THESIS
2
7

LIBRARY
Michigan State
University

This is to certify that the
dissertation entitled

PANEL DATA MODELS WITH UNOBSERVED EFFECTS
AND ENDOGENOUS EXPLANATORY VARIABLES

presented by

IRINA MURTAZASHVILI

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Economics

Jeffrey M. Wooldridge
Major Professor's Signature

July 18, 2007
Date

MSU is an affirmative-action, equal-opportunity employer

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
072409		
JUL 2 7 2010		

6/07 p:/CIRC/DateDue indd-p.1

PANEL DATA MODELS WITH UNOBSERVED EFFECTS AND ENDOGENOUS EXPLANATORY VARIABLES

By

Irina Murtazashvili

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

2007

ABSTRACT

PANEL DATA MODELS WITH UNOBSERVED EFFECTS AND ENDOGENOUS EXPLANATORY VARIABLES

By

Irina Murtazashvili

This dissertation consists of three essays that address issues of estimation in panel data models with unobserved effects and endogenous explanatory variables. The first essay considers estimation of correlated random coefficient (CRC) panel data models with endogenous regressors. This chapter provides a set of conditions sufficient for consistency of a general class of fixed effects instrumental variables (FE-IV) estimators in the context of a CRC panel data model. The usual FE-IV estimator turns out to be fairly robust to the presence of neglected individual-specific slopes. Monte Carlo simulations suggest the proposed FE-IV estimator of Population Averaged Effect (PAE) provided a full set of period dummy variables is included performs better than other estimators in finite samples for the case of (roughly) continuous endogenous explanatory variables.

The second essay continues studying a CRC panel data model from the first chapter but, in addition to allowing some explanatory variables to be correlated with the idiosyncratic error, the joint distribution of the endogenous regressors and the individual heterogeneity conditional on the instruments is allowed to depend on the instruments. The second essay uses a two-step control function approach to account for endogeneity and to consistently estimate average partial effects (APEs) in CRC panel data models with endogenous roughly continuous regressors.The simulation findings indicate that in the finite samples the control function approach to estimating the CRC balanced panel data model with time-constant individual heterogeneity performs better than other estimators under the considered conditions. The pro-

posed method is applied to the problem of estimating the APEs of annual hours of on-job-training on output scrap rates for manufacturing firms in Michigan.

In the third essay, a dynamic binary response panel data model that allows for an endogenous regressor is developed. This estimation approach is of particular value for settings in which one wants to estimate the effects of a treatment which is also endogenous. This model is applied to examine the impact of rural-urban migration on the likelihood that households in rural China fall below the poverty line. The empirical results that migration is important for reducing the likelihood that poor households remain in poverty and that non-poor households fall into poverty. Further, failure to control for unobserved heterogeneity leads to an overestimate of the impact of migrant labor markets on probability of staying poor of those who lived below the poverty lines.

Copyright by

Irina Murtazashvili

2007

ACKNOWLEDGMENTS

I would like to express the deepest appreciation to my adviser, Professor Jeffrey Wooldridge for his generous advice and support. Without his guidance and help this dissertation would not have been possible. I am very grateful for the assistance and advice I received from Professor John Giles who also kindly provided me the data for one of the applications. I wish to thank my committee members, Professor Peter Schmidt, Professor Ana Maria Herrera, and Professor David Tschirley, for valuable comments and fruitful discussions. I also thank other faculty members and doctoral students of the Department of Economics at Michigan State University for support during my graduate studies.

TABLE OF CONTENTS

LIST OF TABLES viii

LIST OF FIGURES ix

1 FIXED EFFECTS INSTRUMENTAL VARIABLES ESTIMATION IN CORRELATED RANDOM COEFFICIENT PANEL DATA MODELS 1

1.1 Introduction 1
1.2 Model Specification and Previous Results 3
1.3 Conditions for Consistent FE-IV Estimation 6
1.4 Examples 10
1.5 Finite Sample Behavior of the FE-IV Estimator 13
1.6 Conclusion 18

2 A CONTROL FUNCTION APPROACH TO ESTIMATION OF CORRELATED RANDOM COEFFICIENT PANEL DATA MODELS 19

2.1 Introduction 19
2.2 Model of Interest for Balanced Panels 21
2.3 Estimating Procedure and Calculation of Standard Errors 28
2.4 Finite Sample Behavior of the Control Function Estimator 32
2.5 Empirical Application to Effects of Job Training on Worker Productivity 40
2.6 Conclusion 48

3 ESTIMATION OF A DYNAMIC BINARY RESPONSE PANEL DATA MODEL WITH AN ENDOGENOUS REGRESSOR, WITH AN APPLICATION TO THE ANALYSIS OF POVERTY PERSISTENCE IN RURAL CHINA 50

3.1 Introduction 50
3.2 Estimation of a Dynamic Binary Response Panel Data Model with an Endogenous Regressor 53
3.2.1 Dynamic Binary Response Panel Data Models 53
3.2.2 A General Approach to Estimation 56
3.2.3 Allowing for Serial Correlation of Errors in the First Stage . . 61
3.2.4 Calculation of Average Partial Effects 62
3.3 Migrant Labor Markets and Poverty Persistence in Rural China . . . 64
3.3.1 Rural-Urban Migration in China 64
3.3.2 The RCRE Household Survey 66

3.3.3 Migration, Consumption Growth and Poverty 68
3.3.4 Estimating the Impact of Migrant Labor Markets on Poverty Persistence 70
3.3.5 Identifiying the Migrant Network 73
3.4 Results 76
3.5 Conclusions 78

APPENDICES **80**

A Tables for Chapter 1 **80**

B Tables for Chapter 2 **85**

C Tables and Figures for Chapter 3 **95**

BIBLIOGRAPHY **105**

LIST OF TABLES

A.1 **Usual Unobserved Effects CRC Model for $\beta = 2$ and $T = 5$** . . 81

A.2 **Usual Unobserved Effects CRC Model for $\beta = 2$ and $T = 10$** . 82

A.3 **Random Trend CRC Model for $\beta = 2$ and $T = 5$** 83

A.4 **Random Trend CRC Model for $\beta = 2$ and $T = 10$** 84

B.1 **Usual Unobserved Effect CRC Model for Continuous y_{2it}** . . 86

B.2 **Random Trend CRC Model for Continuous y_{2it}** 87

B.3 **Usual Unobserved Effect CRC Model for $y_{2it} \in (0, 1)$** 88

B.4 **Random Trend CRC Model for $y_{2it} \in (0, 1)$** 89

B.5 **Standard Errors for the Control Function Approach** 90

B.6 **Summary Statistics from Unbalanced and Balanced Datasets** 91

B.7 **POLS Estimates of the First Stage Regressions** 92

B.8 **FE-IV and CF Estimates of the Second Stage Regressions** . . 93

B.9 **Summary Statistics for the Control Variables** 94

C.1 **Household and Village Characteristics** 100

C.2 **Factors Determining the Size of the Village Migrant Network** 101

C.3 **CF Approach to Estimating Determinants of Poverty Status** 102

C.4 **Linear Probability Model for Determinants of Poverty Status** 103

C.5 **Average Partial Effects of Determinants of Poverty Status** . . 104

LIST OF FIGURES

C.1 **Share of Village Labor Force Employed as Migrants by Year** 96

C.2 **Village Consumption Growth** 97

C.3 **Change in Poverty Headcount** 98

C.4 **Change in Out-Migrants in Village Labor Force** 99

CHAPTER 1

FIXED EFFECTS INSTRUMENTAL VARIABLES ESTIMATION IN CORRELATED RANDOM COEFFICIENT PANEL DATA MODELS

1.1 Introduction

In both cross section and panel data settings, there is substantial interest in estimating population averaged effects (PAEs), including average treatment effects (ATEs), in the correlated random coefficient (CRC) model. Models with both exogenous explanatory variables and endogenous regressors have been investigated in recent years. Angrist (1991) discusses the conditions for consistency of ATE estimates in models with binary endogenous variables and no exogenous covariates. A set of sufficient assumptions required for consistent ATE estimates with (roughly) continuous endogenous regressors in a CRC model can be found in Wooldridge (2003). Both papers study estimation with random sampling from a cross section.

The possibility that treatment effects might depend on individual-specific heterogeneity motivated Imbens and Angrist (1994) to introduce the "local average treatment effect" (LATE) as an evaluation parameter, which provides a useful interpre-

tation of the instrumental variables estimator when the effect of a binary treatment varies across units. That emphasis on LATE led to a reinterpretation of IV estimates in many empirical applications, and spurred a great deal of research on interpreting IV estimators in a variety of contexts. Heckman and Vytlacil (2005) provide a recent unification, including a discussion of whether we should be interested in parameters such as LATE.

The understanding that IV generally consistently estimates LATE in simple settings is useful, but often we are interested in estimating the expected effect for a randomly drawn unit from the underlying population. Plus, strict interpretation of LATE as the average treatment effect among units induced into treatment by the switching of an instrumental variable – such as program eligibility – is limited to special cases. Here we study estimation of population average effects, or average treatment effects, in a general panel data model with heterogeneous slopes. By estimating population average effects we can easily estimate the aggregate effects of various policies, such as increasing the amount of job training among the population of manufacturing workers.

Wooldridge (2005a) studied general fixed effects estimators with strictly exogenous regressors in the CRC model with panel data, and derived conditions under which generalized fixed effects estimators – generalized in the sense that they sweep away unit-specific trends – are consistent for the population averaged effect. In this paper, we study the model in Wooldridge (2005a) but, in addition to allowing correlation between the instruments and the unobserved heterogeneity, we allow some explanatory variables to be correlated with the idiosyncratic error. The main result is a set of sufficient conditions under which fixed effects instrumental variables (FE-IV) estimators consistently estimate the population averaged effect, even when the individual-specific slopes are ignored. The results include the commonly used fixed effects two stage least squares estimator (FE-2SLS) as a special case, but also

more general FE-IV estimators that sweep away individual-specific time trends. The conditions are most likely to apply when the endogenous explanatory variables are at least roughly continuous, as in Wooldridge (2003) for the cross-sectional case.

The remainder of the paper is organized as follows. In Section 1.2 we introduce the model and briefly review existing results. Section 1.3 contains the main consistency result, and Section 1.4 covers examples where the conditions will – and will not – hold. Section 1.5 contains a Monte Carlo study that shows how the FE-IV estimator, with a fully set of time period dummies, outperforms its obvious competitors. The simulation results support the results in Sections 1.3 and 1.4. Section 1.6 contains a brief conclusion.

1.2 Model Specification and Previous Results

The model of interest is a CRC model studied in Wooldridge (2005a). For a random draw i from the population, the model is

$$y_{it} = \mathbf{w}_t \mathbf{a}_i + \mathbf{x}_{it} \mathbf{b}_i + u_{it}, t = 1, ..., T, \tag{1.1}$$

where y_{it} is a dependent variable, $\mathbf{w}_t$ is a $1 \times J$ vector of aggregate time variables, which we treat as nonrandom, $\mathbf{a}_i$ is a $J \times 1$ vector of individual-specific slopes on the aggregate variables, $\mathbf{x}_{it}$ is a $1 \times K$ vector of endogenous covariates that change across time, $\mathbf{b}_i$ is a $K \times 1$ vector of individual-specific slopes, and u_{it} is an idiosyncratic error. As discussed in Wooldridge (2005a), we require $J < T$. So, if we have two time periods, we can only allow a scalar individual-specific intercept, a_i. If $T = 3$, we can allow individual-specific linear trends, too. Higher order trend terms are allowed as T increases.

Equation (1.1) is a correlated random coefficients model when the individual specific slopes, $\mathbf{b}_i$ (as well as the elements in $\mathbf{a}_i$), are allowed to be correlated with $\mathbf{x}_{it}$.

For example, a simple CRC wage equation might look like

$$\log(wage_{it}) = a_{i1} + a_{i2}t + b_{i1}training_{it} + b_{i2}union_{it} + b_{i3}married_{it} + u_{it}, \quad (1.2)$$

where, in addition to the standard level effect a_{i1}, each individual is allowed to have his or her own unobserved growth in wages, a_{i2}. In addition, the time-varying explanatory variables have individual-specific returns. The variable *training* might be hours spent in job training, and the CRC model allows the return to training to be individual-specific and correlated with the amount of training – as a standard model of human capital accumulation would suggest.

Wooldridge (2005a) studied the consistency of fixed effects estimators of (1.1) that sweep out the $\mathbf{a}_i$ but act *as if* $\mathbf{b}_i = \beta$ for all i. To describe Wooldridge's main result, and the extension here, write $\mathbf{b}_i = \beta + \mathbf{d}_i$, and substitute into (1.1):

$$y_{it} = \mathbf{w}_t\mathbf{a}_i + \mathbf{x}_{it}\beta + (\mathbf{x}_{it}\mathbf{d}_i + u_{it}) \equiv \mathbf{w}_t\mathbf{a}_i + \mathbf{x}_{it}\beta + v_{it}, \quad (1.3)$$

where $v_{it} \equiv \mathbf{x}_{it}\mathbf{d}_i + u_{it}$. We eliminate $\mathbf{a}_i$ by regressing, for each i, y_{it} on $\mathbf{w}_t$, $t = 1, ..., T$ and $\mathbf{x}_{it}$ on $\mathbf{w}_t$, $t = 1, ..., T$, and keeping the residuals, $\ddot{y}_{it}$ and $\ddot{\mathbf{x}}_{it}$, respectively. This gives the equations

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\mathbf{b}_i + \ddot{u}_{it} = \ddot{\mathbf{x}}_{it}\beta + (\ddot{\mathbf{x}}_{it}\mathbf{d}_i + \ddot{u}_{it}) = \ddot{\mathbf{x}}_{it}\beta + \ddot{v}_{it}, t = 1, ..., T. \quad (1.4)$$

The fixed effects estimator studied by Wooldridge (2005a) is just the pooled OLS estimator from (1.4). We control the amount of individual-specific detrending by choosing $\mathbf{w}_t$ appropriately.

An assumption used by Wooldridge (2005a) is the standard strict exogeneity assumption conditional on $(\mathbf{a}_i, \mathbf{b}_i)$:

$$\mathrm{E}(u_{it}|\mathbf{x}_{i1}, ..., \mathbf{x}_{iT}, \mathbf{a}_i, \mathbf{b}_i) = 0, t = 1, ..., T. \quad (1.5)$$

Using a simple iterated expectations argument, Wooldridge shows that, under the additional assumption

$$\mathrm{E}(\mathbf{b}_i|\ddot{\mathbf{x}}_{it}) = \mathrm{E}(\mathbf{b}_i), t = 1, ..., T, \tag{1.6}$$

the fixed effects estimator is consistent for the population averaged effect, β.

Consistency of the usual FE estimator relies heavily on assumption (1.5), which rules out traditional simultaneity, time-varying measurement error, correlation between time-varying omitted factors (in u_{it}) and the elements of $\mathbf{x}_{it}$, and models with lagged dependent variables or other kinds of regressors where changes in u_{it} may feed back into changes in $\mathbf{x}_{i,t+h}$ for $h \geq 1$. In the case where $\mathbf{b}_i = \beta$, methods that first eliminate $\mathbf{a}_i$ and then apply instrumental variables – usually, 2SLS – have become a standard tool for the applied economist. Here, we study such estimators but allow for individual-specific slopes, $\mathbf{b}_i$.

Let $\mathbf{z}_{it}$ be a $1 \times L$ vector of instrumental variables, with $L \geq K$. Let $\ddot{\mathbf{z}}_{it}$ be the "detrended" instruments from the individual-specific regressions of $\mathbf{z}_{it}$ on $\mathbf{w}_t$, $t = 1, ..., T$. Then we can estimate (1.4) using instruments $\ddot{\mathbf{z}}_{it}$ for unit i in time period t. Whether we just use pooled 2SLS – the estimator we focus on here – or a more sophisticated generalized method of moments (GMM) estimator, the moment conditions we use are

$$\mathrm{E}(\ddot{\mathbf{z}}_{it}'\ddot{v}_{it}) = \mathbf{0}, t = 1, ..., T. \tag{1.7}$$

In the next section, we study consistency of the FE-2SLS estimator under conditions that relax those in Wooldridge (2005a).

1.3 Conditions for Consistent FE-IV Estimation

In order to ensure that (1.7) holds, we place conditions separately on the relationship between the instruments and idiosyncratic errors and the instruments and the unobserved effects. Plus, of course, there is always a standard rank condition.

ASSUMPTION 1: With the definitions in Section 1.2,

$$\mathrm{E}(u_{it}|\mathbf{z}_{i1}, \mathbf{z}_{i2}, ..., \mathbf{z}_{iT}) = 0, \ t = 1, ..., T. \tag{1.8}$$

Assumption 1 is stronger than we need - as will be clear, $\mathrm{E}(\ddot{\mathbf{z}}_{it}'\ddot{u}_{it}) = \mathbf{0}, t = 1, ..., T$ would suffice – but (1.8) is a natural strict exogeneity assumption on the instruments. Assumption 1 is common in simultaneous equations models with panel data, as well as models with other kinds of endogeneity that induces correlation between $\mathbf{x}_{it}$ and u_{it}, such as omitted variables and measurement error. Assumption 1 rules out lagged dependent variables among the instruments – as well as other non-strictly exogenous instruments – and so its application to dynamic models is limited unless sufficient strictly exogenous instruments are available. When $\mathbf{z}_{it} = \mathbf{x}_{it}$, so that the covariates are strictly exogenous, Wooldridge (2005a) included $\mathbf{a}_i$ and $\mathbf{b}_i$ in the conditioning set, as in (1.5). When the unit-specific trend function is correctly specified, this stronger form of the assumption is essentially harmless.

