin{pmatrix} 1\{A \cup B\} \cdot \text{sgn}(y_1 - \tilde{y}(\gamma) - \Delta x \beta) \\ 1\{A \cup B\} \cdot \text{sgn}(y_1 - \tilde{y}(\gamma) - \Delta x \beta) \cdot y_0 \end{pmatrix}$.

Then, at $\gamma = \gamma_0$, β_0 satisfies the moment conditions $E[m(\gamma_0, \beta_0)] = 0$ uniquely.

6.3.1 Test I

In this subsection, we discuss the first test. To fix ideas, denote the parameters by $\delta \equiv (\delta_3, \dots, \delta_T)'$ and the moments by $m = (m_3, \dots, m_T)'$ and $h = (h_3, \dots, h_T)'$ where

$$\begin{aligned} m_t &= m_t(\delta_t, \gamma) = 1\{S\} \cdot \text{sgn}(y_{i,t-1} - \Delta_t(\gamma, \delta_t) - \tilde{y}_t(\gamma)) \\ h_t &= h_t(\delta_t, \gamma) = 1\{S\} \cdot \text{sgn}(y_{i,t-1} - \Delta_t(\gamma, \delta_t) - \tilde{y}_t(\gamma)) \cdot y_{i,t-2}. \end{aligned}$$

For a fixed value γ_0 , denote the estimates based on the moments $E[m_t(\delta_t, \gamma_0)] = 0$ as $\hat{\delta}_t$. Then the following proposition is useful.

Proposition 3 *Under random i.i.d. sampling and standard regularity conditions, and given consistent estimates of $\hat{\delta}$, we have*

$$n \left(\frac{1}{n} \sum h(\hat{\delta}, \gamma_0) \right)' V^{-1} \left(\frac{1}{n} \sum h(\hat{\delta}, \gamma_0) \right) \overset{A}{\sim} \chi^2(T-2)$$

where

$$\begin{aligned} V &= E \left(h - HM^{-1}m \right) \left(h - HM^{-1}m \right)' \\ H &= \frac{\partial E h}{\partial \delta} \\ M &= \frac{\partial E m}{\partial \delta}. \end{aligned}$$

A consistent estimator of V is given by

$$\hat{V} = \frac{1}{n} \sum \left(\hat{h} - \hat{H} \hat{M}^{-1} \hat{m} \right) \left(\hat{h} - \hat{H} \hat{M}^{-1} \hat{m} \right)'$$

where

$$\begin{aligned} \hat{m}_t &= 1\{S\} \cdot \text{sgn} \left(y_{i,t-1} - \Delta_t(\gamma_0, \hat{\delta}_t) - \tilde{y}_t(\gamma_0) \right) \\ \hat{h}_t &= 1\{S\} \cdot \text{sgn} \left(y_{i,t-1} - \Delta_t(\gamma_0, \hat{\delta}_t) - \tilde{y}_t(\gamma_0) \right) \cdot y_{i,t-2} \end{aligned}$$

and $\hat{H}$ and $\hat{M}$ are consistent estimators of H and M , respectively.

Based on this proposition, we can implement the test as follows: (i) For a fixed value of γ_0 , estimate each of the $T-2$ time dummy parameters δ_t , $t = 3, \dots, T$, by minimizing the corresponding objective function recovered from each of the $T-2$ moment functions $\frac{1}{n} \sum_{i=1}^n m_t(\delta_t, \gamma_0) = 0$, denoted as $\hat{\delta}_t$; (ii) Calculate the sample moments $\frac{1}{n} \sum_{i=1}^n h(\hat{\delta}, \gamma_0)$ and the variance $\hat{V}$; (iii) Compute

the test-statistic $n \left(\frac{1}{n} \sum_{i=1}^n h(\hat{\delta}, \gamma_o) \right)' \hat{V}^{-1} \left(\frac{1}{n} \sum_{i=1}^n h(\hat{\delta}, \gamma_o) \right)$, denoted as Test *A*. We can then compare it with the critical value of a $\chi^2(T-2)$.

Naturally, we can compute a similar test-statistic the other way around. Specifically, we can first estimate δ based on the moment conditions $\frac{1}{n} \sum_{i=1}^n h_t(\delta_t, \gamma_0) = 0$ (by minimizing an appropriate objective function which can be constructed in a similar way). Denote the estimator by $\tilde{\delta}$, and we can then compute $n \left(\frac{1}{n} \sum_{i=1}^n m(\tilde{\delta}, \gamma_o) \right)' \tilde{V}^{-1} \left(\frac{1}{n} \sum_{i=1}^n m(\tilde{\delta}, \gamma_o) \right)$, where $\tilde{V} = \frac{1}{n} \sum \left(\tilde{m} - \tilde{M} \tilde{H}^{-1} \tilde{h} \right) \left(\tilde{m} - \tilde{M} \tilde{H}^{-1} \tilde{h} \right)'$ with the terms defined in the obvious way. Denote this as Test *B*. Asymptotically, Test *A* and *B* are equivalent.