The second component of the error term in (1.4) is $\ddot{\mathbf{x}}_{it}\mathbf{d}_i$, and we need assumptions such that $\ddot{\mathbf{z}}_{it}$ is uncorrelated with $\ddot{\mathbf{x}}_{it}\mathbf{d}_i$. This requires some care because $\ddot{\mathbf{x}}_{it}$ contains endogenous elements. (That is, we allow components of $\mathbf{x}_{it}$ to be endogenous even after removing unit-specific intercepts and trends.) The first assumption mimics the key assumption from Wooldridge (2005a), except that we replace the covariates with the instruments:

ASSUMPTION 2: $\mathbf{b}_i$ is mean independent of all the unit-specific "detrended" $\mathbf{z}_{it}$,

that is,

$$\mathrm{E}(\mathbf{b}_i|\ddot{\mathbf{z}}_{it}) = \mathrm{E}(\mathbf{b}_i) = \beta, t = 1, ..., T. \tag{1.9}$$

Because the $\ddot{\mathbf{z}}_{it}$ are net either of a time average or, more generally, level and trend effects, Assumption 2 maintains mean independence of the heterogeneous slopes and deviations of the instruments from long-run levels or trends. Of course, in the case where the instruments are assumed, in each time period, to be independent of all heterogeneity, Assumption 2 automatically holds. Assumption 2 is practically much weaker than full independence because it allows $\mathbf{b}_i$ to be arbitrarily correlated with systematic components of $\mathbf{z}_{it}$; we cover some examples in Section 1.4. [Wooldridge (2005a) contains a discussion for the case of strictly exogenous $\mathbf{x}_{it}$.]

Generally, the richer is $\mathbf{w}_t$, the more likely (1.9) is to hold. For example, the usual FE-IV estimator takes out time averages from the instruments, and this might not be enough to ensure (1.9) if the instruments are trending differently across units i. On the other hand, adding more aggregate factors to $\mathbf{w}_t$ reduces the variation in $\ddot{\mathbf{z}}_{it}$, generally leading to less efficient IV estimators. Not surprisingly, in deciding what to include in $\mathbf{w}_t$ we confront the usual tradeoff between efficiency and consistency.

Unfortunately, Assumptions 1 and 2 are not enough to conclude that the IV estimator is consistent. Instead, we employ a constant conditional covariance assumption.

ASSUMPTION 3: For $j = 1, ..., K$,

$$\mathrm{Cov}(\ddot{x}_{itj}, b_{ij}|\ddot{\mathbf{z}}_{it}) = \mathrm{Cov}(\ddot{x}_{itj}, b_{ij}), t = 1, ..., T. \tag{1.10}$$

Importantly, (1.10) allows the detrended covariates and the random coefficient to be correlated, and the covariance may change over time; in fact, there is no restriction on the temporal pattern of $\mathrm{Cov}(\ddot{x}_{itj}, b_{ij})$. But the covariance conditional on

the detrended IVs is assumed not to depend on $\ddot{\mathbf{z}}_{it}$. [In any case, the covariances $\mathrm{Cov}(\ddot{x}_{itj}, b_{ij})$ do not depend on i because of random sampling in the cross-sectional dimension. As we are conditioning only on $\ddot{\mathbf{z}}_{it}$, the restriction is that the covariance condition on $\ddot{\mathbf{z}}_{it}$ does not depend on $\ddot{\mathbf{z}}_{it}$; we have no need to place restrictions on other conditional covariances.]

Assumption 3 extends to the panel data case a condition used by Wooldridge (2003) for the pure cross-sectional case. An important difference is that Assumption 3 applies to the detrended covariates and instruments. Importantly, we allow the unconditional covariances to change arbitrarily over time. Of course, if $b_{ij} = \beta_j$ for all i, then (refeq:eq20) is trivially true because both sides are zero.

Assumptions 1 through 3 imply that the key orthogonality conditions (1.7) hold, and these conditions can be used in a generalized method of moments framework. For simplicity, we focus here on the fixed effects two stage least squares estimator, FE-2SLS [interpreted in the general sense of eliminating $\mathbf{a}_i$ from (1.1)]. To ensure consistency of FE-2SLS estimator we add a standard rank condition.

ASSUMPTION 4 (i) $\mathrm{rank}\left(\sum_{t=1}^{T} \mathrm{E}(\ddot{\mathbf{z}}_{it}'\ddot{\mathbf{x}}_{it})\right) = K$;

(ii) $\mathrm{rank}\left(\sum_{t=1}^{T} \mathrm{E}(\ddot{\mathbf{z}}_{it}'\ddot{\mathbf{z}}_{it})\right) = L$.

Practically speaking, the first part of Assumption 4 is most important; it means that, after netting out individual-specific trends, there is still sufficient correlation between the instruments and regressors. Part (ii) requires sufficient variation in the "detrended" instruments. It would be violated if, say, we specify $\mathbf{w}_t = (1, t)$ and $\mathbf{z}_{it}$ contains an element that is constant across t for all i (such as gender) or changes by the same value in each time period (such as a person's age when the length of the sampling period is constant).

PROPOSITION 1: Under Assumptions 1 to 4, the FE-IV estimator is consistent for β, provided a full set of time period dummies is included in (1.4).

PROOF: Under Assumption (refeq:eq19), $\mathrm{E}(d_{ij}|\ddot{\mathbf{z}}_{it}) = 0$, $j = 1, ..., K$ for all t, and so

$$\mathrm{E}(\ddot{x}_{itj}d_{ij}|\ddot{\mathbf{z}}_{it}) = \mathrm{Cov}(\ddot{x}_{itj}, d_{ij}|\ddot{\mathbf{z}}_{it}) = \mathrm{Cov}(\ddot{x}_{itj}, b_{ij}|\ddot{\mathbf{z}}_{it}).$$

But by Assumption 3, the conditional covariances equal the corresponding unconditional covariances, say γ_{tj}, and so $\mathrm{E}(\ddot{x}_{itj}d_{ij}|\ddot{\mathbf{z}}_{it}) = \gamma_{tj}$, $j = 1, ..., J$, $t = 1, ..., T$. Since $\ddot{\mathbf{x}}_{it}\mathbf{d}_i = \ddot{x}_{it1}d_{i1} + \ddot{x}_{it2}d_{i2} + ... + \ddot{x}_{itK}d_{iK}$, we have shown that $\mathrm{E}(\ddot{\mathbf{x}}_{it}\mathbf{d}_i|\ddot{\mathbf{z}}_{it}) = \gamma_{t1} + ... + \gamma_{tK} \equiv \theta_t$. Therefore, we can write $\ddot{\mathbf{x}}_{it}\mathbf{d}_i = \theta_t + r_{it}$ where $\mathrm{E}(r_{it}|\ddot{\mathbf{z}}_{it}) = 0$, $t = 1, ..., T$. Now we plug this expression for $\ddot{\mathbf{x}}_{it}\mathbf{d}_i$ into equation (1.4):

$$\ddot{y}_{it} = \theta_t + \ddot{\mathbf{x}}_{it}\beta + (r_{it} + \ddot{u}_{it}),\ t = 1, ..., T. \tag{1.11}$$

As we have just shown, Assumptions 2 and 3 imply that $\mathrm{E}(r_{it}|\ddot{\mathbf{z}}_{it}) = 0$. Assumption 1 implies that $\mathrm{E}(\ddot{u}_{it}|\ddot{\mathbf{z}}_{it}) = 0$. Thus, the composite error in (1.11) satisfies $\mathrm{E}(r_{it} + \ddot{u}_{it}|\ddot{\mathbf{z}}_{it}) = 0$, $t = 1, ..., T$, and so any IV method that uses instruments $\ddot{\mathbf{z}}_{it}$ at time t consistently estimates β. In particular, under the rank condition in Assumption 4, and standard finite moment conditions, the FE-2SLS estimator is consistent and $\sqrt{N}$-asymptotically normal. This completes the proof.

Proposition 1 contains an important empirical lesson: unless there are very good reasons to the contrary, one should include a full set of time effects in a fixed effects IV analysis. Even if the model does not originally contain separate time period intercepts itself a questionable premise – the estimating equation generally should if one wants to allow correlated random slope coefficients.

Because the error term in (1.11), $r_{it} + \ddot{u}_{it}$, is generally heteroskedastic and serially correlated – at a minimum due to the presence of $\ddot{\mathbf{x}}_{it}\mathbf{d}_i$ – inference should be carried out using a fully robust variance matrix for $\hat{\beta}$. Typically this is straightforward for pooled 2SLS where all instruments have been detrended prior to estimation.

1.4 Examples

To see how Proposition 1 applies, suppose $\mathbf{x}_{it}$ is linearly related to $\mathbf{z}_{it}$ with heterogeneous linear trends for each element of $\mathbf{x}_{it}$:

$$\mathbf{x}_{it} = \mathbf{g}_i\mathbf{\Gamma} + t \cdot \mathbf{h}_i\mathbf{\Psi} + \mathbf{z}_{it}\mathbf{\Pi} + \mathbf{q}_{it}, \ t = 1, ..., T. \tag{1.12}$$

Initially, take $\mathbf{w}_t = (1, t)$, so the regressors and instruments are linearly detrended before applying pooled 2SLS. Assume the instruments also have heterogeneous linear trends, which are removed by individual-specific detrending. Then Assumption 2 simply requires that the idiosyncratic movements in $\mathbf{z}_{it}$ are uncorrelated with $\mathbf{b}_i$, a weak requirement on instrumental variables. For Assumption 3, write $\ddot{\mathbf{x}}_{it} = \ddot{\mathbf{z}}_{it}\mathbf{\Pi} + \ddot{\mathbf{q}}_{it}$, $t = 1, ..., T$, so that $\text{Cov}(\ddot{\mathbf{x}}_{it}, \mathbf{b}_i|\ddot{\mathbf{z}}_{it}) = \text{Cov}[(\ddot{\mathbf{z}}_{it}\mathbf{\Pi} + \ddot{\mathbf{q}}_{it}), \mathbf{b}_i|\ddot{\mathbf{z}}_{it}] = \text{Cov}(\ddot{\mathbf{q}}_{it}, \mathbf{b}_i|\ddot{\mathbf{z}}_{it})$, $t = 1, ..., T$ under Assumption 2. Thus, provided

$$\text{Cov}(\ddot{\mathbf{q}}_{it}, \mathbf{b}_i|\ddot{\mathbf{z}}_{it}) = \text{Cov}(\ddot{\mathbf{q}}_{it}, \mathbf{b}_i), \ t = 1, ..., T, \tag{1.13}$$

we can use $\ddot{\mathbf{z}}_{it}$ as IVs for $\ddot{\mathbf{x}}_{it}$ to obtain a consistent estimate of the PAE, β, in equation (3.4). One might even assume that $(\mathbf{q}_{i1}, ..., \mathbf{q}_{iT}, \mathbf{b}_i)$ is independent of $(\mathbf{z}_{i1}, ..., \mathbf{z}_{iT})$, which is sufficient for (1.13) [as well as for Assumption 2].

It is possible that the FE-IV estimator is consistent even if we only demean the regressors and instruments, provided the instruments satisfy a stronger exogeneity assumption. In other words, even though $\mathbf{x}_{it}$ contains individual-specific linear trends, we ignore that in our estimation procedure. To see why we can still get consistency, demean $\mathbf{x}_{it}$ to get

$$\mathbf{x}_{it} - \bar{\mathbf{x}}_i = [t - (T+1)/2] \cdot \mathbf{h}_i\mathbf{\Psi} + (\mathbf{z}_{it} - \bar{\mathbf{z}}_i)\mathbf{\Pi} + (\mathbf{q}_{it} - \bar{\mathbf{q}}_i), \ t = 1, ..., T. \tag{1.14}$$

Now, if $[(\mathbf{q}_{it} - \bar{\mathbf{q}}_i), \mathbf{b}_i]$ is independent of $(\mathbf{z}_{it} - \bar{\mathbf{z}}_i)$ for each t, and (1.9) holds for $\ddot{\mathbf{z}}_{it} = (\mathbf{z}_{it} - \bar{\mathbf{z}}_i)$ and (1.13) also holds. Therefore,

$$\mathrm{Cov}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i, \mathbf{b}_i | \mathbf{z}_{it} - \bar{\mathbf{z}}_i) = [t - (T+1)/2]\mathbf{\Psi}'\mathrm{Cov}(\mathbf{h}_i, \mathbf{b}_i) = \mathrm{Cov}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i, \mathbf{b}_i)$$

for each t, which means that Assumption 3 holds: while the conditional covariances are not generally zero, or even constant over time, they do not depend on $\mathbf{z}_{it} - \bar{\mathbf{z}}_i$. So, the FE-IV estimator will be consistent provided we include a full set of year dummies in estimation.

What happens if we have a binary endogenous variable, x_{it}? Assumption 3 is unlikely to hold. To see why, take the case $w_t \equiv 1$, $t = 1, ..., T$, which corresponds to the usual unobserved effects model with correlated random coefficients. Then, $\ddot{x}_{it} = x_{it} - \bar{x}_i, t = 1, ..., T$, and we need $\mathrm{E}(\ddot{x}_{it} d_i | \ddot{\mathbf{z}}_{it})$ not to depend on $\ddot{\mathbf{z}}_{it}$. Now, by iterated expectations,

$$\mathrm{E}(\ddot{x}_{it} d_i | \ddot{\mathbf{z}}_{it}) = \mathrm{E}[\mathrm{E}(\ddot{x}_{it} d_i | d_i, \mathbf{z}_i) | \ddot{\mathbf{z}}_{it}] = \mathrm{E}[d_i \mathrm{E}(\ddot{x}_{it} | d_i, \mathbf{z}_i) | \ddot{\mathbf{z}}_{it}]. \tag{1.15}$$

Standard models for binary responses, with $\mathbf{z}_{it}$ strictly exogenous conditional on d_i, would have $\mathrm{P}(x_{it} = 1 | d_i, \mathbf{z}_i)$ depending on d_i and $\mathbf{z}_{it}$, in a nonlinear way. For concreteness, suppose $\mathrm{P}(x_{it} = 1 | d_i, \mathbf{z}_i)$ follows a probit model,

$$\mathrm{P}(x_{it} = 1 | d_i, \mathbf{z}_i) = \mathrm{P}(x_{it} = 1 | d_i, \mathbf{z}_{it}) = \Phi(\alpha_0 + \alpha_1 d_i + \mathbf{z}_{it}\alpha_2). \tag{1.16}$$

Then

$$\mathrm{E}(\ddot{x}_{it} | d_i, \mathbf{z}_i) = \Phi(\alpha_0 + \alpha_1 d_i + \mathbf{z}_{it}\alpha_2) - T^{-1} \sum_{r=1}^{T} \Phi(\alpha_0 + \alpha_1 d_i + \mathbf{z}_{ir}\alpha_2) \equiv g_t(d_i, \mathbf{z}_i) \tag{1.17}$$

and so, by (1.15),

$$\mathrm{E}(\ddot{x}_{it} d_i | \ddot{\mathbf{z}}_{it}) = \mathrm{E}[d_i g_t(d_i, \mathbf{z}_i) | \ddot{\mathbf{z}}_{it}] \tag{1.18}$$

Even if d_i is independent of $\ddot{\mathbf{z}}_{it}$ – a sensible strengthening of Assumption 2 – (1.18) generally depends on $\ddot{\mathbf{z}}_{it}$. Thus, assuming $\mathrm{E}(\ddot{x}_{it}d_i|\ddot{\mathbf{z}}_{it})$ does not depend on $\ddot{\mathbf{z}}_{it}$ is rather strong for a binary endogenous explanatory variable x_{it}. [Heckman (1997) contains a detailed discussion of the behavioral implications of this assumption in different empirical studies.] In a cross-sectional context, Wooldridge (1997) proposes a modified set of assumptions that are sufficient for consistent estimation of the ATE, β, with a binary endogenous variable, but, applied to the current setup, $\mathrm{P}(x_{it} = 1|d_i, \ddot{\mathbf{z}}_{it})$ would have to follow a linear probability model.

In a cross-sectional setting, Card (2001) shows that the analogue of Assumption 3 can also be violated in the case of roughly continuous explanatory variables due to heteroskedasticity in the variance matrix of $(\mathbf{x}_i, \mathbf{b}_i)$ given $\mathbf{z}_i$. (With a pure cross section, there are no time subscripts and, of course, no unit-specific demeaning or detrending.) In an earnings equation where $\mathbf{x}_i$ includes schooling, Card rejects $\mathrm{Cov}(\mathbf{x}_i, \mathbf{b}_i|\mathbf{z}_i) = \mathrm{Cov}(\mathbf{x}_i, \mathbf{b}_i)$ using IQ score as a proxy for unobserved ability (an element of $\mathbf{b}_i$) and a binary indicator for college proximity as an instrument for education. In our panel data setup, Assumption 1 allows $\mathrm{Cov}(\mathbf{x}_{it}, \mathbf{b}_i|\mathbf{z}_{it})$ to depend on $\mathbf{z}_{it}$, as it generally would if $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$ contain persistent heterogeneity correlated with $\mathbf{b}_i$. Using a generalized fixed effects approach, we need only assume $\mathrm{Cov}(\ddot{\mathbf{x}}_{it}, \mathbf{b}_i|\ddot{\mathbf{z}}_{it})$ does not depend on $\ddot{\mathbf{z}}_{it}$, and this is much more plausible when we think the unit-specific detrending successfully eliminates the time-constant heterogeneity in $\ddot{\mathbf{x}}_{it}$ and $\ddot{\mathbf{z}}_{it}$.

More recently, in a cross-sectional setting, Wooldridge (2005b) proposes conditions that allow $\mathrm{Cov}(\mathbf{x}_i, \mathbf{b}_i|\mathbf{z}_i)$ to depend on $\mathbf{z}_i$, but these do not apply directly to the panel data case with time-constant heterogeneity that can be correlated with the covariates and instruments.

1.5 Finite Sample Behavior of the FE-IV Estimator

In this section we provide evidence on the finite sample properties of FE-IV estimator of the population averaged effect in a CRC panel data model. Because one of the most commonly used applications of CRC panel data models is the usual unobserved effects model with a random coefficient, we first assume $w_t \equiv 1$, $t = 1, ...T$ in (1.1), as in the second part of the first example from Section 1.4. Also, for scalar processes x_{it} and z_{it}, we assume a linear relationship between x_{it} and z_{it}, with a linear trend for x_{it}. We use Monte Carlo simulations to draw the data and check the properties of the estimator. The number of replications is 500, and the results of the experiment are presented for cross-sectional sample sizes of 100, 400, and 800 for two time horizons, $T = 5$ and $T = 10$. The population average values are $\beta = 2$ and $\alpha = 3$.

For $t = 1, ..., T$, the endogenous explanatory variable is generated as

$$x_{it} \equiv \lambda_{xz} z_{it} + \lambda_{xu} u_{it} + \lambda_{xa} a_i + \xi b_i + \xi t d_i + \sqrt{1 - \lambda_{xz}^2 - \lambda_{xu}^2 - \lambda_{xa}^2 - \xi^2 (1+t)^2} e_{it}, \tag{1.19}$$

where u_{it}, $e_{it} \sim$ Normal $(0, 1)$, $a_i \sim$ Normal $(\alpha, 1)$, $b_i = \beta + d_i$, $d_i \sim$ Normal $(0, \sigma_b^2)$, and $\lambda_{xz}, \lambda_{xu}, \lambda_{xa}$, and ξ are constants. Further, the instrument is generated as $z_{it} = \lambda_{za} a_i + \sqrt{1 - \lambda_{za}^2} m_{it}$ – where a_i is defined above – $m_{it} \sim$ Normal $(t, 1)$, and λ_{za} is the population correlation coefficient between z_{it} and a_i, $t = 1, ...T$.

In our reported simulations we use $\sigma_b^2 = 1$. When $\lambda_{za} = 0$, the coefficients λ_{xz}, λ_{xu}, and λ_{xa} from (1.19) are the population correlation coefficients between x_{it} and z_{it}, x_{it} and u_{it}, and x_{it} and a_i, $t = 1, ..., T$, respectively. The population correlation between x_{it} and b_i when $\lambda_{za} = 0$ is $\xi(1+t)$, $t = 1, ..., T$. We use the coefficient on the error term in (1.19) to ensure that x_{it} has unit variance when $\lambda_{za} = 0$. When $\lambda_{za} \neq 0$,

$\text{Var}(x_{it}) = 1 + 2\lambda_{xz}\lambda_{za}\lambda_{xa}$, which is only slightly greater than one for our choices of the λ parameters. The relevant covariances are $\text{Cov}(x_{it}, u_{it}) = \lambda_{xu}$, $\text{Cov}(x_{it}, a_t) = \lambda_{xz}\lambda_{za} + \lambda_{xa}$, and $\text{Cov}(x_{it}, z_{it}) = \lambda_{xz} + \lambda_{xa}\lambda_{za}$. For the endogenous explanatory variable defined in (1.19), Assumption 3 is met: $\text{Cov}(\ddot{x}_{it}, b_i|\ddot{z}_{it}) = \text{Cov}(\ddot{x}_{it}, b_i) = \xi(1+t)$, $t = 1, ..., T$.

The dependent variable y_{it} is generated as:

$$y_{it} = a_i + x_{it}b_i + u_{it}, \ t = 1, ..., T, \tag{1.20}$$

where a_i, b_i, u_{it}, and x_{it} are defined above. Among other estimators, we obtain the FE-IV estimator in (1.20) acting as if $b_i = \beta$. Based on the first example from Section 4, we know this FE-IV estimator is consistent for x_{it} generated as in (1.19) provided we include a full set of time dummies, even though we only demean the regressor and the instrument while ignoring the individual-specific linear trend in the regressor.

Tables A.1 and A.2 present simulation results for the correlated random coefficient model for $\lambda_{xu} = .40$, $\lambda_{xa} = .20$, $\lambda_{xz} = .20$, and $\lambda_{za} = .25$. The implied correlation between x_{it} and z_{it} is about .245, which seems to be a reasonable value for panel data. For comparison, we used a data set provided with Wooldridge (2002) on domestic route air fares for 1,149 routes in the United States for 1997 through 2000. (The data set is called AIRFARE.) The correlation between the log of air fare (an endogenous explanatory variable in a passenger demand equation) and the instrumental variable candidate, the concentration ratio on the route, is about $-.22$, which has a magnitude in the range of .245.

Table A.1 reports the simulation outcomes for $T = 5$, where $\xi = .12$, while Table A.2 covers the case $T = 10$, where $\xi = .06$. When $\xi = .12$, the correlation between x_{i1} and b_i is slightly less than .24; when $\xi = .06$, the correlation is just below .12. Columns 1 through 6 contain the mean, standard deviation (SD), root mean squared error (RMSE), lower quartile (LQ), median, and upper quartile (UQ) of the

PAE estimates from 500 replications. Rows of the table report statistics for usual pooled ordinary least squares (POLS) estimates on the original data, the usual fixed effects estimates (FE-OLS), which is just pooled OLS on the time-demeaned data, pooled instrumental variables (IV) estimates using the original data, the fixed effects-instrumental variables estimates without period dummy variables (FE-IV without dummies), and fixed effects instrumental variables estimates when a full set of period dummy variables is included (FE-IV with dummies).

From the table we see that the POLS estimates are roughly 1.5 times larger than the true value of β in the 100, 400 and 800 observation samples. One source of bias of the POLS estimates is the correlation between the unobserved heterogeneity a_i and the regressor x_{it}. A second source of bias in the POLS estimates is the endogeneity of the regressor x_{it}, with correlation coefficient ρ_{xu} very close to .4. A third source of bias (and inconsistency) is the correlation between x_{it} and b_i.

The within transformation eliminates a_i, and so the correlation between x_{it} and a_i is not a source of bias for the usual FE-OLS estimator. But FE-OLS still produces a biased estimator of β for the last two reasons mentioned above. The bias in the FE-OLS estimator is much lower than for POLS, but the bias is still on the order of 30 percent.

The pooled IV estimator – that is, without removing time averages and without time period dummies – actually has a larger bias than the FE-OLS estimator, a finding that is not too surprising because the instruments are correlated with a_i. Using the FE transformation combined with IV eliminates the dependence between z_{it} and a_i because $z_{it} = \lambda_{za} a_i + \sqrt{1 - \lambda_{za}^2} m_{it}$. Therefore, the FE-IV estimator (without time dummies) has a smaller bias and considerably smaller RMSE than the pooled IV estimator. More importantly, the FE-IV estimator with period dummies has the lowest RMSE among all estimators for all the sample sizes and both time horizons. Plus, the RMSE of the FE-IV estimator with time dummies falls quickly as the sample

size, N, grows. Without period dummies, the FE-IV estimates of β are biased by at least 20 percent, and the bias does not disappear as $N \longrightarrow \infty$. As T increases, the RMSE of the FE-IV estimator without dummies estimates decreases but it is still higher than the one for the FE-IV estimates when the period dummy variables are included. Thus, even though the structural model (1.20) does not contain a time trend, inclusion of a full set of period dummies ensures the consistency of the FE-IV estimation.

Not surprisingly, the FE-OLS estimator has a smaller standard deviation than the FE-IV estimator (both without time dummies). Typically, methods that treat regressors as exogenous have substantially less sampling variation than their IV counterparts because the correlation between the instrument and regressor is typically well below one, as in the current simulation.

The difference between the FE-IV estimates with and without time dummies illustrates the trade-off between bias and variance. The FE-IV estimates without time period dummy variables are always less variable than the FE-IV with time dummies. This is hardly surprising, as including more explanatory variables – the time dummies in this case – that are correlated with the instrument induces multicollinearity into the IV estimates. The instrument, z_{it}, is constructed to be correlated with time dummies, and so the FE-IV estimator with time dummies is less precise than that without. But, of course, the estimator without time dummies suffers from substantial bias even though the structural model does not contain separate period intercepts. The RMSE for the FE-IV estimator that includes a full set of dummies is much lower than the estimator that does not.

We also conducted simulations with more variability in the random coefficient, namely, $\sigma_b^2 = 4$, so that the standard deviation of b_i is double that in Tables A.1 and A.2. The results of these simulations are not included here but are available on request. With more variability in b_i, the bias induced by failing to include time

dummies in the FE-IV estimation is more pronounced (even though, remember, the structural model does not include time effects). For example, with $T = 5$, and $N = 800$, the RMSE of the FE-IV estimator without dummies is about 1.36, compared with about .22 for the estimator that does include the dummies.

For the next set of simulations, we take $\mathbf{w}_t \equiv (1, t)$, $t = 1, ..., T$, in (1.1), so that each cross-sectional unit has its own linear trend. In particular, we generate y_{it} as

$$y_{it} = a_{i0} + a_{i1}t + x_{it}b_i + u_{it}, \ t = 1, ..., T, \tag{1.21}$$

where a_{i0} and a_{i1} are independent Normal$(\alpha, 1)$ random variables and b_i, and u_{it} are defined above. The endogenous explanatory variable x_{it} is generated as

$$x_{it} \equiv \lambda_{xz}z_{it}+\lambda_{xu}u_{it}+\lambda_{xa}(a_{i0}+a_{i1})+\xi b_i+\xi t d_i+\sqrt{1-\lambda_{xz}^2-\lambda_{xu}^2-2\lambda_{xa}^2-\xi^2(1+t)^2}e_{it}, \tag{1.22}$$

and the instrument is generated as $z_{it} = \lambda_{za}a_{i0}+\sqrt{1-\lambda_{za}^2}m_{it}$. Again, the coefficient on e_{it} is chosen so that $\text{Var}(x_{it}) = 1$ if $\lambda_{za} = 0$. We use the same values for the λ parameters as in Tables A.1 and A.2, and we take $\sigma_b = 1$. (Simulation findings for the case $\sigma_b = 2$ are available on request.) Because the structural model (1.21) contains a time trend, the default is to include a full set of time period dummies in the various estimation methods. For comparison, we include the FE-IV estimator without time period dummies.

The rows of Tables A.3 and A.4 report statistics for POLS with time dummies, fixed effects with time dummies, pooled instrumental variables with time dummies, fixed effects instrumental variables estimates with time dummies, and fixed effects instrumental variables estimates without time dummies. As in Tables A.1 and A.2, the simulation findings are unambiguous: fixed effects IV with a full set of time dummies is superior, by far, to the other estimation methods, for all combinations of N and T. Perhaps not surprisingly, when y_{it} is itself trending, the consequences of

omitting aggregate time effects is much more detrimental than in the previous case.

The simulation findings are perhaps not too surprising: the only estimator that is essentially unbiased for the PAE removes the unobserved effect (or, more generally, the individual-specific trends), includes a full set of aggregate time effects, and instruments for the endogenous explanatory variable. Nevertheless, it is useful to see that the theoretical findings in Section 1.3 have practically important implications: the FE-IV estimator with time dummies is robust to correlation between the random coefficients and the explanatory variable, at least for assumptions that can be met by continuous endogenous explanatory variables.

1.6 Conclusion

This paper suggests a set of conditions sufficient for applying the standard IV approach to the estimation of population averaged effects in a correlated random coefficient panel data model with (roughly) continuous endogenous explanatory variables. Assumptions 1 through 4 ensure consistent FE-IV estimation of the population averaged slopes, β, even ignoring individual-specific slopes. Monte Carlo simulations suggest the proposed FE-IV estimator of PAE provided a full set of period dummy variables is included performs better than other estimators in finite samples for the case of (roughly) continuous endogenous explanatory variables.

A natural direction for future work is to relax homoskedasticity of $\mathrm{E}(\ddot{\mathbf{x}}_{it}\mathbf{d}_i|\ddot{\mathbf{z}}_{it})$; Card (2001) showed how the analogous assumption can fail in a cross-sectional environment. Recently, Murtazashvili (2006) shows how this assumption can be relaxed using a control function approach by putting restrictions on the reduced forms of the endogeneous elements of $\mathbf{x}_{it}$ – restrictions that can be met for roughly continuous variables – and by modeling the conditional covariances.

CHAPTER 2

A CONTROL FUNCTION APPROACH TO ESTIMATION OF CORRELATED RANDOM COEFFICIENT PANEL DATA MODELS

2.1 Introduction

Recently, a lot of attention has been devoted to estimation of average partial effects (APEs) in correlated random coefficient (CRC) models, in both cross section and panel data settings. Studies are primarily conducted in a cross-sectional setup with few exceptions for panel data. CRC panel data models are investigated for both exogenous and endogenous explanatory variables. Wooldridge (2005a) discusses fixed effects estimation of a CRC model for the case of exogenous independent variables in a panel data setting. Murtazashvili and Wooldridge (2005) address fixed effects instrumental variables (FE-IV) estimation of APEs with (at least roughly) continuous endogenous regressors in CRC panel data models.[1] One of the main conditions for consistent estimates of APEs in their study is an assumption of independence of co-

[1]We refer to the continuous variables with some discrete characteristics as roughly continuous, and provide a discussion about this kind of variables in the next section.

variance between detrended endogenous regressors and individual heterogeneity, conditional on the transformed IVs, from the detrended instruments. Card (2001) shows for cross-sectional data that this assumption can be violated in the case of roughly continuous endogenous explanatory variables due to heteroskedasticity in variance-covariance matrix of explanatory variables and individual heterogeneity conditional on the instruments. He rejects this assumption using IQ as a proxy for unobserved ability and a binary indicator for college proximity as an instrument for education in the human capital earnings model. Wooldridge (2005b) proposes conditions weaker than those in Murtazashvili and Wooldridge (2005) for obtaining consistent APEs estimates for (roughly) continuous regressors with the Card's problem in a cross-sectional setup.

In this paper, we study the model in Murtazashvili and Wooldridge (2005) but, in addition to allowing some explanatory variables to be correlated with the idiosyncratic error, we correct for the drawback described in Card (2001) while still allowing the endogenous regressors to be (roughly) continuous. We use a control function approach, which introduces residuals from the reduced form for the endogenous regressors as covariates in the structural model. We propose a two-step method to account for endogeneity and to consistently estimate APEs in CRC panel data models with endogenous (roughly) continuous regressors. The motivation for our two-step panel data procedure comes from a cross section study by Wooldridge (2005b). Further, we relax the assumptions in Wooldridge (2005a) and Murtazashvili and Wooldridge (2005) by allowing the individual slopes in a CRC model to vary over time. Both cases of time-constant and time-varying individual slopes are covered in this paper.

Monte Carlo simulations indicate that in the finite samples the control function (CF) approach we propose for estimating the CRC balanced panel data model with time-invariant individual heterogeneity performs better than other estimators when the joint distribution of the individual heterogeneity and the endogenous regressors

conditional on the detrended instruments depends on the instrumental variables.

We apply the proposed method to the problem of estimating the average partial effects of annual hours of on-job-training on output scrap rates for manufacturing firms in Michigan using the firm level data for 1987 through 1989. The control function approach we propose delivers the APEs of the annual hours of job training on the output scrap rates that are larger in magnitudes and statistically more significant than the APEs' estimates from the FE-IV approach.

2.2 Model of Interest for Balanced Panels

For a random draw i from the population, the structural model is

$$y_{1it} = \mathbf{w}_t \mathbf{a}_i + \mathbf{x}_{it} \mathbf{b}_i + u_{it}, \ t = 1, ..., T, \tag{2.1}$$

where $\mathbf{w}_t$ is a $1 \times J$ vector of aggregate time variables which we treat as nonrandom – $\mathbf{a}_i$ is a $J \times 1$ vector of individual-specific slopes on the aggregate variables, $\mathbf{x}_{it}$ is a $1 \times K$ vector of exogenous covariates, $\mathbf{z}_{1it}$, and an endogenous covariate, y_{2it}, that change across time, in general, $\mathbf{x}_{it} = \mathbf{f}(\mathbf{z}_{1it}, y_{2it})$, $\mathbf{b}_i$ is a $K \times 1$ vector of individual-specific slopes, and u_{it} is an idiosyncratic error. For simplicity, assume $\mathbf{x}_{it} = (\mathbf{z}_{1it}, y_{2it})$. Let $\mathbf{z}_{it} = (\mathbf{z}_{1it}, \mathbf{z}_{2it})$ be a $1 \times L$ vector of instrumental variables, with $L \geq K$, i.e., we assume the vector $\mathbf{z}_{2it}$ contains at least one element. We assume a sample of size N randomly drawn from the population, and T being fixed in the asymptotic analysis. For the purpose of this paper, we assume a balanced panel.

Our object of interest is $\beta = \mathrm{E}(\mathbf{b}_i)$, the $K \times 1$ vector of average partial effects, i.e., vector of partial effects averaged over the population distribution of any unobserved heterogeneity. The APEs are usually of primary interest to empirical analysts. Another empirical question of possible interest is estimating $\mathbf{b}_i$s themselves. However, the estimation of $\mathbf{b}_i$s, when we treat them as parameters, is not precise unless T is large. As an alternative, we turn to estimation of average partial effects in our model.

Following Murtazashvili and Wooldridge (2005) we study estimators of β that are based on the assumption that the slopes $\mathbf{b}_i$ are constant, but we study the properties of these estimators in the context of model (2.1). We write $\mathbf{b}_i = \beta + \mathbf{d}_i$, and $\mathrm{E}(\mathbf{d}_i) = 0$, by definition. In other words, we assume that that individual heterogeneities have constant means, β, and random error terms, $\mathbf{d}_i$. Substitution into (2.1) gives

$$\begin{aligned} y_{1it} &= \mathbf{w}_t\mathbf{a}_i + \mathbf{x}_{it}\beta + (\mathbf{x}_{it}\mathbf{d}_i + u_{it}) \\ &\equiv \mathbf{w}_t\mathbf{a}_i + \mathbf{x}_{it}\beta + v_{1it}, \end{aligned} \tag{2.2}$$

where $v_{1it} \equiv \mathbf{x}_{it}\mathbf{d}_i + u_{it}$. We estimate β in (2.1) allowing the entire vector $\mathbf{a}_i$ to vary by i, and to be arbitrarily correlated with $\mathbf{x}_{it}$. Following a cross-sectional definition from Heckman and Vytlacil (1998), we call (2.1) a *correlated random coefficient* model because of the possible correlation between $\mathbf{b}_i$ and $\mathbf{x}_{it}$.

In this paper we develop a two-step estimation method motivated by Wooldridge (2005b) for obtaining consistent estimates of the average partial effects. The method we employ for obtaining consistent estimates of APEs is called a *control function* approach, which was pioneered by Smith and Blundell (1986) and Rivers and Vuong (1988). The main idea of the control function method is to add control variables into the structural model to control for the endogeneity problem (regardless of its exact nature). To use the control function approach in our case, we need to make assumptions about the nature of the endogeneity in the random coefficient model. Since we have two sources of endogeneity in our model - the correlation between the unobserved heterogeneities and the regressor y_{2it}, and the correlation between that regressor and the structural error, we are interested in modeling the relationships among the random coefficients, exogenous covariates, and the error from the reduced form equation for the endogenous explanatory variable.