Those two test-statistics are computed for the black and white samples separately, for a sequence of γ values ranging from 0 to 0.9 with an increment of 0.01. The results are plotted in Figure 7a and 7b.³³ The 95% critical value for the chi-squared with degrees of freedom $T-2$ (or 15) is also plotted as a reference line. Values of γ at which the test-statistics are above the reference line are rejected at the 5% level and values of γ at which the test-statistics are on or below the reference line are not rejected. Therefore, values of γ corresponding to the intersections of the $\chi^2(15)$ and the test-statistics yield the 95% confidence interval for γ . As expected, the two test-statistics *A* and *B* are very similar. The confidence interval for the black sample is approximately (0.2, 0.5) and for white sample, it is estimated at (0.35, 0.45).

³³A bandwidth of 0.1 is used for calculating both of the test-statistics plotted in Figure 7a and 7b. A smaller bandwidth (0.05) was also used, and the results (not reported here) are very similar.

Figure 7a: Test-Statistics A and B for Different γ (Black Sample)

Figure 7b: Test-Statistics A and B for Different γ (White Sample)

6.3.2 Test II

Denote the estimators of δ based on $E(m) = 0$ and $E(h) = 0$ as $\beta_1 = (\hat{\delta}_3, \dots, \hat{\delta}_T)'$ and $\beta_2 = (\tilde{\delta}_3, \dots, \tilde{\delta}_T)'$, respectively. It is useful to note that an optimal linear combination of these two estimators is asymptotically equivalent to the optimal GMM estimator.

Proposition 4 Denote $\Sigma = V \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ and $\Sigma^{-1} = \begin{pmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{pmatrix}$. Then the optimal linear combination of the two estimates is

$$\beta_{opt} = A \cdot \beta_1 + (I - A) \cdot \beta_2$$

where $A = [\Sigma^{11} + \Sigma^{12} + \Sigma^{21} + \Sigma^{22}]^{-1} [\Sigma^{11} + \Sigma^{21}]$. In addition, β_{opt} is asymptotically equivalent to the optimal GMM estimators.

Since β_{opt} is asymptotically equivalent to the optimal GMM, the criterion function of the optimal GMM evaluated at β_{opt} is asymptotically distributed as chi-squared with degrees of freedom k , which in this case equals $T - 2$. I denote this test as Test C. The results are plotted in Figure 8a and 8b for the black and white samples, respectively.³⁴

The confidence intervals constructed from the three test-statistics are similar to each other, all of which suggest that the AR(1) coefficient γ in the earnings model is higher for white than for black, although the lesser precision for the black sample prevents a sharper conclusion.

³⁴A bandwidth of 0.1 is used to calculate the two test-statistics reported in Figure 8a and 8b. Again, using a smaller bandwidth does not change the results much.

Figure 8a: Test-Statistic C for Different γ (Black Sample)

Figure 8b: Test-Statistic C for Different γ (White Sample)

6.3.3 Test-Statistics Based on Bootstrap Variances

Note that in calculating all three tests above, it is necessary to estimate H and M , which are the derivatives of the expected value of the moment functions with respect to the parameter δ . Numerical derivatives are used. It turns out that the results are not very sensitive to the choice of bandwidth, although the test-statistics become very non-smooth especially at high values of γ in the black sample, which may be due to the non-smoothness of the moment function (because of the nature of the sign function). In an attempt to improve the estimates of H and M , a second-order polynomial was used to approximate the moment function and the derivative of the polynomial was used. The results (not reported here) are very similar to those obtained using numerical derivatives.

I have also experimented with calculating the test-statistics for Test A and B based on the bootstrap estimates of the variance V by 1,000 replications. The results are plotted in Figure 9a and 9b for the black and white samples, respectively. The confidence intervals are very similar to those in Figure 7a and 7b, although the graphs are smoother when I use this method to calculate the test statistics. For the sake of time-saving, I will use numerical derivatives in the remaining analysis.