First, we assume there is some strictly monotonic function $h(\cdot)$ defined on the

support set of y_{2it}, such that

$$h(y_{2it}) = \delta_{20} + \mathbf{z}_{it}\delta_{21} + \overline{\mathbf{z}}_i\delta_{22} + v_{2it}, \ t = 1, ..., T, \tag{2.3}$$

$$\mathrm{E}(v_{2it}|\mathbf{z}_{i1}, ..., \mathbf{z}_{iT}) = 0, \ t = 1, ..., T, \tag{2.4}$$

where $\overline{\mathbf{z}}_i = T^{-1}\sum_{t=1}^{T}\mathbf{z}_{it}$, v_{2it}'s are error terms, and

$$\begin{aligned} \mathrm{E}(u_{it}|\mathbf{z}_{i1}, ..., \mathbf{z}_{iT}, v_{2i1}, ..., v_{2iT}) &= \mathrm{E}(u_{it}|v_{2i1}, ..., v_{2iT}) = \\ &= \rho_1 v_{2it} + \rho_2 \overline{v}_{2i}, \ t = 1, ..., T, \end{aligned} \tag{2.5}$$

where ρ_1 and ρ_2 are scalars, and $\overline{v}_{2i} = T^{-1}\sum_{t=1}^{T} v_{2it}$. Assumption (2.5) is stronger than just assuming that u_{it} is uncorrelated with $\mathbf{z}_i$. There are two parts to this assumption. The first equality says that u_{it} is conditional mean independent of $\mathbf{z}_i$ given $v_{2i1}, ..., v_{2iT}$. This will always be true if $(u_{it}, v_{2i1}, ..., v_{2iT})$ and $\mathbf{z}_i$ are independent. The second equality states that $\mathrm{E}(u_{it}|v_{2i1}, ..., v_{2iT})$ is linear. Assumption (2.5) holds if $v_{2it} = a_{2i} + e_{2it}$, where $\{(u_{it}, e_{2it})\}$ is independently and identically distributed and all conditional expectations are linear. Thus, we maintain (2.5) is a valid extension to the CRC panel data models. We follow Rivers and Vuong (1988) and call equation (2.3) a reduced form equation.

Strict monotonicity of $h(\cdot)$ implies that y_{2it} is a well-defined function of $\{\mathbf{z}_{i1}, ..., \mathbf{z}_{iT}\}$ and v_{2it}. Further, assumptions (2.3) and (2.4) mean that when some function $h(\cdot)$ is applied to the endogenous explanatory variable, y_{2it}, the latter has a linear conditional mean given all the instruments. In other words, linearity of $\mathrm{E}(y_{2it}|\mathbf{z}_{i1}, ..., \mathbf{z}_{iT})$ might not be an appropriate assumption, while we want y_{2it} to be included linearly in the regression equation. Assumption (2.4) always holds if v_{2it} is independent of $\mathbf{z}_i$. In the standard case of continuous y_{2it} with a large support set assumptions (2.3) and (2.4) are very reasonable in many possible situations. But if the endogenous covariate has characteristics that are not quite suitable for a continuous variable these assumptions do not generally hold. For example, assume a

continuous variable y_{2it} with a large support set is defined according to (2.3) when $h(\cdot)$ is identity so that $v_{2it}|\mathbf{z}_i \tilde{}\text{Normal}(0, \sigma_{it}^2)$, where $\sigma_{it}^2 = \text{Var}(v_{2it}|\mathbf{z}_i)$ is a conditional variance that depends on $\mathbf{z}_i$. In this case we can standardize v_{2it} to be a variable $\frac{v_{2it}}{\sigma_{it}}$, which is independent of $\mathbf{z}_i$, guaranteeing that assumption (2.4) is satisfied. However, assumption (2.4) is unlikely to hold if y_{2it} has some "discrete"-type characteristics. For instance, let y_{2it} be a binary variable so that $y_{2it}|\mathbf{z}_i$ follows a probit model. Even having standardized the error term for this variable, v_{2it}, we cannot hope to obtain a new one, which is independent of $\mathbf{z}_i$.

For the purpose of our study, we will refer to the continuous variables with some discrete characteristics as *roughly continuous* to distinguish them from the traditional continuous variables and emphasize that these roughly continuous variables do not always have fine behaviors of continuous variables. Possible examples of these variables would be income, education, experience, etc. Garen (1984) discusses estimation of models in the presence of selection bias when the choice variable is continuous and the choice set is ordered. He suggests treating level of education in the human capital earnings model as such a continuous variable: on the one hand, schooling is traditionally thought of as a continuous variable, on the other hand, only integers of that variable are observed.

Which functions can we use as a strictly monotonic function $h(\cdot)$ in transformation (2.3)? For a trivial case of a continuous y_{2it} with a large support set, we can use $h(y_{2it}) = y_{2it}$. When the nature of y_{2it} is more "exotic," the choice of $h(\cdot)$ is not so straightforward. For instance, Wooldridge (2005b) suggests using $h(y_{2it}) = \ln(\frac{y_{2it}}{1-y_{2it}})$, when y_{2it} is a fraction in the open unit interval, and $h(y_{2it}) = \ln(y_{2it})$, when $y_{2it} > 0$. Assumptions (2.3) and (2.4) rule out probit, logit, and Tobit models because y_{2it} has discrete characteristics.

For example, if we are interested in estimating whether there is an effect of per-pupil spending on math test pass rates for fourth graders in Michigan, and the en-

dogenous variable of our interest is per-pupil spending, then per-pupil spending is a roughly continuous variable and choosing a log-transformation of per-pupil spending is appropriate. If we employ the logged per-pupil spending as the endogenous covariate in the model, then logged per-pupil spending can be thought of as a continuous variable and function $h(y_{2it}) = y_{2it}$ with $y_{2it} = \ln(\text{per-pupil spending})$ is clearly adequate.

Second, we need to make assumptions about the distribution of $(\mathbf{a}_i, v_{2it})$ conditional on the instruments. We assume

$$\mathrm{E}(\mathbf{a}_i|\mathbf{z}_{i1}, ..., \mathbf{z}_{iT}, v_{2i1}, ..., v_{2iT}) = \mathrm{E}(\mathbf{a}_i|\overline{\mathbf{z}}_i, \overline{v}_{2i}), \tag{2.6}$$

and

$$\mathrm{E}(\mathbf{a}_i|\overline{\mathbf{z}}_i, \overline{v}_{2i}) = \alpha + \mathbf{A}_1\overline{\mathbf{z}}_i\prime + (\mathbf{A}_2 + \mathbf{A}_3\overline{\mathbf{z}}_i\prime)\overline{v}_{2i}, \tag{2.7}$$

where α and $\mathbf{A}_2$ are $J \times 1$, $\mathbf{A}_1$ and $\mathbf{A}_3$ are $J \times L$ matrices of constants, respectively, $\overline{\mathbf{z}}_i$ and $\overline{v}_{2i}$ are defined above. Assumption (2.6) means that $\overline{\mathbf{z}}_i$ and $\overline{v}_{2i}$ can be thought of as sufficient statistics for describing the relationship between $\mathbf{a}_i$ and the history of $\{\mathbf{z}_{it}, v_{2it} : t = 1, ..., T\}$. Assumption (2.7) specifies a particular functional form for the relationship among $\mathbf{a}_i$, $\overline{\mathbf{z}}_i$, and $\overline{v}_{2i}$. Interactions among the exogenous variables $\mathbf{z}_{it}$ and v_{2it} might be important. In a cross-sectional context, Card (2001) shows that the joint distribution of $(\mathbf{a}_i, v_{2it})$ given $\mathbf{z}_{it}$ can depend on $\mathbf{z}_{it}$ due to heteroskedasticity in $\mathrm{Var}(\mathbf{a}_i, v_{2it}|\mathbf{z}_{it})$. He shows that using IQ as a proxy for unobserved ability and a binary indicator for college proximity as an instrument for education in the human capital earnings model. Assumptions (2.4) and (2.7) can still be true even when the conditional variance-covariance matrix, $\mathrm{Var}(\mathbf{a}_i, v_{2it}|\mathbf{z}_i)$, is heteroskedastic.

Third, we need to make assumptions about the expected value of $\mathbf{d}_i$, conditional on $\{\mathbf{z}_{it}\}$ and $\{v_{2it}\}$. We assume

$$\mathrm{E}(\mathbf{d}_i|\mathbf{z}_{i1}, ..., \mathbf{z}_{iT}, v_{2i1}, ..., v_{2iT}) = \mathrm{E}(\mathbf{d}_i|\overline{\mathbf{z}}_i, \overline{v}_{2i}), \tag{2.8}$$

and, in particular,

$$\mathrm{E}(\mathbf{d}_i|\mathbf{z}_{i1},...,\mathbf{z}_{iT},v_{2i1},...,v_{2iT}) = \mathbf{B}_1(\overline{\mathbf{z}}_i - \psi)\prime + (\mathbf{B}_2 + \mathbf{B}_3\overline{\mathbf{z}}_i\prime)\overline{v}_{2i}, \tag{2.9}$$

where $\psi \equiv \mathrm{E}(\overline{\mathbf{z}}_i)$, $\mathbf{B}_1$ and $\mathbf{B}_3$ are $K \times L$, and $\mathbf{B}_2$ is $K \times 1$ matrices of constants, $\overline{\mathbf{z}}_i$ and $\overline{v}_{2i}$ are defined above. In Murtazashvili and Wooldridge (2005), one of the conditions for consistency of β estimates states that the covariance between $\mathbf{d}_i$ and the detrended $\mathbf{x}_{it}$ conditional on the detrended $\mathbf{z}_{it}$ equals its unconditional version, that is, it does not depend on the detrended $\mathbf{z}_{it}$. For the reasons mentioned earlier, this assumption might be too restrictive for the case of roughly continuous endogenous explanatory variables. In this study, we relax this assumption not only by dealing with the original data, but also by allowing the covariance between $\mathbf{d}_i$ and $\mathbf{x}_{it}$ conditional on the instruments to be a function of $\mathbf{z}_{it}$. The conditions we employ in this paper assure the consistency of β estimates in the case of roughly continuous endogenous explanatory variables.

Then, we take the expectation of equation (2.2) with respect to $(\mathbf{z}_{i1},...,\mathbf{z}_{iT},v_{2i1},...,v_{2iT})$, employ that y_{2it} is a deterministic function of $(\mathbf{z}_{i1},...,\mathbf{z}_{iT},v_{2i1},...,v_{2iT})$, and use assumptions (2.3) through (2.9). The resulting estimating equation is:

$$\begin{aligned}\mathrm{E}(y_{1it}|\mathbf{z}_{i1},...,\mathbf{z}_{iT},v_{2i1},...,v_{2iT}) &= \mathbf{w}_t\alpha + (\overline{\mathbf{z}}_i \otimes \mathbf{w}_t)\alpha_1 + \overline{v}_{2i}\mathbf{w}_t\alpha_2 + \overline{v}_{2i}(\overline{\mathbf{z}}_i \otimes \mathbf{w}_t)\alpha_3 + \\ &+\mathbf{x}_{it}\beta + ((\overline{\mathbf{z}}_i - \psi) \otimes \mathbf{x}_{it})\beta_1 + \overline{v}_{2i}\mathbf{x}_{it}\beta_2 + \overline{v}_{2i}(\overline{\mathbf{z}}_i \otimes \mathbf{x}_{it})\beta_3 + \rho_1 v_{2it},\end{aligned} \tag{2.10}$$

where $t = 1,...,T$. Here, $H = 2(1+L)(J+K)+1$ is the total number of all the independent second-stage variables. Equation (2.10) is an estimating equation for obtaining consistent estimates of APE, β. Importantly, the components of $\mathbf{z}_{2it}$ – the instrumental variables excluded from the structural equation (2.2) – do not enter the estimating equation (2.10) in levels or interacted only with $\mathbf{z}_{1it}$. Generally, if we had any of these introduced in (2.10) we would lose identification. [See Wooldridge (2005b) for more details.]

In some cases we might think that assumption (2.8) is too restrictive. In some potential applications we might want to allow the random coefficient to vary not only across i but also across t. In other words, for a random draw i from the population, the structural model becomes

$$y_{1it} = \mathbf{w}_t\mathbf{a}_i + \mathbf{x}_{it}\mathbf{b}_{it} + u_{it}, \quad t = 1, ..., T, \tag{2.11}$$

where $\mathbf{b}_{it}$ is a $K \times 1$ vector of time varying individual-specific slopes. We write $\mathbf{b}_{it} = \beta + \mathbf{q}_{it}$, and $\mathrm{E}(\mathbf{q}_{it}) = 0$, by definition. In other words, we assume that that individual heterogeneities have constant means, β, and random error terms, $\mathbf{q}_{it}$. Further, we assume that $\mathbf{q}_{it}$s consist of both time-constant and time-varying zero mean components, i.e., $\mathbf{q}_{it} = \mathbf{d}_i + \mathbf{r}_{it}$. Substitution into (2.1) gives

$$\begin{aligned} y_{1it} &= \mathbf{w}_t\mathbf{a}_i + \mathbf{x}_{it}\beta + (\mathbf{x}_{it}\mathbf{q}_{it} + u_{it}) \\ &\equiv \mathbf{w}_t\mathbf{a}_i + \mathbf{x}_{it}\beta + v_{1it}, \end{aligned} \tag{2.12}$$

where $v_{1it} \equiv \mathbf{x}_{it}\mathbf{q}_{it} + u_{it}$. Then, the estimation equation for the model (2.12) will need to expand in comparison with the estimation equation (2.10) to reflect the time varying nature of the individual multiplicative heterogeneity. Assumptions (2.8) and (2.9) can be replaced with the following assumptions about the error term and the distribution of $(\mathbf{q}_{it}, v_{2it})$ conditional on the instruments:

$$\mathrm{E}(\mathbf{q}_{it}|\mathbf{z}_{i1}, ..., \mathbf{z}_{iT}, v_{2i1}, ..., v_{2iT}) = \mathrm{E}(\mathbf{q}_{it}|\overline{\mathbf{z}}_i, \overline{v}_{2i}, v_{2it}, \mathbf{z}_{it}), \tag{2.13}$$

which says that $\mathrm{E}(\mathbf{q}_{it}|\mathbf{z}_{i1}, ..., \mathbf{z}_{iT}, v_{2i1}, ..., v_{2iT})$ depends only on the time t values and time averages. Since we maintain $\mathbf{q}_{it}$s consist of time-constant and time-varying components $\mathbf{d}_i$ and $\mathbf{r}_{it}$, respectively, assumption (2.13) reflects the nature of $\mathbf{q}_{it}$. And, finally, we assume

$$\begin{aligned} &\mathrm{E}(\mathbf{q}_{it}|\mathbf{z}_{i1}, ..., \mathbf{z}_{iT}, v_{2i1}, ..., v_{2iT}) = \mathrm{E}(\mathbf{d}_i + \mathbf{r}_{it}|\overline{\mathbf{z}}_i, \overline{v}_{2i}, v_{2it}, \mathbf{z}_{it}) = \\ &= \{\mathbf{B}_1(\overline{\mathbf{z}}_i - \psi)\prime + (\mathbf{B}_2 + \mathbf{B}_3\overline{\mathbf{z}}_i\prime)\overline{v}_{2i}\} + (\mathbf{B}_4 + \mathbf{B}_5\mathbf{z}_{it}\prime)v_{2it}, \end{aligned} \tag{2.14}$$

where $\psi \equiv \mathrm{E}(\overline{\mathbf{z}}_i)$, $\mathbf{B}_2$ and $\mathbf{B}_4$ are $K \times 1$, $\mathbf{B}_j$, $j = 1,3,5$, are $K \times L$ matrices of constants, respectively, $\overline{\mathbf{z}}_i$ and $\overline{v}_{2i}$ are defined above. Clearly, the right hand side of equation (2.14) is identical to equation (2.9) when $\mathbf{B}_4 = \mathbf{B}_5 = 0$.

Then, similar to the case of the time-invariant individual heterogeneity, we take the expectation of equation (2.12) with respect to $(\mathbf{z}_{i1}, ..., \mathbf{z}_{iT}, v_{2i1}, ..., v_{2iT})$, employ that y_{2it} is a deterministic function of $(\mathbf{z}_{i1}, ..., \mathbf{z}_{iT}, v_{2i1}, ..., v_{2iT})$, and use assumptions (2.3) through (2.7), (2.13), and (2.14). The resulting estimating equation is:

$$\begin{aligned}\mathrm{E}(y_{1it}|\mathbf{z}_{i1}, ..., \mathbf{z}_{iT}, v_{2i1}, ..., v_{2iT}) &= \mathbf{w}_t\alpha + (\overline{\mathbf{z}}_i \otimes \mathbf{w}_t)\alpha_1 + \overline{v}_{2i}\mathbf{w}_t\alpha_2 + \\ +\overline{v}_{2i}(\overline{\mathbf{z}}_i \otimes \mathbf{w}_t)\alpha_3 + \mathbf{x}_{it}\beta + ((\overline{\mathbf{z}}_i - \psi) \otimes \mathbf{x}_{it})\beta_1 &+ \overline{v}_{2i}\mathbf{x}_{it}\beta_2 + \overline{v}_{2i}(\overline{\mathbf{z}}_i \otimes \mathbf{x}_{it})\beta_3 + \\ +v_{2it}\mathbf{x}_{it}\beta_4 + v_{2it}(\mathbf{z}_{it} \otimes \mathbf{x}_{it})\beta_5 &+ \rho_1 v_{2it}, \quad (2.15)\end{aligned}$$

where t=1,...,T, $\alpha_1 = \mathrm{vec}[\mathbf{A}_1]$, $\alpha_2 = \mathrm{vec}[\mathbf{A}_2] + (\rho_2\ 0\ ...\ 0)\prime$, where $(\rho_2\ 0\ ...\ 0)\prime$ is a $J \times 1$ vector, $\alpha_3 = \mathrm{vec}[\mathbf{A}_3]$, $\beta_j = \mathrm{vec}[\mathbf{B}_j]$, $j = \overline{1,5}$. Once again, equation (2.15) is an estimating equation for obtaining consistent estimates of APE, β. When $w_t \equiv 1$, $t = 1, ..., T$, equation (2.15) simplifies to:

$$\begin{aligned}\mathrm{E}(y_{1it}|\mathbf{z}_{i1}, ..., \mathbf{z}_{iT}, v_{2i1}, ..., v_{2iT}) &= \alpha + \overline{\mathbf{z}}_i\alpha_1 + \overline{v}_{2i}\alpha_2 + \overline{v}_{2i}\overline{\mathbf{z}}_i\alpha_3 + \\ +\mathbf{x}_{it}\beta + ((\overline{\mathbf{z}}_i - \psi) \otimes \mathbf{x}_{it})\beta_1 &+ \overline{v}_{2i}\mathbf{x}_{it}\beta_2 + \overline{v}_{2i}(\overline{\mathbf{z}}_i \otimes \mathbf{x}_{it})\beta_3 + \\ +v_{2it}\mathbf{x}_{it}\beta_4 + v_{2it}(\mathbf{z}_{it} \otimes \mathbf{x}_{it})\beta_5 &+ \rho_1 v_{2it}, \ t = 1, ..., T. \quad (2.16)\end{aligned}$$

2.3 Estimating Procedure and Calculation of Standard Errors

We employ the control function approach that uses the reduced form error terms, v_{2it}, as "control variables" for heterogeneity and endogeneity in the structural model. A two-step method that consistently estimates the parameters from equation (2.11) is the following:

1. Run the POLS regression of

$$h(y_{2it}) \text{ on } 1,\ \mathbf{z}_{it},\ \overline{\mathbf{z}}_i,\ i = 1, ..., N,\ t = 1, ..., T, \tag{2.17}$$

and save the residuals, $\widehat{v}_{2it}$, $i = 1, ..., N$, $t = 1, ..., T$. Obtain $\overline{\widehat{v}}_{2i} = T^{-1}\sum_{t=1}^{T}\widehat{v}_{2it}$, $i = 1, ..., N$.