Figure 9a: Test-Statistics A and B Based on Bootstrap Variances (Black Sample)

Figure 9b: Test-Statistics A and B Based on Bootstrap Variances (White Sample)

6.4 Standard Asymptotic Variance

To be able to directly compare the results in this section to the linear GMM estimates obtained earlier, I also compute the point estimate of γ along with the standard error. Denote the value of γ at which the test-statistic based on the quadratic criterion function (Test C) is minimized, by $\gamma_{\min}$. Assuming the full set of the moment conditions identify all the parameters, the estimate $\hat{\gamma} = \gamma_{\min}$, along with the corresponding $\hat{\delta}$, will give an estimator which is asymptotically equivalent to the optimal GMM. We can therefore compute the standard asymptotic variance-covariance matrix for the estimates $\hat{\theta} = (\hat{\gamma}, \hat{\delta}_3, \dots, \hat{\delta}_T)'$ as $AVar(\hat{\theta}) = \frac{1}{n} (\hat{\Gamma}' \hat{V}^{-1} \hat{\Gamma})^{-1}$ where $\hat{V} = \frac{1}{n} \sum_{i=1}^n m(\hat{\theta}) m(\hat{\theta})'$ and $\hat{\Gamma}$ is a consistent estimator of $\frac{\partial E(m(\theta))}{\partial \theta}$ evaluated at $\theta = \hat{\theta}$. The results are reported in Table 2.

Table 2		
Estimates of AR(1) Coefficient from Non-Linear GMM: Dependent Variable is Log Real Annual Earnings (in Thousands)		
	Black	White
Number of Observations	N=226	N=1883
Number of Time Periods	T=17	T=17
Nonlinear GMM Estimate of γ (S.E.)	0.210 (0.129)	0.380 (0.051)

The estimated autoregressive parameter γ is much higher for the white men's sample than for the black's. This is in contrast to the results using the linear GMM estimator that ignores the censoring.³⁵ The higher persistence of the earnings process for white men than for black men is consistent with the notion that black men were less likely to have jobs that offered stable wage growth, and their economic well-being was more likely to be subject to fluctuations in the period 1957-73.

7 Some Further Results

7.1 Unbalanced Panel

So far, I have restricted the sample to individuals who had positive social security earnings in each of the 17 years from 1957 to 1973. However, assuming that the observations with zero earnings are randomly missing, the estimation method is applicable to all observations that have positive earnings in three consecutive time periods. In this section, I repeat the exercise for the previous section but with the larger unbalanced panel data sample. The test-statistics for Test A and B using the unbalanced panel are plotted in Figure 10a and 10b for the samples of black and white men, respectively.

³⁵I have also constructed confidence intervals for the parameter γ for the linear GMM estimation in the same way as in the previous section. Appendix Figure 2a and 2b plot the same three test-statistics for the linear model. Compared to Figure 7a and 7b, the confidence intervals from the linear model are of similar length to those from the nonlinear model.

Figure 10a: Test-Statistics A and B Using Unbalanced Panel (Black Sample)

Figure 10b: Test-Statistics A and B Using Unbalanced Panel (White Sample)

Compared to the results in Figure 7a and 7b, the confidence intervals in Figure 10a and 10b are tighter, which is not surprising given that they are based on a larger sample size.³⁶ Interestingly, the confidence intervals for the parameter γ are located closer to zero when the unbalanced panel is used. This might be because the earnings process for individuals who had positive earnings in all 17 years are more persistent (represented by higher γ) than those with more sporadic positive earnings from year to year.

7.2 Age Effects

The basic model specification considered in the previous sections does not fully model the age or cohort effects. As documented in the literature, earnings usually have a distinct and characteristic life-cycle profile, in particular, the profile is hump-shaped, rising to its maximum in the middle years of life, and declining somewhat thereafter. If the earnings-age profile is indeed nonlinear, then this nonlinearity is not taken into account by only including a set of time dummies. This is because the individuals in the sample are not all born in the same year, and hence they have different age in a given year (see Deaton (1997) for more detailed discussion of the decomposition of age, cohort, and time effects). To explore this possibility, I consider a more complicated specification where both time and age effects are considered (in the current case, age effects are the same as cohort effects). More specifically, the explanatory variables are augmented by a set of cohort dummies. Let $w = (w_2, \dots, w_{10})$ be a set of indicator variables representing each of the birth cohort from 1931 to 1939 (the omitted category is individuals born in 1930). Combined with the time dummies, the set of cohort dummy variables should capture the age effects.

Compared to the basic model, the expanded one is different only in the presence of the cohort dummies in the explanatory variables. Re-define the variable

$$\Delta_t \equiv \Delta_t(\gamma, \delta_t, \rho) = \gamma y_{t-2} - \delta_t + w' \rho$$

where ρ is a 9×1 vector of parameters on the age dummies. Now we have altogether $T - 2 + 9$ parameters: $T - 2$ for the time dummies (same as before) and 9 for the age dummies. For fixed value of γ_0 , the model can be estimated using a set of moment conditions which correspond to the first order conditions of a well-behaved objective function. Specifically, let $m_t(\gamma_0, \delta_t, \rho) =$

³⁶The number of observations for the unbalanced panel sample is 474 and 3,089 for the black and white, respectively.

$1\{S\} \cdot \text{sgn}(y_{t-1} - \Delta_t(\gamma_0, \delta_t, \rho) - \tilde{y}_t(\gamma_0))$, then the moments are

$$\begin{aligned} E[m_t(\gamma_0, \delta_t, \rho)] &= 0 \text{ for } t = 3, \dots, T \\ \sum_{t=3}^T E[m_t(\gamma_0, \delta_t, \rho) \cdot w_k] &= 0 \text{ for } k = 2, \dots, 10. \end{aligned} \quad (12)$$

These moments are, loosely speaking, the first order conditions of the minimization of a well-behaved objective function $\sum_{t=3}^T E[Q_t(\gamma_0, \delta_t, \rho)]$, where $Q_t(\cdot)$ is defined in a similar fashion as in section 6.2 with the term $\gamma y_{t-2} - \delta_t + w' \rho$ replacing $\gamma y_{t-2} - \delta_t$ everywhere. After obtaining the estimates $(\hat{\delta}, \hat{\rho})$, we can plug them into the sample analog of the moments

$$\begin{aligned} E[h_t(\gamma_0, \delta_t, \rho)] &= 0 \text{ for } t = 3, \dots, T \\ \sum_{t=3}^T E[h_t(\gamma_0, \delta_t, \rho) \cdot w_k] &= 0 \text{ for } k = 2, \dots, 10 \end{aligned} \quad (13)$$

where $h_t(\gamma_0, \delta_t, \rho) = 1\{S\} \cdot \text{sgn}(y_{t-1} - \Delta_t(\gamma_0, \delta_t, \rho) - \tilde{y}_t(\gamma_0)) \cdot y_{t-2}$. The sample analog of the moments (13) will be asymptotically distributed as $N(0, V)$. With a consistent estimator of the variance-covariance matrix V , we can compute the quadratic form test-statistic, denoted as Test A, which is asymptotically distributed as $\chi^2(T - 2 + 9)$. Similarly, we can obtain estimates $(\hat{\delta}, \hat{\rho})$ from (13), plug them into (12), and compute another quadratic form test-statistic based on the distribution of the sample analog, denoted as Test B, which should also be asymptotically distributed as $\chi^2(T - 2 + 9)$. The results are plotted in Figure 11a and 11b.³⁷

³⁷A bandwidth of 0.1 is used.

Figure 11a: Test-Statistics A and B with Age Effects (Unbalanced Panel - Black Sample)

Figure 11b: Test-Statistics A and B with Age Effects (Unbalanced Panel - White Sample)

Compared to the results in Figure 10a and 10b, the inclusion of age effects does not seem to change the confidence intervals for the AR(1) coefficient much. Again, the estimate of γ for the black sample is less precise and lower than that for the white sample, suggesting less persistency in the individual earnings process for blacks. The small difference between the results from the two model specifications (with and without age effects in addition to time dummies) indicates that for the relatively narrowly defined 10-year birth cohort, the earnings-age profile does not seem to exhibit strong nonlinearity.

8 Conclusion

This paper makes a contribution to the literature on the estimation of nonlinear dynamic panel data models. Specifically, I propose a method for estimating a censored dynamic panel data model with individual fixed effects and a lagged *latent* dependent variable. Censoring destroys a certain symmetry between the latent variables. But we can artificially truncate the observations in such a way that the symmetry is restored. Based on the restored symmetry, orthogonality conditions are constructed and GMM estimation is implemented.

As an empirical application, I use the matched data from 1973 and 1978 March CPS and Social Security Administration Earnings Records to estimate the earnings dynamics for a sample of men living in the South from 1957-1973 by race. I construct asymptotic confidence intervals for the autoregressive parameter γ . The main empirical finding is that although the AR(1) parameter estimated by linear GMM does not seem to differ by race, the white men's earnings process appears to be more persistent than black men's (conditional on individual heterogeneity) when censoring is taken into account.