2. Run the POLS regression of

$$\begin{aligned} &y_{1it} \text{ on } \mathbf{w}_t,\ \text{vec}[(\overline{\mathbf{z}}_i \otimes \mathbf{w}_t)]',\ \overline{\widehat{v}}_{2i}\mathbf{w}_t,\ \text{vec}[(\overline{\mathbf{z}}_i \otimes \mathbf{w}_t)]'\overline{\widehat{v}}_{2i}, \\ &\mathbf{x}_{it},\ \text{vec}[(\overline{\mathbf{z}}_i - \overline{\mathbf{z}}) \otimes \mathbf{x}_{it}]',\ \overline{\widehat{v}}_{2i}\mathbf{x}_{it},\ \text{vec}[(\overline{\mathbf{z}}_i \otimes \mathbf{x}_{it})]'\overline{\widehat{v}}_{2i}, \\ &\widehat{v}_{2it}\mathbf{x}_{it},\ \text{vec}[(\mathbf{z}_{it} \otimes \mathbf{x}_{it})]'\widehat{v}_{2it},\ \widehat{v}_{2it}, \end{aligned} \tag{2.18}$$

where $i = 1, ..., N$, $t = 1, ..., T$, $\overline{\mathbf{z}} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{z}_{it}$, and obtain $\hat{\beta}$ and the other parameter estimates. Terms containing the vec operator are used to denote all possible interactions among the variables. For example, term $\text{vec}[(\mathbf{z}_{it} \otimes \mathbf{x}_{it})]'\widehat{v}_{2it}$ in (2.18) consists of $K * L$ interaction terms.

If we want to test whether the data exhibit the properties of time-varying or time-constant individual heterogeneity, we can employ a test of joint significance of β_j, $j = 4, 5$ in (2.15). The null hypothesis of time-constant individual heterogeneity is $\text{H}_0 : \beta_4 = \beta_5 = \mathbf{0}$. A fully robust adjusted Wald statistic is appropriate. If the Wald test rejects the null hypothesis then the model with time-varying individual heterogeneity – (2.15) – should be estimated.

To test for endogeneity of y_{2it} and individual heterogeneity we can simply test for joint significance of all the second-stage terms other than $\mathbf{w}_t$ and $\mathbf{x}_{it}$. By construction, the errors from the second stage of the estimating procedure are zero mean independent of all the explanatory variables on that stage. As a result, the POLS estimates of the second-stage parameters will be consistent, and a standard F test of joint significance of all the second-stage terms containing the first-stage residuals and time-demeaned exogenous variables, $\overline{\mathbf{z}}_i$, will be a valid test. If the coefficients of

all the terms from the second stage that contain the generated regressors and time-averaged instruments are statistically jointly different from zero, there is endogeneity and heterogeneity problem, and neglecting it will lead to misspecification.

If the null hypothesis of no endogeneity and no individual heterogeneity is rejected, the standard errors in (2.18) should be adjusted for the first-stage estimation of $\delta_2 = (\delta_{20}, \delta_{21}\prime, \delta_{22}\prime)\prime$, a $(2L+1) \times 1$ vector of the first stage parameters in (2.17). Define $\mathbf{g}_{it}$ to be a $1 \times H$, $H = (2J + 3K)(1 + L) + 1$, vector of all the independent second-stage variables, i.e., $\mathbf{g}_{it} = (\mathbf{w}_t, \text{vec}[(\bar{\mathbf{z}}_i \otimes \mathbf{w}_t)]', \bar{v}_{2i}\mathbf{w}_t, \text{vec}[(\bar{\mathbf{z}}_i \otimes \mathbf{w}_t)]'\bar{v}_{2i}, \mathbf{x}_{it}, \text{vec}[(\bar{\mathbf{z}}_i - \bar{\mathbf{z}}) \otimes \mathbf{x}_{it}]', \bar{v}_{2i}\mathbf{x}_{it}, \text{vec}[(\bar{\mathbf{z}}_i \otimes \mathbf{x}_{it})]'\bar{v}_{2i}, v_{2it}\mathbf{x}_{it}, \text{vec}[(\mathbf{z}_{it} \otimes \mathbf{x}_{it})]'v_{2it}, v_{2it})$. Let $\hat{\mathbf{g}}_{it}$ to be a $1 \times H$ vector $\mathbf{g}_{it}$ that contains the estimated first-stage residuals, $\widehat{v}_{2it}$: $\hat{\mathbf{g}}_{it} = (\mathbf{w}_t, \text{vec}[(\bar{\mathbf{z}}_i \otimes \mathbf{w}_t)]', \bar{\widehat{v}}_{2i}\mathbf{w}_t, \text{vec}[(\bar{\mathbf{z}}_i \otimes \mathbf{w}_t)]'\bar{\widehat{v}}_{2i}, \mathbf{x}_{it}, \text{vec}[(\bar{\mathbf{z}}_i - \bar{\mathbf{z}}) \otimes \mathbf{x}_{it}]', \bar{\widehat{v}}_{2i}\mathbf{x}_{it}, \text{vec}[(\bar{\mathbf{z}}_i \otimes \mathbf{x}_{it})]'\bar{\widehat{v}}_{2i}, \widehat{v}_{2it}\mathbf{x}_{it}, \text{vec}[(\mathbf{z}_{it} \otimes \mathbf{x}_{it})]'\widehat{v}_{2it}, \widehat{v}_{2it})$. Then, the estimating equation (2.18) can be rewritten as $y_{1it} = \mathbf{g}_{it}\theta + e_{it}$, where $\text{E}(e_{it}|\mathbf{z}_{i1}, ..., \mathbf{z}_{iT}, v_{2i1}, ..., v_{2iT}) = 0$, and θ is a column of all the parameters from the estimating equation. Define $\mathbf{y}_{1i}$ to be the $T \times 1$ vector of y_{1it}, let $\mathbf{G}_i$ be the matrix with t^{th} row $\mathbf{g}_{it}$, and $\hat{\mathbf{G}}_i$ be the matrix with t^{th} row $\hat{\mathbf{g}}_{it}$. Then, θ can be estimated as:

$$\hat{\theta} = (\sum_{i=1}^{N} \hat{\mathbf{G}}_i'\hat{\mathbf{G}}_i\ )^{-1}(\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{\mathbf{g}}_{it}' y_{1it}). \tag{2.19}$$

Write $y_{1it} = \hat{\mathbf{g}}_{it}\theta + (\mathbf{g}_{it} - \hat{\mathbf{g}}_{it})\theta + e_{it} = \hat{\mathbf{g}}_{it}\theta + \theta'(\mathbf{g}_{it} - \hat{\mathbf{g}}_{it})' + e_{it}$. Plugging this in (2.19) and multiplying through by $\sqrt{N}$ gives

$$\sqrt{N}(\hat{\theta} - \theta) = \hat{\mathbf{A}}^{-1} N^{-1/2} \sum_{i=1}^{N}\sum_{t=1}^{T} \hat{\mathbf{g}}_{it}'[\theta'(\mathbf{g}_{it} - \hat{\mathbf{g}}_{it})' + e_{it}],$$

where $\hat{\mathbf{A}} = N^{-1}\sum_{i=1}^{N} \hat{\mathbf{G}}_i'\hat{\mathbf{G}}_i$. Using the Law of Large Numbers, we know that $\hat{\mathbf{A}} \xrightarrow{p} \mathbf{A} \equiv \text{E}(\mathbf{G}_i'\mathbf{G}_i)$. Further, a mean value expansion gives

$$N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{\mathbf{g}}_{it}' e_{it} = N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T} \mathbf{g}_{it}' e_{it} + [N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} (\nabla_{\delta_2}\mathbf{g}_{it})e_{it}]\sqrt{N}(\hat{\delta}_2 - \delta_2) + o_p(1),$$

where $\nabla_{\delta_2}\mathbf{g}_{it}$ is the $H \times (2L+1)$ Jacobian of $\mathbf{g}'_{it}$ with respect to the parameters δ_2 from the first stage of the estimating procedure. For each (i,t), $\nabla_{\delta_2}\mathbf{g}_{it}$ is a block matrix of the form:

$$\begin{pmatrix} -1 \\ -\overline{z}_{1i} \\ \vdots \\ -\overline{z}_{Li} \\ -\overline{z}_{1i} \\ \vdots \\ -\overline{z}_{Li} \end{pmatrix} \begin{pmatrix} 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \mathbf{w}_t & \mathbf{w}_t & \dots & \mathbf{w}_t & \mathbf{w}_t & \dots & \mathbf{w}_t \\ (\overline{\mathbf{z}}_i \otimes \mathbf{w}_t) & (\overline{\mathbf{z}}_i \otimes \mathbf{w}_t) & \dots & (\overline{\mathbf{z}}_i \otimes \mathbf{w}_t) & (\overline{\mathbf{z}}_i \otimes \mathbf{w}_t) & \dots & (\overline{\mathbf{z}}_i \otimes \mathbf{w}_t) \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \mathbf{x}_{it} & \mathbf{x}_{it} & \dots & \mathbf{x}_{it} & \mathbf{x}_{it} & \dots & \mathbf{x}_{it} \\ (\overline{\mathbf{z}}_i \otimes \mathbf{x}_{it}) & (\overline{\mathbf{z}}_i \otimes \mathbf{x}_{it}) & \dots & (\overline{\mathbf{z}}_i \otimes \mathbf{x}_{it}) & (\overline{\mathbf{z}}_i \otimes \mathbf{x}_{it}) & \dots & (\overline{\mathbf{z}}_i \otimes \mathbf{x}_{it}) \\ \mathbf{x}_{it} & \mathbf{x}_{it} & \dots & \mathbf{x}_{it} & \mathbf{x}_{it} & \dots & \mathbf{x}_{it} \\ (\mathbf{z}_{it} \otimes \mathbf{x}_{it}) & (\mathbf{z}_{it} \otimes \mathbf{x}_{it}) & \dots & (\mathbf{z}_{it} \otimes \mathbf{x}_{it}) & (\mathbf{z}_{it} \otimes \mathbf{x}_{it}) & \dots & (\mathbf{z}_{it} \otimes \mathbf{x}_{it}) \\ 1 & 1 & \dots & 1 & 1 & \dots & 1 \end{pmatrix}.$$

Each row of the jacobian matrix corresponds to each addendum in estimating equation (2.15). Because $\mathrm{E}(e_{it}|\mathbf{z}_{i1},...,\mathbf{z}_{iT},v_{2i1},...,v_{2iT}) = 0$, $\mathrm{E}((\nabla_{\delta_2}\mathbf{g}_{it})'e_{it}) = 0$. It follows that

$$N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\nabla_{\delta}\mathbf{g}_{it})e_{it} = o_p(1),$$

and, since $\sqrt{N}(\hat{\delta}_2 - \delta_2) = O_p(1)$, we get $N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\mathbf{g}}'_{it}e_{it} = N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{g}'_{it}e_{it} + o_p(1)$.

Next, using similar reasoning,

$$\begin{aligned} N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\mathbf{g}}'_{it}\theta'(\mathbf{g}_{it} - \hat{\mathbf{g}}_{it})' &= -[N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{g}'_{it}\theta'(\nabla_{\delta_2}\mathbf{g}_{it})]\sqrt{N}(\hat{\delta}_2 - \delta_2) + o_p(1) = \\ &= -\mathbf{B}\sqrt{N}(\hat{\delta}_2 - \delta_2) + o_p(1), \end{aligned}$$

where $\mathbf{B} = \mathrm{E}(\sum_{t=1}^{T}\mathbf{g}'_{it}\theta'(\nabla_{\delta_2}\mathbf{g}_{it}))$. Further, based on the first stage of the estimation procedure - (2.17) - we know that

$$\sqrt{N}(\hat{\delta}_2 - \delta_2) = \mathbf{C}^{-1}N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{z}^F_{it})'v_{2it} + o_p(1),$$

where $\mathbf{C} \equiv \sum_{t=1}^{T}\mathrm{E}[(\mathbf{z}^F_{it})'\mathbf{z}^F_{it}]$, $\mathbf{z}^F_{it} = (1, \mathbf{z}_{it}, \overline{\mathbf{z}}_i)$ is a $1 \times (2L+1)$ vector of the first stage explanatory variables, i.e., it is a vector containing a constant, exogenous ex-

planatory variables, $\mathbf{z}_{it}$, and, time averages of the exogenous explanatory variables, $\bar{\mathbf{z}}_i = T^{-1}\sum_{t=1}^{T}\mathbf{z}_{it}$, and $\mathrm{E}((\mathbf{z}_{it}^F)'v_{2it}) = 0$, $t = 1, ..., T$. Thus, collecting all the terms we obtain

$$\sqrt{N}(\hat{\theta} - \theta) = \mathbf{A}^{-1}N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}[\mathbf{g}_{it}'e_{it} - \mathbf{B}\mathbf{C}^{-1}(\mathbf{z}_{it}^F)'v_{2it}] + o_p(1).$$

By the Central Limit Theorem,

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} \mathrm{Normal}(\mathbf{0}, \mathbf{A}^{-1}\mathbf{M}\mathbf{A}^{-1}),$$

where $\mathbf{M} \equiv \mathrm{Var}\sum_{t=1}^{T}(\mathbf{g}_{it}'e_{it} - \mathbf{B}\mathbf{C}^{-1}(\mathbf{z}_{it}^F)'v_{2it})$. Therefore, the asymptotic variance of $\hat{\theta}$, $\mathrm{Avar}(\hat{\theta})$, is estimated as

$$\widehat{\mathrm{V}} \equiv \hat{\mathbf{A}}^{-1}\hat{\mathbf{M}}\hat{\mathbf{A}}^{-1}/N, \tag{2.20}$$

where $\hat{\mathbf{A}}$ is defined above,

$$\hat{\mathbf{M}} = N^{-1}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}(\hat{\mathbf{g}}_{it}'\hat{e}_{it} - \hat{\mathbf{B}}\hat{\mathbf{C}}^{-1}(\mathbf{z}_{it}^F)'\hat{v}_{2it})\right)\left(\sum_{t=1}^{T}(\hat{\mathbf{g}}_{it}'\hat{e}_{it} - \hat{\mathbf{B}}\hat{\mathbf{C}}^{-1}(\mathbf{z}_{it}^F)'\hat{v}_{2it})\right)',$$

$\hat{\mathbf{B}} = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\mathbf{g}}_{it}'\hat{\theta}'(\nabla_{\delta_2}\hat{\mathbf{g}}_{it})$, $\hat{\mathbf{C}} = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{z}_{it}^F)'\mathbf{z}_{it}^F$, and $\hat{e}_{it} = y_{1it} - \hat{\mathbf{g}}_{it}\hat{\theta}$.

2.4 Finite Sample Behavior of the Control Function Estimator

In this section we provide evidence on the finite sample properties of the control function estimator of the APE in CRC balanced panel data models. We assume that the unobserved heterogeneity is time constant. This assumption allows us to compare the proposed estimation method with other available estimators in the same context, and time constant slopes are commonly assumed in many empirical applications. So, we consider two CRC panel data models with time-constant unobserved heterogeneity

described by equation (2.10). First, we study the usual unobserved effect CRC model with a random coefficient, i.e., we assume $w_t \equiv 1$, $t = 1, ..., T$. Second, we employ the random trend CRC model with $\mathbf{w}_t \equiv (1, t)$, $t = 1, ..., T$, so that each cross-sectional unit has its own linear trend. We use Monte Carlo simulations to draw the data and check the properties of the estimator. The number of replications is 500, and the results of the experiments are presented for samples of 500 and 1000 observations for a time horizon $T = 5$. The population values of the model parameters are set at $\beta = 2$ and $\alpha = 1$. We consider two options for a scalar endogenous explanatory variable y_{2it}: (1) a continuous y_{2it} with a large support set, i.e., it is a traditional continuous variable, and (2) y_{2it} being a fraction in the open unit interval, $y_{2it} \in (0, 1)$, i.e., it is a roughly continuous variable.

For the usual unobserved effect CRC model the dependent variable y_{1it} is generated as:

$$y_{1it} = a_i + y_{2it} b_i + u_{it}, \quad t = 1, ..., T, \tag{2.21}$$

where

$$a_i \equiv \alpha + \lambda_{1a} \overline{z}_i + \lambda_{2a} \overline{v}_{2i} + \lambda_{3a} \overline{v}_{2i} \overline{z}_i + \lambda_{4a} e_i^a, \tag{2.22}$$

$$b_i \equiv \beta + \lambda_{1b} (\overline{z}_i - \overline{z}) + \lambda_{2b} \overline{v}_{2i} + \lambda_{3b} \overline{v}_{2i} \overline{z}_i + \lambda_{4b} e_i^b, \tag{2.23}$$

and

$$u_{it} \equiv \lambda_{1u} v_{2it} + \lambda_{2u} \overline{v}_{2i} + \lambda_{3u} e_{it}^u, \tag{2.24}$$

where $z_{it} \sim \text{Normal}(t, 1)$, $v_{2it} \sim \text{Normal}(0, 1)$, e_i^a, $e_i^b \sim \text{Normal}(0, 1)$, $e_{it}^u \sim \text{Normal}(0, 1)$, $\overline{z}_i = T^{-1} \sum_{t=1}^{T} z_{it}$, $\overline{z} = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} z_{it}$, $\overline{v}_{2i} = T^{-1} \sum_{t=1}^{T} v_{2it}$, λ_{1a}, λ_{2a}, λ_{3a}, λ_{4a}, λ_{1b}, λ_{2b}, λ_{3b}, λ_{4b}, λ_{1u}, λ_{2u}, and λ_{3u} are constants.