In general, our knowledge of the nonlinear (dynamic) panel models is still limited. The main challenge is the presence of the unobserved individual-specific effects. The approach in this paper is to “difference away” the effects. Alternatively, one could take the random effects approach. The latter often requires one to make distributional assumptions on the underlying data process and to deal with the initial conditions problem. As a result, the random effects approach might be subject to misspecification errors. As another alternative, one could treat the individual-specific effects as parameters to be estimated. This approach suffers from the classical incidental parameters

problem. For the linear dynamic panel data models, this approach yields inconsistent estimators and the asymptotic bias is of order $O(1/T)$ (see Nickell (1981) for the theoretical proof). For the censored dynamic panel data models, combining this approach with, for example, the Censored Least Absolute Deviations (CLAD) estimator developed in Powell (1984) suggests that we can estimate the basic model (without age effects) by

$$\begin{aligned} & \min_{\gamma, \delta, \alpha} \sum_{i=1}^n \sum_{t=2}^{T+1} 1\{y_{it-1} > 0\} \cdot |y_{it} - \max(0, \gamma y_{it-1} + \delta_t + \alpha_i)| \\ &= \min_{\gamma, \delta} \sum_{i=1}^n \min_{\alpha_i} \sum_{t=2}^{T+1} 1\{y_{it-1} > 0\} \cdot |y_{it} - \max(0, \gamma y_{it-1} + \delta_t + \alpha_i)|. \end{aligned} \quad (14)$$

I implement this method using the balanced panel from Section 6 and obtain the estimates $\hat{\gamma} = 0.499$ (S.E.=0.031) for the black men sample and $\hat{\gamma} = 0.572$ (S.E.=0.019) for the white men sample.³⁸ I also implement this method using 8 randomly selected time periods for each individual. This yields $\hat{\gamma} = 0.540$ (S.E.=0.052) for the black men sample and $\hat{\gamma} = 0.611$ (S.E.=0.025) for the white men sample. The difference between the estimates based on $T = 16$ (the full sample) and $T = 8$ suggests that the incidental parameters problem is important. If one assumes that the bias resulting from this method is of the same order here as in the linear model, then one could jackknife the estimators to get an asymptotically unbiased estimator. In the example here, the estimates would be $\tilde{\gamma} = 0.46$ for the black and $\tilde{\gamma} = 0.53$ for the white. Investigating whether such an approach could be justified is an interesting topic for future research.

³⁸Note that (14) can be written as $\min_{\gamma, \delta} \sum_{i=1}^n q(\gamma, \delta)$ where $q(\gamma, \delta)$ is the minimized function value of the second summation for a given parameter value of (γ, δ) . The estimator of (γ, δ) can therefore be thought of as an extreme estimator, and it is thus easy to compute the asymptotic variance of the estimator.

9 Appendix

9.1 Appendix Table

Appendix Table 1						
Estimates of AR(1) Coefficient from Censored Regression Models by Selected Year: Dependent Variable is Log Real Annual Earnings (in Thousands)						
Year	Black			White		
	OLS	Tobit (MLE)	CLAD*	OLS	Tobit (MLE)	CLAD*
1960	0.493 (0.036)	0.704 (0.057)	0.682 (0.147)	0.542 (0.019)	0.727 (0.020)	0.748 (0.034)
1965	0.802 (0.042)	0.930 (0.045)	0.879 (0.073)	0.695 (0.029)	1.119 (0.035)	0.738 (0.068)
1970	0.990 (0.061)	1.156 (0.063)	1.033 (0.076)	0.850 (0.022)	1.116 (0.026)	0.913 (0.056)

*Bootstrap standard errors (#replications=3,000).

9.2 Appendix Figures

Appendix Figure 1

Appendix Figure 2a: Test-Statistics A and B for the Linear GMM (Black Sample)

Appendix Figure 2b: Test-Statistics A and B for the Linear GMM (White Sample)

9.3 Derivation of Proposition 3

Define the estimator of the parameter δ , whose dimension is $(T - 2) \times 1$, by the solutions to the $T - 2$ moment conditions

$$\frac{1}{n} \sum m(\hat{\delta}, \gamma_o) = 0.$$

By the standard GMM argument (see, for example, Newey and McFadden (1994)), we have

$$\sqrt{n}(\hat{\delta} - \delta_o) \stackrel{LD}{=} -M^{-1} \cdot \frac{1}{\sqrt{n}} \sum m(\delta_o, \gamma_o)$$

where $M = \frac{\partial E m}{\partial \delta}$.

It is then straightforward to show that, after plugging the estimate $\hat{\delta}$ into the $T - 2$ moment conditions h , we have

$$\begin{aligned} & \sqrt{n} \cdot \frac{1}{n} \sum h(\hat{\delta}, \gamma_o) \\ & \stackrel{LD}{=} \frac{1}{\sqrt{n}} \sum h(\delta_o, \gamma_o) + H \cdot \sqrt{n}(\hat{\delta} - \delta_o) \\ & \stackrel{LD}{=} \frac{1}{\sqrt{n}} \sum h(\delta_o, \gamma_o) - HM^{-1} \frac{1}{\sqrt{n}} \sum m(\delta_o, \gamma_o) \\ & \stackrel{LD}{=} N(0, V) \end{aligned}$$

where

$$V = E(h - HM^{-1}m)(h - HM^{-1}m)'$$

and

$$H = \frac{\partial E h}{\partial \delta}.$$

Therefore we have

$$n \left(\frac{1}{n} \sum h(\hat{\delta}, \gamma_o) \right)' V^{-1} \left(\frac{1}{n} \sum h(\hat{\delta}, \gamma_o) \right) \stackrel{A}{\approx} \chi^2(T - 2).$$

Q.E.D.

9.4 Derivation of Proposition 4

Let $\Gamma_1 = \frac{\partial E m}{\partial \beta}$, $\Gamma_2 = \frac{\partial E h}{\partial \beta}$, and $\begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} = \begin{pmatrix} E(mm') & E(mh') \\ E(hm') & E(hh') \end{pmatrix}$. Note that

$$V \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \left[\begin{pmatrix} \Gamma_1' & 0 \\ 0 & \Gamma_2' \end{pmatrix} \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}^{-1} \begin{pmatrix} \Gamma_1 & 0 \\ 0 & \Gamma_2 \end{pmatrix} \right]^{-1}$$

$$\begin{aligned}
&= \begin{pmatrix} \Gamma_1' V^{11} \Gamma_1 & \Gamma_1' V^{12} \Gamma_2 \\ \Gamma_2' V^{21} \Gamma_1 & \Gamma_2' V^{22} \Gamma_2 \end{pmatrix}^{-1} \\
&= \begin{pmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{pmatrix}^{-1}.
\end{aligned}$$

Let

$$\beta_{opt} = A \cdot \beta_1 + (I - A) \cdot \beta_2 = \begin{pmatrix} A & I - A \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

where $A = k^{-1} [\Sigma^{11} + \Sigma^{21}]$ and $\kappa = [\Sigma^{11} + \Sigma^{12} + \Sigma^{21} + \Sigma^{22}]$, then

$$\begin{aligned}
V(\beta_{opt}) &= \begin{pmatrix} A & I - A \end{pmatrix} V \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \begin{pmatrix} A \\ I - A \end{pmatrix} \\
&= \begin{pmatrix} A & I - A \end{pmatrix} \begin{pmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{pmatrix} \begin{pmatrix} A \\ I - A \end{pmatrix} \\
&= k^{-1} \left\{ \left[(\Sigma^{11} + \Sigma^{21}) \Sigma_{11} + (\Sigma^{12} + \Sigma^{22}) \Sigma_{21} \right] \cdot (\Sigma^{11} + \Sigma^{21}) \right. \\
&\quad \left. + \left[(\Sigma^{11} + \Sigma^{21}) \Sigma_{12} + (\Sigma^{12} + \Sigma^{22}) \Sigma_{22} \right] \cdot (\Sigma^{12} + \Sigma^{22}) \right\} k^{-1} \\
&= k^{-1} C k^{-1}.
\end{aligned}$$

We will demonstrate that $V(\beta_{opt})$ equals the variance of the optimal GMM estimator. The latter is

$$\begin{aligned}
V(\beta_{GMMOpt}) &= \left[\begin{pmatrix} \Gamma_1' & \Gamma_2' \end{pmatrix} \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}^{-1} \begin{pmatrix} \Gamma_1 \\ \Gamma_2 \end{pmatrix} \right]^{-1} \\
&= \left[\Gamma_1' V^{11} \Gamma_1 + \Gamma_1' V^{12} \Gamma_2 + \Gamma_2' V^{21} \Gamma_1 + \Gamma_2' V^{22} \Gamma_2 \right]^{-1} \\
&= \left[\Sigma^{11} + \Sigma^{12} + \Sigma^{21} + \Sigma^{22} \right]^{-1} \\
&= \kappa^{-1}.
\end{aligned}$$

It therefore suffices to show that $C = \kappa$.

By definition of the inverse of a matrix,

$$\Sigma \begin{pmatrix} \Sigma^{11} \\ \Sigma^{21} \end{pmatrix} = \begin{pmatrix} I \\ 0 \end{pmatrix}$$

and

$$\Sigma \begin{pmatrix} \Sigma^{12} \\ \Sigma^{22} \end{pmatrix} = \begin{pmatrix} 0 \\ I \end{pmatrix}$$