For a continuous y_{2it} on a large support set we define the endogenous explanatory

variable y_{2it} to be $y_{2it} \equiv h(y_{2it}) \equiv g_{2it}$, where we generate g_{2it} according to:

$$g_{2it} \equiv \lambda_{g_2 z} z_{it} + \xi z_{it} d_i + \lambda_{g_2 v_2} v_{2it}, \tag{2.25}$$

where $v_{2it} \sim \text{Normal}(0,1)$, $d_i = b_i - \beta$, $\lambda_{g_2 z}$, ξ, and $\lambda_{g_2 v_2}$ are constants. For $y_{2it} \in (0,1)$, we use the following equality to define the endogenous regressor:

$$y_{2it} \equiv \frac{\exp(g_{2it})}{1 + \exp(g_{2it})}. \tag{2.26}$$

If we set ξ to be 0 in (2.25) then the condition for consistency of the FE-IV estimator of CRC panel data models in Murtazashvili and Wooldridge (2005) will be satisfied. When $\xi \neq 0$, the covariance between the detrended endogenous explanatory variable, $\ddot{y}_{2it}$, and the unobserved heterogeneity, $b_i = \beta + d_i$, conditional on the detrended instrument, $\ddot{z}_{it}$, is not equal its unconditional version: $\text{Cov}(\ddot{y}_{2it}, b_i | \ddot{z}_{it}) \neq \text{Cov}(\ddot{y}_{2it}, b_i)$. Thus, for $\xi \neq 0$, the FE-IV estimation in Murtazashvili and Wooldridge (2005) does not deliver consistent estimators of the model parameters. While (2.25) does not meet the requirements for consistent FE-IV estimation of (2.21) when $\xi \neq 0$, it does satisfy (2.3) through (2.14) and does allow using the CF approach to obtain consistent parameters' estimates in (2.21).

For the random trend CRC model the dependent variable y_{1it} is generated as:

$$y_{1it} = a_{1i} + a_{2i} t + y_{2it} b_i + u_{it}, \; t = 1, ..., T, \tag{2.27}$$

where both a_{1i} and a_{2i} are generated according to (2.22), b_i, y_{2it}, and u_{it} are also defined above.

Why would we think that the data generating process we propose in (2.22) through (2.25) is representative of something that we might actually see in practice? One of possible empirical examples can be a study by Hall and Jones (1999). The authors attempt to explain the differences in output per worker by differences in institutions and government policies, which they call social infrastructure. Even though Hall and

Jones (1999) do a cross-sectional investigation, their idea can be easily extended to a panel data setup. Social infrastructure is thought to be endogenous. First of all, it can depend itself on the level of GDP per worker in a country. Secondly, we do not observe social infrastructure directly, and need to deal with a measurement error problem. Hall and Jones (1999) suggest using Western European influence around the world as an instrumental variable for social infrastructure. Specifically, a distance of a country from the equator and a fraction of population speaking a European language are used as measures of Western European influence. Clearly, the distance of a country from the equator is time-invariant. Instead,we can use a time-varying fraction of population speaking a European language as an IV in a panel data setting. While both models (2.21) and (2.27) can be thought appropriate, perhaps, structural equation (2.27) should seem more suitable for modeling a behaviour of output per worker, since we want to allow each country to have its own time trend. Further, endogeneity of social infrastructure explains equations (2.24) and (2.25). Country-specific unobserved cultural characteristics, both additive and multiplicative, might be related to the fraction of population speaking a European language. It is Western Europe who distributed to the rest of the world the ideas of Adam Smith and the importance of property rights (among others). As a result, countries that were influenced by Western Europe the most are more likely to have favorable social infrastructure. This would explain the linear terms in equations (2.22) and (2.23). Importantly, it is possible that the joint distributions of (a_i, v_{2it}) given $\mathbf{z}_i$ and (b_i, v_{2it}) given $\mathbf{z}_i$ can depend on $\mathbf{z}_i$ due to heteroskedasticity in $\text{Var}(a_i, v_{2it}|\mathbf{z}_i)$ or $\text{Var}(b_i, v_{2it}|\mathbf{z}_i)$, where $j = 1$ or 2, as discussed by Card (2001) for the human capital earnings model. That is why we might think that the interaction terms in (2.22), (2.23), and (2.25) are required.

Table B.1 and Table B.2 present experimental results for the CRC model with a continuous scalar endogenous explanatory variable with a large support set and a scalar instrument z_{it}. Table B.1 reports the simulation outcomes for the usual

unobserved effect CRC model, while Table B.2 covers the case of the random trend CRC model. For the usual unobserved effect model, column 2 contains the sample correlation coefficients among the endogenous regressor, y_{2it}, and the instrument, z_{it}, the error, u_{it}, the unobserved additive effect, a_i, and the unobserved multiplicative heterogeneity, b_i, denoted $\hat{\rho}_{y_2z}$, $\hat{\rho}_{y_2u}$, $\hat{\rho}_{y_2a}$, $\hat{\rho}_{y_2b}$, respectively, because analytical expressions are not readily available. For the random trend model, we report the sample correlations between y_{2it} and a_{1i}, and between y_{2it} and a_{2i} separately. We denote these sample correlations $\hat{\rho}_{y_2a_1}$, $\hat{\rho}_{y_2a_2}$, respectively. $\hat{\rho}_{y_2b}$ is reported for $t = 1$.[2]

Columns 3 through 10 contain the mean, regular standard error (Reg. SE), robust standard error (Rob. SE)[3], standard deviation (SD), root mean squared error (RMSE), lower quartile (LQ), median, and upper quartile (UQ) of the APE estimates from 500 replications. Rows of the table report statistics for the usual pooled ordinary least squares (POLS) estimates on the original data, the usual fixed effects estimates (FE-OLS), which is just pooled OLS on the time-demeaned data, the instrumental variables (IV) estimates using the original data, the fixed effects-instrumental variables estimates (FE-IV), and the estimates from the control function approach (CF). Adjusted standard error (Adj. SE) is reported for the CF approach.

It is easy to see that when $\xi = 0$ and the endogenous explanatory variable y_{2it} is continuous on a large support set, i.e., y_{2it} is defined by (2.25) for $\xi = 0$, the (conditional and unconditional) covariance between the detrended endogenous regressor and the unobserved heterogeneity is constant over time. Even though Murtazashvili and Wooldridge (2005) emphasize that the FE-IV estimator should contain a full set of time dummies to deliver consistent estimates, they do so allowing the covariances

[2]When $\xi \neq 0$, Table B.1 and Table B.2 are obtained for $\lambda_{1a} = \lambda_{2a} = \lambda_{3a} = 0.29$, $\lambda_{4a} = 0.84$, $\lambda_{1b} = \lambda_{2b} = \lambda_{3b} = 0.2$, $\lambda_{4b} = 0.99$, $\lambda_{1u} = \lambda_{2u} = 0.37$, $\lambda_{3u} = 0.88$, $\lambda_{g_2z} = 0.44$, $\xi = 0.55$, and $\lambda_{g_2v_2} = 0.71$. When $\xi = 0$, $\lambda_{1a} = \lambda_{2a} = \lambda_{3a} = 0.31$, $\lambda_{4a} = 0.82$, $\lambda_{1b} = \lambda_{2b} = \lambda_{3b} = 0.61$, $\lambda_{4b} = 0.91$, $\lambda_{1u} = \lambda_{2u} = 0.2$, $\lambda_{3u} = 0.96$, $\lambda_{g_2z} = 0.26$, and $\lambda_{g_2v_2} = 0.97$ are used for Table B.1 and Table B.2.

[3]Robust standard errors are calculated using the scaling factor from Stata 9.0, i.e., they are clustered on individuals.

to vary with time while still being independent of the detrended instruments. Thus, for the usual unobserved effect model, when we define y_{2it} according to (2.25), there is no need to include time dummies to obtain consistent FE-IV estimates of β when $\xi = 0$. As a result, all the estimates we consider for the usual unobserved effect model including the FE-IV estimates are based on the regressions without the time dummies. For the random trend CRC model, all the reported estimates (but the CF) are based on the regressions with the time dummies.

There are three sources of bias in the estimates under consideration. First, the correlation between the unobserved heterogeneity a_i and the regressor y_{2it} results in the biased estimates of the model parameters. Second, the endogeneity of the regressor y_{2it} also explains why the estimates we consider are biased. Finally, the correlation between the regressor y_{2it} and the random coefficient b_i leads to the bias (and inconsistency) in the estimates, as well. As long as $\xi = 0$ in (2.25), the correlation between the endogenous explanatory variable and the random coefficient does not result in the inconsistency of the FE-IV estimator. When $\xi = 0$, both the FE-IV and the CF methods deliver consistent estimates of β. When $\xi \neq 0$, the FE-IV estimates of β are both biased and inconsistent. The CF estimates, while being biased, are the only consistent estimates considered for $\xi \neq 0$.

Columns 4 and 5 contain regular and robust standard errors of the estimates. To be exact, we report the averages of the regular and robust standard errors of the estimates obtained from 500 replications. The regular SE are the standard errors calculated under assumption that there are no heteroskedasticity and serial correlation in the error terms. The robust SE are adjusted for both serial correlation and heteroskedasticity that are possibly present in the errors. Standard errors reported for the CF approach are the standard errors, which are computed according to formula (2.20), and which are the standard error adjusted for the first stage estimation and which are robust to arbitrary serial correlation and heteroskedasticity. As expected,

the simulations show that the robust standard errors for the first four estimators are surely better estimates of the standard deviations than the regular standard errors are.

Studying Table B.1 and Table B.2 for both sample sizes in case of $\xi \neq 0$, we conclude that the CF estimates have the smallest biases and the smallest RMSEs. Murtazashvili and Wooldridge (2005) show that the FE-IV estimation results in consistent estimates of β, when $\xi = 0$. Table B.1 and Table B.2 indicate that the CF estimator and the FE-IV estimator have very similar RMSEs for $\xi = 0$. Closeness in RMSEs comes from similarity in both biases and standard deviations of these estimators. The CF estimator has a smaller standard deviation than the FE-IV estimator in all the cases considered in Tables B.1 and B.2. For instance, when $\xi = 0$ and $N = 1000$, the standard deviation of the FE-IV estimator is about 21% higher than the standard deviation of the CF estimator. Efficiency of the CF estimators comes from the assumed specific functional forms for the endogenous variable and the random coefficients of equation (2.21).

For the next set of simulations, we take y_{2it} being a fraction in the open unit interval. Because the structural model (2.27) contains a trend, the default is to include a full set of time period dummies in every estimation technique but the CF approach. Table B.3 and Table B.4 show the simulation results for $y_{2it} \in (0, 1)$.[4] Now, the bias in the CF estimate (and all other estimates) is more pronounced, even though the CF estimating method results in the smallest bias of $\hat{\beta}$ among all the estimators. For example, when we consider the random trend model, for $y_{2it} \in (0, 1)$ when $\xi \neq 0$, with $\hat{\rho}_{y_2 b} = .581$, and $N = 500$, the CF estimate of β is 2.229 with the RMSE of .750, compared to $\hat{\beta}_{CF} = 2.039$ with the RMSE of .182 with $\hat{\rho}_{y_2 b} = .578$,

[4]When $\xi \neq 0$, Table B.3 and Table B.4 are obtained for $\lambda_{1a} = \lambda_{2a} = \lambda_{3a} = 0.21$, $\lambda_{4a} = 0.92$, $\lambda_{1b} = 0.2$, $\lambda_{2b} = \lambda_{3b} = 0.7$, $\lambda_{4b} = 0.94$, $\lambda_{1u} = \lambda_{2u} = 0.43$, $\lambda_{3u} = 0.84$, $\lambda_{g_2 z} = 0.87$, $\xi = 0.48$, and $\lambda_{g_2 v_2} = 0.11$. When $\xi = 0$, $\lambda_{1a} = \lambda_{2a} = \lambda_{3a} = 0.31$, $\lambda_{4a} = 0.82$, $\lambda_{1b} = 0.2$, $\lambda_{2b} = \lambda_{3b} = 0.7$, $\lambda_{4b} = 0.94$, $\lambda_{1u} = \lambda_{2u} = 0.22$, $\lambda_{3u} = 0.96$, $\lambda_{g_2 z} = 0.26$, and $\lambda_{g_2 v_2} = 0.97$ are used for Table B.3 and Table B.4 .

and $N = 500$, for y_{2it} with a large support set.

As expected, when $\xi = 0$, it is the CF approach that has the smallest bias among all the estimators under consideration since we simulate our dataset to satisfy the assumptions (2.5) through (2.7). However, when $\xi = 0$, the evidence on the RMSEs of the CF and the FE-IV estimators is mixed. On the one hand, the RMSE of the FE-IV method is either clearly smaller or only marginally bigger than the RMSE of the CF method. For example, for the random trend model with a roughly continuous regressor, the RMSE of the FE-IV estimator is 0.599 vs. the RMSE of 0.875 of the CF estimator when $N = 500$. The random trend model with a continuous explanatory variable results in the RMSE of 0.235 for the FE-IV estimator and the RMSE of 0.223 for the CF estimator when $N = 500$. On the other hand, the bias of the FE-IV estimator is clearly much more severe. The differences between the FE-IV and the CF approaches for the random trend model with $y_{2it} \in (0, 1)$ and $\xi = 0$ illustrate the trade-off between bias and efficiency. For $\xi = 0$, both the CF method and the FE-IV approach are consistent. Further, the simulations in Tables B.3 and B.4 show that the FE-IV estimates are always less variable than the CF estimates for the random trend model with $y_{2it} \in (0, 1)$. However, when $\xi = 0$, the bias in the CF estimator is significantly less than the bias in the FE-IV. Overall, the simulation findings in Tables B.3 and B.4 support the idea that the CF estimating method produces more desirable estimates of β when $\xi \neq 0$.

Applied economists are quite often reluctant to use control function methods since control function approaches require the calculation of the adjusted standard errors, which is not routinely done in standard econometric packages. Table B.5 contains detailed information (which can be partially seen in Tables B.1 through B.4) on the standard errors of the control function estimates of the APEs for the two models and the two cases of the endogenous explanatory variable considered. Columns 1 shows whether ξ is different from zero. Column 2 reports the cross-sectional sample size.

Columns 3 through 7 contain the mean, regular standard error (Reg. SE), robust standard error (Rob. SE), adjusted standard error (Adj. SE), and standard deviation (SD) of the APE estimates from the CF approach from 500 replications. The regular SE are the standard errors from the second stage estimation for the CF approach without adjustment for heteroskedasticity and serial correlation and without taking the first stage estimation into account. The robust SE are the second stage standard errors from the CF method, which are robust to both serial correlation and heteroskedasticity, and which are obtained ignoring the first stage estimation. The adjusted SE are the only standard errors which are adjusted for the first stage estimation (they are calculated according to formula (2.20)). Clearly, besides being the only theoretically appropriate estimates of the standard errors, the adjusted standard errors based on (2.20) approximate the standard deviations the best among the three standard errors considered.

To summarize, the simulation findings verify that when the joint distribution of (a_i, b_i, v_{2it}) given z_{it} depends on z_{it} the most robust estimator of the average partial effect in a correlated random coefficient balanced panel data model is the control function estimator from the two-step estimating method (2.17) - (2.18).

2.5 Empirical Application to Effects of Job Training on Worker Productivity

The method we propose for estimating the average partial effects from a correlated random coefficient panel data model is developed for large N small T framework. However, real-life data limitations quite often do not allow researchers to use "truly" large N datasets. Here, we follow a common real-life situation with a not so large N dimension of the available data. Suppose we want to estimate an average partial effect

of job training on worker performance measured by output scrap rates. Holzer, Block, Cheatham, and Knott (1993) explore the effects of a state-financed training grant program for manufacturing firms in Michigan using constant coefficient models. They use a three-year panel of data (1987–1989) from a unique survey of firms in Michigan that applied for training grants under the state's Michigan Job Opportunity Bank-Upgrade (MJOB) program. This program was designed to provide one-time grants to eligible firms. An eligible firm was defined as a manufacturing company with 500 or fewer employees that was implementing new technology and had not received a grant before. Let us estimate the effects of on-job-training on worker productivity allowing for both additive and multiplicative unobserved firm-specific effects.

Why would we think that the random coefficient panel data model might be appropriate in this context? A possible justification for using a RC model can be that some unobserved firm characteristics might cause firms to respond heterogeneously to the job training. For instance, an unobserved "atmosphere" in each firm might result in a heterogeneous effect of the annual hours of training per employee. Workers might feel supported and encouraged more in firms where the management promotes and advocates additional schooling and team efforts. Contrary, employees of firms with no policy on education beyond workers' current level might be discouraged to improve their present skills and effort. As a results, the same annual hours of job training in the two types of firms can lead to different outcomes of the output scrap rates. Since the effect of the job training on the worker performance might be related to the extent of the unobserved support from a firm, we should consider a correlated version of the random coefficient model

To fit the method from Section 2.2, we balance data from Holzer, Block, Cheatham, and Knott (1993), and obtain a sample on 45 firms that applied for an MJOB grant during 1988 and 1989. (The dataset is provided with Wooldridge (2002), and it is called JTRAIN.) Of these firms, 27 had received a grant and 18 had

not. The balancing of the data is made based on the availability of the data for the scrap rate (per 100 items) and the annual hours of job training per employee.

Balancing the data raises concerns that the final dataset might be a non-random sample. Table B.6 contains summary statistics from the unbalanced and balanced datasets for each of the following groups: the entire sample, firms that received a grant in either 1988 or 1989, and firms that did not receive a grant. Comparison of the two panels of Table B.6 suggests that even though the proportion of the firms that received the grant and the firms that did not receive the grant changed, there are virtually no differences between the firms in the two datasets with regard to the scrap rates and the annual hours of job training. The balanced data seem to be very close to the unbalanced dataset in preserving the information on the scrap rates and the annual hours of training per worker. Of course, we should also be concerned that some unobserved firms' characteristics played a role in formation of these two samples. Since the MJOB program distributed grants to eligible firms on a first-come, first-serve basis, we believe the grant distribution to be a fairly random process, and assume firms are not selected into the two samples based on their unobserved characteristics. Given these assumptions, we feel sufficiently confident in relying on the balanced dataset to proceed with our analysis.

Our goal is to evaluate the average partial effect of another hour of job training on worker productivity relaxing the traditional assumption of a constant effect of the annual hours of training per worker on the output scrap rate. In the context of the correlated random coefficient approach, a simple panel data model of our interest is

$$\log(scrap_{it}) = \alpha + b_{1i}hrsemp_{it} + \delta_1 d88_t + \delta_2 d89_t + a_{1i} + u_{it}, \tag{2.28}$$

where $scrap_{it}$ is firm's i's scrap rate in year t, $hrsemp_{it}$ is annual hours of on-job-training per employee, a_{1i} is a firm-specific unobserved effect, and u_{it} is an unobserved disturbance for firm i at year t. We also allow different year intercepts in our structural model. The unobserved firm fixed effect, a_{1i}, can contain unmeasured worker ability,

capital, and managerial skill, which we think of as being roughly constant over the time period we consider. Since the unobserved firm effect includes the worker ability, the annual hours of job training can be correlated with the unobserved effect. For example, firm managers might want to train workers with lower skills more to improve their productivity. Or, on the contrary, they might be interested in improving the productivity of relatively high skilled workers even more in order to utilize new hi-tech equipment that requires very well trained employees. Further, we should be concerned if u_{it} is correlated with $hrsemp_{it}$. For example, a firm might hire more skilled workers and reduce the on-job-training requirements at the same time. A possibility of measurement error in $hrsemp_{it}$ should also be considered since there might be some incentives for recipients of a grant to overstate or non-recipients to understate their training changes. If any (or both) of these is the case, we need to deal with the endogeneity of the annual hours of training in equation (2.28). Here, we exploit the fact that some firms received MJOB grants. We assume that grant designation in year t is uncorrelated with the error term u_{it} in every time period. This seems to be a reasonable assumption, since firms are eligible to receive a grant only once, and grants were distributed on a first-come, first-serve basis, which we believe to be a fairly random process. Thus, whether a grant is received or not in year t should not be related to changes in the output scrap rates in any other year directly and only through the changes in the annual hours of job training.[5] Thus, we use a dummy variable indicating whether or not a grant was received as an instrumental variable for the annual hours of training per worker provided that $hrsemp_{it}$ and $grant_{it}$ are

[5]Using a constant coefficient approach, we regress a change in the log of the scrap rates as occurring between years $t-1$ and t on a change in the annual hours of on-job-training, a change in a dummy variable indicating whether a grant was received, and a lag of this variable. The changes are taken to eliminate the firm fixed effect. The results of the regression suggest that none of the three variables are either individually or jointly statistically significant at any conventional level of significance (R^2 for the regression is 0.032 with F-statistic=0.96).

correlated.

Clearly, the variable $hrsemp_{it}$ takes on only non-negative values, and we should consider taking a logarithmic transformation of this variable to run the first stage regression. At the same time, about 27% of all the observations have zero values for the annual hours of job training. Normally, we would transform a variable x that has zero observations using $\log(1 + x)$ transformation. However, for the purpose of our method, there is no gain in using this transformation, since the new variable, $\log(1 + x)$, will take on only positive values. Thus, we choose to use the variable reporting the annual hours of on-job-training in levels.

The first stage regression results with and without different year intercepts are reported in Table B.7. Columns (1) and (2) report the results from the first stage regression with and without different year intercepts when no other variables but $grant_{it}$ are used as explanatory variables. Table B.8 reports the estimation results for equation (2.28) by FE-IV and CF methods.

Overall, the APE estimates of the annual hours of on-job-training by the CF approach for equation (2.28) are bigger than the corresponding estimates from the same equation by the FE-IV method [See columns (1) and (2) vs. columns (5) and (6) of Table B.8]. For example, the CF approach in regression (6) suggests that 10 more hours of job training per worker are estimated to reduce the scrap rate by about 37%. For the firms in the sample, the average amount of job training over the three-year period is 15.6 hours per employee, with a minimum of zero and a maximum of 154. Comparing regressions (6) and (2) from Table B.8, we can say that the FE-IV estimate of the APE of the average annual hours of job training per worker is about 18 times smaller in magnitude and is statistically insignificant. Overall, the estimates from the CF method (regressions (5) and (6)) are more statistically significant and substantially larger in magnitudes for the two regression specifications considered than the corresponding FE-IV estimates (regressions (1) and (2)).

What do the differences in the two estimation methods suggest? They suggest that even if the unobserved "atmosphere" of a firm is independent of the grant designation, the correlation between hours of on-site job training and the firm-specific unobserved effect is different for firms that received a grant and firms that did not. Indeed, it is natural to think that the effect of the unobserved worker ability, managerial skills, and a firm's "atmosphere" on hours of job training might be stronger among those firms that received grants. In other words, using the language from Section 2.2, the joint distribution of (a_i, b_i, v_{2it}) is different for those firms that received a grant and those that did not.

To address the question of rationale of the model with both additive and multiplicative individual heterogeneities in the context of this application, we test whether the coefficients of different sets of variables from the second stage are statistically different from zero. Several Wald statistics are calculated. First, the coefficients on all the explanatory variables but w_t and x_{it} (and year dummies if included) are restricted to zero. For specificity, let us use regression (6) from Table B.8. The Wald statistic for this regression equals 13.00, which allows us to reject the null hypothesis that ignoring endogeneity and heterogeneity is inappropriate for our data with p-value of 0.072 (a critical chi-squared value is 12.02 with 7 degrees of freedom at 10% level). Second, to keep in mind the endogeneity of the main explanatory variable, we restrict to zero the coefficients on all the explanatory variables but w_t, x_{it}, and $\hat{v}_{2it}$ (and year dummies if included). The Wald statistic for this test is 13.33, which exceeds a critical value of 12.59 with 6 degrees of freedom at 5% level (p-value is 0.038). And, thirdly, to reflect assumption (2.5), the coefficients on all the explanatory variables but w_t, x_{it}, $\hat{v}_{2it}$, and $\bar{\hat{v}}_{2i} w_t$ (and year dummies if included) are restricted to zero. Now, the Wald statistic is 13.23, which is above a critical value of 11.07 with 5 degrees of freedom at 5% level (p-value is 0.021).

Finally, we check whether the interaction terms between the averages of the first-

stage residuals, $\bar{\hat{v}}_{2i}$, and the averages of the exogenous variables, $\bar{z}_i$, are jointly different from zero. To do so, we consider three possibilities: first, when all such terms are jointly significant; second, when only terms originated from our assumption on the additive heterogeneity, a_{1i}, are jointly different from zero; and, third, when only terms introduced from the assumption on the multiplicative heterogeneity, b_{1i}, are jointly important. The resulting Wald statistics for regression (6) are 11.74, 2.47, and 11.32, respectively. Thus, we can reject the first and the last null hypothesizes at 1% level with chi-squared critical values of 9.21 and 6.63 for 2 and 1 degrees of freedom, respectively. And we cannot reject the hypothesis that the interaction term between the additive unobserved fixed effect and the explanatory variable is significant at any conventional level of significance (a chi-squared critical value for 1 degrees of freedom for 10% level is 2.71). Thus, there is evidence at 1% level of significance that the conditional variance-covariance matrix, $\text{Var}(b_{1i}, x_{it}|\mathbf{z}_i)$, is heteroskedastic. Overall, we can conclude that the CF approach should perform favorably in comparison with the FE-IV method for the output scrap rates application.

We are unlikely to draw conclusions about the causal effect of on-job-training on the output scrap rate having only one explanatory variable – $hrsemp_{it}$ – unless other control variables are also accounted for. We consider a logarithm of the dollar value of the annual sales – $lsales_{it}$, a logarithm of the number of employees – $lemploy_{it}$, and a logarithm of the average annual employee salary – $lavgsal_{it}$ – as additional explanatory variables. The summary statistics for these variables are reported in Table B.9. The first stage regression results with and without different year intercepts are provided in Table B.7. Regressions (3) and (4) employ additional explanatory variables available. In these regressions, control variables $lsales_{it}$ and $lemploy_{it}$ and their time averages are both individually and jointly statistically insignificant at any conventional levels. The logarithm of the average annual employee salary, $lavgsal_{it}$, and its time average are individually and jointly significant at least at 10% level of

significance. Based on these results we consider the average annual salary as the only valid additional explanatory variable. Thus, the next model of our interest is

$$\log(scrap_{it}) = \alpha + b_{1i}hrsemp_{it} + b_{2i}lavgsal_{it} + \delta_1 d88_t + \delta_2 d89_t + a_{1i} + u_{it}. \quad (2.29)$$

The estimation results for this model are provided in column (8) of Table B.8. Column (7) of the same table contains the results for the case without year dummies. The CF estimates of the APE of the annual hours of job training slightly decrease when we use two explanatory variables in comparison with the case with only one regressor besides year dummies [See columns (7) and (8) vs. columns (5) and (6) of Table B.8]. Once again, the estimates from the CF method (regressions (7) and (8)) are more statistically significant and substantially larger in magnitudes for the two regression specifications considered than the corresponding FE-IV estimates (regressions (3) and (4)).

Finally, we might think that the assumption of a heterogenous response to the annual average salary per worker stretches our imagination too much. Indeed, the same amount of the annual salary per employee is likely to have the same impact on the output scrap rate across different firms. Since the firms are located in the same area (Michigan), workers of these firms are expected to get comparable value out of the same monetary compensation for their work. Consequently, we consider the following model:

$$\log(scrap_{it}) = \alpha + b_{1i}hrsemp_{it} + \beta_2 lavgsal_{it} + \delta_1 d88_t + \delta_2 d89_t + a_{1i} + u_{it}. \quad (2.30)$$

The results for this model using the CF approach when the year dummies are both excluded and included are provided in columns (9) and (10) of Table B.8, respectively. The estimates from the CF method are bigger and more statistically significant than the corresponding FE-IV estimates reported in columns (3) and (4) of Table B.8.

Interestingly, Table B.8 shows that the estimates of the adjusted standard errors for the CF method are very close to the estimates of the robust standard errors that

are calculated ignoring the first stage estimation. Contrary, the simulation results [See Table B.5] indicate that the adjustment in the standard errors for the first stage estimation is not trivial.

Based on results reported in columns (5) through (10) of Table B.8 we can conclude that the CF approach estimates of the APE of the annual hours of job training per worker are robust to different model specifications. Further, they are larger in magnitudes and statistically more significant than the FE-IV estimates (regressions (1) through (4) of Table B.8) for all models considered.

2.6 Conclusion

This paper studies CRC balanced panel data models with endogenous regressors as in Murtazashvili and Wooldridge (2005). However, in addition to allowing some explanatory variables to be correlated with the idiosyncratic error, we also let the joint distribution of the endogenous regressors and the individual heterogeneity conditional on the instruments depend on the instruments. In particular, we allow the endogenous regressors to be roughly continuous. We use a control function approach, which introduces residuals from the reduced form for the endogenous regressors as covariates in the structural model. We propose a two-step method to account for heterogeneity and endogeneity and to consistently estimate APEs in CRC panel data models with endogenous (roughly) continuous regressors. Further, we relax the assumptions in Murtazashvili and Wooldridge (2005) by allowing the individual slopes in a CRC model to vary over time.

Monte Carlo simulations indicate that in the finite samples the control function approach to estimating the CRC balanced panel data model with time-invariant individual heterogeneity performs better than other estimators when the joint distribution of the individual heterogeneity and the endogenous regressors conditional on

the instruments depends on these instruments.

Finally, we apply the new method to the problem of estimating the APEs of the annual hours of on-job-training on the output scrap rates for manufacturing firms in Michigan extending the work of Holzer, Block, Cheatham, and Knott (1993) to allow for a firm-specific effect. The control function approach we propose delivers the APEs of the annual hours of job training on the output scrap rates that are larger in magnitudes and statistically more significant than the APEs' estimates from the FE-IV approach.

CHAPTER 3

ESTIMATION OF A DYNAMIC BINARY RESPONSE PANEL DATA MODEL WITH AN ENDOGENOUS REGRESSOR, WITH AN APPLICATION TO THE ANALYSIS OF POVERTY PERSISTENCE IN RURAL CHINA

3.1 Introduction

Dynamic binary response models have considerable appeal for a diverse range of policy analyses in which identifying or controlling for state dependence is important and one is interested in a binary outcome.[1] When the outcome is also affected by an endogenous treatment, then an additional complication arises in efforts to identify the effects of the treatment on the outcome and on state dependence. In this paper, we

[1]The range of research areas for which dynamic binary response models have proven important include: labor force participation (Heckman and Willis, 1977; Hyslop, 1999), the probability of receiving welfare (Bane and Ellwood, 1986), the experience social exclusion (Poggi, 2007), and the identification of adverse selection in insurance markets (Chiappori and Salanie, 2000).

propose a parametric approach to estimating binary response dynamic panel data models with endogenous contemporaneous regressors. Our method combines the approach to solving the unobserved heterogeneity and the initial conditions problems in non-linear dynamic models (Wooldridge, 2005c) with a control function approach to controlling for endogeneity of contemporaneous explanatory variables in cross-sectional non-linear models (e.g., Rivers and Vuong, 1988; Smith and Blundell, 1986).

Among other possible applications, the relevance and potential strength of our approach can be demonstrated in analyses of how migration in developing countries affects the poverty status of residents living in migrant source communities. In this setting, we are faced with two important sources of endogeneity: first, the migration decision of community residents may be driven by negative shocks that also raise the probability that households are poor. Second, we expect there to be correlation between migration decisions and the unobserved characteristics of individuals and communities, which may also affect poverty status. Our approach allows us to consistently estimate parameters of a dynamic binary response panel data model with unobserved heterogeneity when some of the continuous contemporaneous explanatory variables are endogenous. To account for the endogeneity in migration from home communities, we employ a control function approach in which residuals from the reduced form for the endogenous regressor are introduced as covariates in the structural model. To deal with the dynamic nature of the model, we consider two possibilities. We first use a "pure" random effects approach that allows the unobserved heterogeneity to be independent of the observed exogenous covariates and initial conditions. Next, we relax this strong assumption by employing the dynamic correlated random effects model introduced by Wooldridge (2005c). This approach is not only more relevant for analyses of poverty persistence, but also more flexible and computationally straightforward than alternative approaches currently in use.

We then implement our empirical approach using panel household and village

data from rural China. Following the market-oriented reforms introduced in the early 1980s, there was a pronounced decline in the proportion of China's population living below the poverty line (Ravallion and Chen, 2007). While much of the literature examining growth in China's rural areas has focused on incentive effects related to reform and on the role of local non-farm employment, there has been relatively little research demonstrating the relationship between reduction of barriers to migration from villages and the probability that households within the village have consumption levels below the poverty line. Our empirical analysis demonstrates an economically significant causal relationship between reduction of barriers to migration and poverty reduction in rural China.

The paper proceeds as follows. In the Section 3.2 below, we first review approaches to estimation of dynamic binary response panel data models, and then propose an approach to estimating these models when there is an endogenous regressor. In Section 3.3, we introduce the rural China setting, and develop a specific implementation of the empirical model and strategy for identifying the effect of migration on poverty within China's villages. In Section 3.4, we discuss our estimation results and the performance of the model, and then in Section 3.5 we summarize our results and discuss directions for future research.

3.2 Estimation of a Dynamic Binary Response Panel Data Model with an Endogenous Regressor

3.2.1 Dynamic Binary Response Panel Data Models

Dynamic binary response panel data models with unobserved heterogeneity have been used extensively in theoretical and empirical studies. Both parametric and semiparametric methods have been proposed to solve the initial conditions problem and to obtain consistent estimates of model parameters when all explanatory variables other than the lagged dependent variable are strictly exogenous.[2] Semiparametric methods allow estimation of parameters without specifying a distribution of the unobserved heterogeneity, but they are often overly restrictive with respect to the strictly exogenous covariates. Honoré and Kyriazidou (2000), for example, propose an approach that does not allow for discrete explanatory variables. More importantly, because the semiparametric methods do not specify the distribution of the unobserved heterogeneity, the absolute importance of any of the explanatory variables in a dynamic binary response panel data model cannot be determined. Models with no assumption on either the unobserved effects or the initial conditions, or their relationship to other covariates, are best described as *fixed effects* models, and the semiparametric approach of Honoré and Kyriazidou (2000) falls into this class of models.[3]

[2]With a structural binary outcome model that allows for unobserved effects, one must be concerned that bias could be introduced through a systematic relationship between an unobserved effect and the initial value of the dependent variable. This is known as the initial conditions problem.

[3]We follow Chay and Hyslop (2000) in classifying models requiring no assumption on unobservable effects or initial conditions as *fixed effect* models, and refer to *random effect* models as those in which one specifies a distribution of unobserved effects and initial conditions given exogenous explanatory variables.

Due to their computational simplicity, parametric methods have received greater attention than semiparametric methods. There are four main parametric approaches that have been employed for estimation of the dynamic nonlinear panel data models with strictly exogenous covariates other than the lagged dependent variable. All four approaches use conditional maximum likelihood estimation (CMLE) analysis. The first approach treats the initial conditions for each cross-sectional unit - y_{i0} - as nonrandom variables. If, in addition, unobserved effects, c_i, are also assumed to be independent of $\mathbf{z}_i$, one obtains the density of $(y_{i1}, y_{i2}, ..., y_{iT})$ given the initial conditions, y_{i0}, and the exogenous explanatory variables, $\mathbf{z}_i = (\mathbf{z}_{i1}, \mathbf{z}_{i2}, ..., \mathbf{z}_{iT})$, by integrating out the c_i. We refer to the relationship between the observed exogenous covariates and the unobserved heterogeneity in the first method a *"pure" random effects* relationship because we assume c_i to be independent of $\mathbf{z}_i$ and y_{i0}. While this method does provide a way to obtain consistent estimates of the model parameters, nonrandomness of the initial conditions requires the very strong assumption of independence between the initial conditions and the unobserved effects.