It is then straightforward to verify that

$$(\Sigma^{11} + \Sigma^{21}) \Sigma_{11} + (\Sigma^{12} + \Sigma^{22}) \Sigma_{21} = (\Sigma^{11} \Sigma_{11} + \Sigma^{12} \Sigma_{21}) + (\Sigma^{21} \Sigma_{11} + \Sigma^{22} \Sigma_{21}) = I + 0 = I$$

and

$$\left(\Sigma^{11} + \Sigma^{21}\right) \Sigma_{12} + \left(\Sigma^{12} + \Sigma^{22}\right) \Sigma_{22} = \left(\Sigma^{11} \Sigma_{12} + \Sigma^{12} \Sigma_{22}\right) + \left(\Sigma^{21} \Sigma_{12} + \Sigma^{22} \Sigma_{22}\right) = 0 + I = I.$$

It then immediately follows that $C = (\Sigma^{11} + \Sigma^{21}) + (\Sigma^{12} + \Sigma^{22}) = \kappa$. Q.E.D.

9.5 Illustration of the Relationship Between the Moment Conditions of Plain Difference and the Quadratic Objective Function

In the text we discussed how the moment conditions based on (8) with $\xi(e_1, e_2) = \text{sgn}(e_1 - e_2)$ can be trivially satisfied at the parameter $\delta = \pm\infty$. The same is true with $\xi(e_1, e_2) = e_1 - e_2$. Appendix Figure 3 below illustrates this using the same data as were used for Figure 4.

Appendix Figure 3: Plain Difference Moment Condition as Function of δ when $T = 3$ and $\gamma = 0.5$

The moment condition has the form $E[g(\lambda_0)] = 0$ where

$$g(\lambda) = 1 \left\{ -z_2 \leq \lambda \leq \frac{z_1 + \gamma z_2}{1 - \gamma} \right\} \cdot (z_1 - z_2 - \lambda).$$

By integrating this, we can recover an objective function such that the first order conditions for minimizing it will be the moment conditions $E[g(\lambda_0)] = 0$. Therefore, instead of defining the estimator of λ as the value that satisfies $\frac{1}{n} \sum_{i=1}^n g_i(\tilde{\lambda}) = 0$, we define estimator $\bar{\lambda}$ as the minimizer

of an objective function $\frac{1}{n} \sum_{i=1}^n G_i(\bar{\lambda})$, where

$$G(\lambda) = \begin{cases} z_1^2 & \text{if } \lambda < -z_2 \\ (z_1 - z_2 - \lambda)^2 & \text{if } -z_2 \leq \lambda \leq \frac{z_1 + \gamma z_2}{1 - \gamma} \\ \left(\frac{z_2 + \gamma z_1}{1 - \gamma}\right)^2 & \text{if } \frac{z_1 + \gamma z_2}{1 - \gamma} < \lambda. \end{cases}$$

The function $G(\lambda)$ is continuous and almost everywhere differentiable.

Specifically, for a fixed value of γ_0 , define the estimator of δ_t as

$$\bar{\delta}_t = \arg \min_{\delta_t} R_n(\delta_t)$$

where

$$\begin{aligned} R_n(\delta_t) &= \frac{1}{n} \sum_{i=1}^n 1 \{ (y_{it}, y_{it-1}, y_{it-2}) > 0 \} \\ &\cdot \left\{ [y_{it-1} - \tilde{y}_t(\gamma_0) - (\gamma_0 y_{it-2} - \delta_t)]^2 \cdot 1 \left\{ -\tilde{y}_t(\gamma_0) < \gamma_0 y_{it-2} - \delta_t < \frac{\gamma_0 \tilde{y}_t(\gamma_0) + y_{it-1}}{1 - \gamma_0} \right\} \right. \\ &+ y_{it-1}^2 \cdot 1 \{ \gamma_0 y_{it-2} - \delta_t < -\tilde{y}_t(\gamma_0) \} \\ &\left. + \left[\frac{\tilde{y}_t(\gamma_0) + \gamma_0 y_{it-1}}{1 - \gamma_0} \right]^2 \cdot 1 \left\{ \gamma_0 y_{it-2} - \delta_t > \frac{\gamma_0 \tilde{y}_t(\gamma_0) + y_{it-1}}{1 - \gamma_0} \right\} \right\}. \end{aligned}$$

Again, using the white men sample of 1965-67 and fixing $\gamma = 0.5$, the function $R_n(\delta_t)$ is plotted in Appendix Figure 4 below.

Appendix Figure 4: Objective Function $R_n(\delta_t)$ as a Function of δ when $T = 3$ and $\gamma = 0.5$

References

- [1] Abowd, J. and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57, 411-445.
- [2] Anderson, T. and C. Hsiao (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47-82.
- [3] Arellano, M. and O. Bover (1997): "Estimating Dynamic Limited Dependent Variable Models From Panel Data," *Investigaciones Economicas*, 21, 141-65.
- [4] Arellano, M., O. Bover, and J. Labeaga (1997): "Autoregressive Models with Sample Selectivity for Panel Data," Working Paper No. 9706, CEMFI.
- [5] Arellano, M. and B. Honoré (1999): "Panel Data Models: Some Recent Developments," unpublished manuscript, Princeton University.
- [6] Ashenfelter, O. (1978): "Estimating the Effect of Training Programs on Earnings," *Review of Economics and Statistics*, 60, 47-57.
- [7] Bertrand, M. (1997): "The Microstructure of the Phillips Curve," unpublished manuscript, Harvard University.
- [8] Bound, J. and A. Krueger (1991): "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics*, 9, 1-24.
- [9] Card, D. and D. Sullivan (1988): "Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment," *Econometrica*, 56, 497-530.
- [10] Chay, K. (1996): "An Empirical Analysis of Black Economic Progress Over Time," Ph.D. dissertation, Princeton University.
- [11] Chay, K. and B. Honoré (1998): "Estimation of Semiparametric Censored Regression Models: An Application to Changes in Black-White Earnings Inequality During the 1960s," *Journal of Human Resources*, 33, 5-38.
- [12] Deaton, A. (1992): *Understanding Consumption*, New York: Oxford University Press.

- [13] Deaton, A. (1997): "The Analysis of Household Surveys: A Microeconometric Approach to Development Policy," Baltimore: Johns Hopkins University Press.
- [14] Farber, H. and R. Gibbons (1996): "Learning and Wage Dynamics," *Quarterly Journal of Economics*, 111, 1007-47.
- [15] Hansen, L., J. Heaton, and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262-80.
- [16] Heckman, J. (1981): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time – Discrete Data Stochastic Process," in *Structural Analysis of Discrete Panel Data with Econometric Applications*, ed. by C. Manski and D. McFadden, Cambridge: MIT Press.
- [17] Hirano, K. (1999): "Semiparametric Bayesian Models for Dynamic Earnings Data," unpublished manuscript, UCLA.
- [18] Honoré, B. (1992): "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60, 533-65.
- [19] Honoré, B. (1993a): "Efficiency Considerations in Tobit Models with Fixed Effects," in *Panel Data and Labour Market Dynamics*, ed. by P. Jensen and N. Westergaard-Nielsen, Amsterdam: North-Holland.
- [20] Honoré, B. (1993b): "Orthogonality Conditions for Tobit Models with Fixed Effects and Lagged Dependent Variables," *Journal of Econometrics*, 59, 35-61.
- [21] Honoré, B. and L. Hu (1999): "Estimation of Censored Regression Models with Endogeneity," unpublished manuscript, Princeton University.
- [22] Honoré, B., E. Kyriazidou, and C. Udry (1997): "Estimation of Type 3 Tobit Models Using Symmetric Trimming and Pairwise Comparisons," *Journal of Econometrics*, 76, 107-128.
- [23] Honoré, B. and J. Powell (1994): "Pairwise Difference Estimators of Censored and Truncated Regression Models," *Journal of Econometrics*, 64, 241-78.

- [24] Hoynes, H. (1999): "The Employment, Earnings, and Income of Less Skilled Workers Over the Business Cycle," Working Paper, Northwestern University and University of Chicago Joint Center for Poverty Research.
- [25] Kyriazidou, E. (1997): "Estimation of a Panel Data Sample Selection Model," *Econometrica*, 65, 1335-1364.
- [26] Kyriazidou, E. (1999): "Estimation of Dynamic Panel Data Sample Selection Models," unpublished manuscript, UCLA.
- [27] Maddala, G. (1971): "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," *Econometrica*, 39, 341-358.
- [28] Maddala, G. (1987): "Limited Dependent Variable Models Using Panel Data," *Journal of Human Resources*, 22, 307-38.
- [29] Nerlove, M. (1967): "Experimental Evidence on the Estimation of Dynamic Economic Relationships from a Time Series of Cross-Sections," *Economic Studies Quarterly*, 18, 42-74.
- [30] Nerlove, M. (1971): "Further Evidence on the Estimation of Dynamic Economic Relationships from a Time Series of Cross-Sections," *Econometrica*, 39, 359-387.
- [31] Newey, W. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics*, ed. by R. Engle and D. McFadden, Volume 4, 2111-2245, Amsterdam, London and New York: Elsevier, North-Holland.
- [32] Neyman, J. and E. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1-32.
- [33] Nickell, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417-26.
- [34] Pakes, A. and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027-57.
- [35] Powell, J. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303-25.

- [36] Powell, J. (1986): "Symmetrically Trimmed Least Squares Estimation for Tobit Models," *Econometrica*, 54, 1435-60.