A second parametric approach would involve treating the initial conditions as random and specify a density of y_{i0} given $(\mathbf{z}_i, c_i)$. With this density, one can then obtain the joint distribution of all the outcomes, $(y_{i0}, y_{i1}, y_{i2}, ..., y_{iT})$, conditional on unobserved heterogeneity, c_i, and strictly exogenous observables, $\mathbf{z}_i$. One important drawback of this approach lies with the difficulty of specifying the density of y_{i0} given $(\mathbf{z}_i, c_i)$.[4]

A third method, proposed by Heckman (1981), suggests approximating a density of the initial conditions, y_{i0}, given $(\mathbf{z}_i, c_i)$ and specifying a density of the unobserved effects given the strictly exogenous explanatory variables. The density of $(y_{i0}, y_{i1}, y_{i2}, ..., y_{iT})$ given $\mathbf{z}_i$ can then be obtained. While Heckman's approach avoids the drawback of the second method, it is computationally challenging. Since both

[4]More details on this approach and potential drawbacks can be found in Wooldridge (2002), page 494.

the second and the third methods explicitly specify a distribution of the unobserved heterogeneity conditional on strictly exogenous observables and a distribution of the initial conditions conditional on the unobserved effects and the exogenous covariates, they can be classified as *random effects* models.

Finally, an approach proposed by Wooldridge (2005c) recommends obtaining a joint distribution of $(y_{i1}, y_{i2}, ..., y_{iT})$ conditional on $(y_{i0}, \mathbf{z}_i)$ rather than a distribution of $(y_{i0}, y_{i1}, y_{i2}, ..., y_{iT})$ conditional on $\mathbf{z}_i$ as in Hechman's approach. For this method to work, we need to specify a density of c_i given $(y_{i0}, \mathbf{z}_i)$.[5] This fourth approach is more flexible and requires fewer computational resources than Heckman's technique. In this method, we call the relationship between the observed exogenous covariates and the unobserved heterogeneity a *"correlated" random effects* relationship because we allow c_i to be a linear function of $\mathbf{z}_i$ and y_{i0}.

In the next section we develop a theoretical method that consistently estimates parameters of a dynamic binary response panel data model when the contemporaneous explanatory variables are not strictly exogenous. To do so, we employ a *control function* approach, originally introduced by Smith and Blundell (1986) and Rivers and Vuong (1988). The main idea of the control function approach is to add control variables into the structural model to control for endogeneity. Since we will consider a model with two possible sources of endogeneity – the correlation between the unobserved heterogeneity and a regressor, and the correlation between a regressor and the structural error, we model the relationships among the unobserved effect, exogenous covariates, and the error from the reduced form equation for the endogenous explanatory variable.

[5]The specification of this density in Wooldridge's method is motivated by Chamberlain's (1980) device.

3.2.2 A General Approach to Estimation

Our specification of the binary response model assumes that for a random draw i from the population, there is an underlying latent variable model:

$$y_{1it}^{*} = \mathbf{z}_{1it}\beta_1 + \beta_2 y_{2it} + \rho y_{1i,t-1} + c_{1i} + u_{1it}, \tag{3.1}$$

$$y_{1it} = 1[y_{1it}^{*} \geq 0], t = 1, ..., T. \tag{3.2}$$

where $\mathbf{z}_{1it}$ is a $1 \times (K-1)$ vector of strictly exogenous covariates, which may contain a constant term, y_{2it} is an endogenous covariate, c_{1i} is an unobserved effect, and u_{1it} is an idiosyncratic serially uncorrelated error such that $\text{Var}(u_{1it}) = 1$. $1[\cdot]$ is an indicator function. We assume a sample of size N randomly drawn from the population, and that T, the number of time periods, is fixed in the asymptotic analysis. For simplicity, we assume a balanced panel.

Let β denote $(\beta_1', \beta_2, \beta_3, \rho)\prime$, which is a $1 \times (K+1)$ vector of all the parameters. Importantly, this model allows the probability of success at time t to depend not only on unobserved heterogeneity, c_{1i}, but also on the outcome in $t-1$. A key assumption is that the dynamics are correctly specified and dynamic completeness of the model implies that the error term is serially uncorrelated. Thus, assuming that model (3.1) is correctly specified dynamically, we assume that the error term u_{1it} is serially uncorrelated. Allowing, u_{1it} to have arbitrary serial correlation, would suggest including more lags of the dependent variable (3.1). For example, in the simplest case of a linear model, when an error term, u_{it}, follows AR(1) process, a simple calculation shows that a dependent variable, say y_{it}, actually depends on not only $y_{i,t-1}$ but also $y_{i,t-2}$. Similarly, in the context of our model, one should have a good reason to expect a serially correlated error term u_{1it} and yet to include only one lag of y_{1it}.

Further, we make additional assumptions on strict exogeneity of the contemporaneous explanatory variables. First, some of the contemporaneous covariates, $\mathbf{z}_{1it}$, are assumed to be strictly exogenous (conditional on c_{1i}). Second, we allow some of the explanatory variables, here represented by the scalar y_{2it}, to be endogenous.

$$\begin{aligned} y_{2it} &= \mathbf{z}_{1it}\delta_1 + \mathbf{z}_{2it}\delta_2 + c_{2i} + u_{2it} \\ &= \mathbf{z}_{it}\delta + \overline{\mathbf{z}}_i\lambda + a_{2i} + u_{2it} \\ &= \mathbf{z}_{it}\delta + \overline{\mathbf{z}}_i\lambda + v_{2it}, \end{aligned} \tag{3.3}$$

where $t = 1, ..., T$, c_{2i} is an unobserved effect, and u_{2it} is an idiosyncratic serially uncorrelated error with $\text{Var}(u_{2it}) = \sigma_2^2$. Let $\mathbf{z}_{it} = (\mathbf{z}_{1it}, \mathbf{z}_{2it})$ be a $1 \times L$ vector of instrumental variables, with $L \geq K$, i.e., we assume the vector $\mathbf{z}_{2it}$ contains at least one element. We employ the Mundlak-Chamberlain device for the unobserved effect, c_{2i}, and this is reflected in line two of equation (3.3). We replace c_{2i} with its projection onto the time averages of all the exogenous variables: $c_{2i} = \overline{\mathbf{z}}_i\lambda + a_{2i}$. Then, the new composite error term is $v_{2it} = a_{2i} + u_{2it}$. Further, $\overline{\mathbf{z}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathbf{z}_{it}$, and $\delta = (\delta_1', \delta_2')\prime$. We follow Rivers and Vuong (1988) and refer to (3.3) as a reduced form equation.

Next, consider the relationship between u_{1it} and u_{2it}. We assume that (u_{1it}, u_{2it}) has a zero mean, bivariate normal distribution and is independent of $\mathbf{z}_i = (\mathbf{z}_{1i}, \mathbf{z}_{2i}) = (\mathbf{z}_{i1}, \mathbf{z}_{i2}, ..., \mathbf{z}_{iT})$. Note that under joint normality of (u_{1it}, u_{2it}), with $\text{Var}(u_{1it}) = 1$, we can write

$$\begin{aligned} u_{1it} &= \theta u_{2it} + e_{1it} \\ &= \theta(v_{2it} - a_{2i}) + e_{1it}, \end{aligned} \tag{3.4}$$

where $\theta = \eta/\sigma_2^2$, $\eta = \text{Cov}(u_{1it}, u_{2it})$, $\sigma_2^2 = \text{Var}(u_{2it})$, and e_{1it} is a serially uncorrelated random term, which is independent of $\mathbf{z}_i$ and u_{2it}. The absence of serial correlation of the e_{1it} follows from the fact that u_{1it} and u_{2it} are both assumed not to suffer from

serial correlation. If there were no lagged dependent variables on the right hand side of equation (3.1), there would be little need to worry about possible serial correlation in the error term u_{2it} of equation (3.3), as long as we assume that u_{1it} is also serially uncorrelated. However, we are interested in a dynamic model, and the assumption of no serial correlation in u_{2it} is crucial for equation (3.4). Since equation (3.3) is essentially a reduced form equation for the endogenous variable y_{2it}, the assumption of no serial correlation in u_{2it} (and in e_{1it}, as a result) is appropriate in the context of our model.

Equation (3.4) is essentially an assumption regarding the contemporaneous endogeneity of y_{2it}. It suggests that the contemporaneous v_{2it} is sufficient for explaining the relation between u_{1it} and v_{2it}. In other words, once we somehow account for endogeneity of y_{2it} in period t, we might think that y_{2it} becomes "completely" exogenous, and we can estimate the parameters of interest using standard methods valid for exogenous explanatory variables. However, there is the possibility of an additional feedback from the endogenous variable y_2 in different time periods to the main dependent variable of interest, y_1, at time t. This possibility arises because we let the reduced form equation for the endogenous variable, y_{2it}, contain a time-constant unobserved effect, a_{2i}.

From assumption (3.4), $e_{1it} \sim \text{Normal}(0, \sigma^2_{e_1})$, where $\sigma^2_{e_1} = 1 - \xi^2$, since $\text{Var}(u_{1it}) = 1$, and $\xi = \text{Corr}(u_{1it}, u_{2it})$, we can write

$$\begin{aligned} y_{1it} &= 1[\mathbf{x}_{1it}\beta + c_{1i} + \theta(v_{2it} - a_{2i}) + e_{1it} \geq 0] \\ &= 1[\mathbf{x}_{1it}\beta + \theta v_{2it} + (c_{1i} - \theta a_{2i}) + e_{1it} \geq 0 \\ &= 1[\mathbf{x}_{1it}\beta + \theta v_{2it} + c_{0i} + e_{1it} \geq 0], \end{aligned} \tag{3.5}$$

where $t = 1, ..., T$, $\mathbf{x}_{1it} = (\mathbf{z}_{1it}, y_{2it}, y_{1i,t-1})$, $\beta = (\beta_1', \beta_2, \rho)\prime$, and $c_{0i} = c_{1i} - \theta a_{2i}$ is a composite unobserved effect. Since the unobserved effect c_{0i} is present in equation (3.5), we should consider the relation between the unobserved effect c_{0i} and the

explanatory variables in equation (3.5). Importantly, the composite unobserved effect c_{0i} is a function of v_{2it}, where $t = 1, ..., T$, by construction:

$$c_{0i} = c_{1i} - \theta a_{2i} = c_{1i} - \theta(v_{2it} - u_{2it}), t = 1, ..., T.$$

Thus, in order to obtain consistent estimates of the parameters from equation (3.5), we must take into account the relation between c_{0i} and v_{2it} in different time periods.

First, we use a "pure" random effects approach, i.e., we assume that

$$c_{0i}|\mathbf{z}_i, y_{1i0}, \mathbf{v}_{2i} \sim \text{Normal}(\alpha_0 \overline{v}_{2i}, \sigma^2_{a_1}), t = 1, ..., T, \tag{3.6}$$

which can be written as $c_{0i} = \alpha_0 \overline{v}_{2i} + a_{1i}$, $t = 1, ..., T$, where $a_{1i}|\mathbf{z}_i, y_{1i0}, \mathbf{v}_{2i} \sim$ Normal$(0, \sigma^2_{a_1})$ and is independent of $(\mathbf{z}_i, y_{1i0}, \mathbf{v}_{2i})$, where $\overline{v}_{2i} = \frac{1}{T}\sum_{t=1}^{T} v_{2it}$, and $\mathbf{v}_{2i} = (v_{i1}, v_{i2}, ..., v_{iT})$. While a limiting assumption in many potential applications, the "pure" RE assumption (3.6) may be relevant for certain cases. In particular, when every individual in the initial time period is in the same state (e. g., we are interested in the population of people who smoke), assumption (3.6) might be appropriate. Further, since we assume that the composite unobserved effect, c_{0i}, is independent of the initial condition, y_{1i0}, it is natural to think that v_{2it}'s in different time periods have equal impacts on c_{0i}. Consequently, we employ $\overline{v}_{2i}$ as a sufficient statistic for describing the relation between c_{0i} and v_{2it}'s in different time periods.

Then, under assumptions (3.1)-(3.4) and (3.6), we can rewrite equation (3.5) as

$$y_{1it} = 1[\mathbf{x}_{1it}\beta + \theta v_{2it} + \alpha_0 \overline{v}_{2i} + a_{1i} + e_{1it} \geq 0]. \tag{3.7}$$

Clearly, the estimates of $\beta = \frac{\beta}{\sqrt{\sigma^2_{e_1}+\sigma^2_{a_1}}}$, $\theta = \frac{\theta}{\sqrt{\sigma^2_{e_1}+\sigma^2_{a_1}}}$, and $\alpha_0 = \frac{\alpha_0}{\sqrt{\sigma^2_{e_1}+\sigma^2_{a_1}}}$ can be obtained using standard random effects probit software by including $\overline{\hat{v}}_{2i}$ in each time period into the list of the explanatory variables along with $\mathbf{x}_{1it}$ and $\hat{v}_{2it}$, where $\overline{\hat{v}}_{2i} = \frac{1}{T}\sum_{t=1}^{T} \hat{v}_{2it}$.

However, as we discussed earlier, the assumption of independence between the unobserved effect and the initial conditions and the exogenous covariates is often too

restrictive. In particular, the "pure" random effects assumption is unrealistic in the context of the application to poverty persistence that we will examine below. For instance, unobserved dimensions of ability are very likely to be related to poverty status not only in the initial period, but also in future periods.

Rather than using a "pure" random effects approach, we prefer building on the dynamic "correlated" random effects model introduced by Wooldridge (2005c). Instead of the conditional distribution of c_{0i} assumed in (3.6), we now assume that

$$c_{0i}|\mathbf{z}_i, y_{1i0}, \mathbf{v}_{2i} \sim \text{Normal}(\mathbf{v}_{2i}\alpha_0 + \mathbf{z}_i\alpha_1 + \alpha_2 y_{1i0}, \sigma^2_{a_1}), \tag{3.8}$$

which follows from writing $c_{0i} = \mathbf{v}_{2i}\alpha_0 + \mathbf{z}_i\alpha_1 + \alpha_2 y_{1i0} + a_{1i}$, where $a_{1i}|\mathbf{z}_i, y_{1i0}, \mathbf{v}_{2i} \sim \text{Normal}(0, \sigma^2_{a_1})$ and independent of $(\mathbf{z}_i, y_{1i0}, \mathbf{v}_{2i})$. Since we allow for a nonzero correlation between the composite unobserved effect, c_{0i}, and the initial condition, y_{1i0}, v_{2it}'s in different time periods might have different effects on c_{0i}. Thus, we let v_{2it}'s from different time periods have unequal "weights" for explaining c_{0i}. Assumption (3.8) is an extension of Chamberlain's assumption for a static probit model to the dynamic setting. To allow for correlation between c_{0i} and $\mathbf{z}_i$ and y_{1i0}, we assume a conditional normal distribution with linear expectation and constant variance. Assumption (3.8) is a restrictive assumption since it specifies a distribution for c_{0i} given $\mathbf{z}_i, y_{1i0}, \mathbf{v}_{2i}$. However, it is an improvement on the "pure" random effects approach in that it allows for some dependence between the unobserved effect and the vector of all explanatory variables across all time periods.

Then, under assumptions (3.1)-(3.4) and (3.8), we can rewrite equation (3.5) as

$$\begin{aligned} y_{1it} &= 1[\mathbf{x}_{1it}\beta + \theta v_{2it} + c_{0i} + e_{1it} \geq 0] \\ &= 1[\mathbf{x}_{1it}\beta + \theta v_{2it} + \mathbf{v}_{2i}\alpha_0 + \mathbf{z}_i\alpha_1 + \alpha_2 y_{1i0} + a_{1i} + e_{1it} \geq 0]. \end{aligned} \tag{3.9}$$

Equation (3.9) suggests that we can estimate $\beta = \frac{\beta}{\sqrt{\sigma^2_{e_1}+\sigma^2_{a_1}}}$ and $\theta = \frac{\theta}{\sqrt{\sigma^2_{e_1}+\sigma^2_{a_1}}}$ along with $\alpha_0 = \frac{\alpha_0}{\sqrt{\sigma^2_{e_1}+\sigma^2_{a_1}}}$, $\alpha_1 = \frac{\alpha_1}{\sqrt{\sigma^2_{e_1}+\sigma^2_{a_1}}}$ and $\alpha_2 = \frac{\alpha_2}{\sqrt{\sigma^2_{e_1}+\sigma^2_{a_1}}}$ using standard random

effects probit software by including $\hat{\mathbf{v}}_{2i}$, $\mathbf{z}_i$, and y_{1i0} in each time period into the list of the explanatory variables along with $\mathbf{x}_{1it}$ and $\hat{v}_{2it}$.

3.2.3 Allowing for Serial Correlation of Errors in the First Stage

If the first stage error, u_{2it}, is serially correlated, we must modify our two-step estimating procedure. To be specific, assume u_{2it} follows an AR(1) process: $u_{2it} = \pi u_{2i,t-1} + e_{2it}$, where e_{2it} is a white noise error with $\text{Var}(e_{2it}) = \sigma^2_{e_2}$. Then, under assumption (3.4),

$$\begin{aligned}\text{Cov}(e_{1it}, e_{1it-1}) &= \text{Cov}(u_{1it} - \theta u_{2it}, u_{1i,t-1} - \theta u_{2i,t-1}) \\ &= \text{Cov}(u_{1it} - \pi\theta u_{2it} - \theta e_{2it}, u_{1i,t-1} - \theta u_{2i,t-1}) = \pi\theta^2 \text{E}(u^2_{2i,t-1}),\end{aligned}$$

which is more than 0, unless either $\pi = 0$ or $\theta = 0$. Clearly, assumption (3.4) is no longer appropriate and needs to be corrected.

Define the variance-covariance matrix of $\mathbf{v}_{2i}$ as $\mathbf{\Omega} \equiv \text{E}(\mathbf{v}'_{2i}\mathbf{v}_{2i})$, a $T \times T$ matrix that we assume to be positive definite. Then,

$$\mathbf{\Omega} \equiv \text{E}(\mathbf{v}_{2i}\mathbf{v}'_{2i}) = \sigma^2_{a_2}\mathbf{j}_T\mathbf{j}'_T + \sigma^2_2 \begin{pmatrix} 1 & \pi & \pi^2 & \cdots & \pi^{T-2} & \pi^{T-1} \\ \pi & 1 & \pi & \cdots & \pi^{T-3} & \pi^{T-2} \\ \pi^2 & \pi & 1 & \cdots & \pi^{T-4} & \pi^{T-3} \\ \vdots & & & \ddots & & \vdots \\ \pi^{T-2} & \pi^{T-3} & \pi^{T-4} & \cdots & 1 & \pi \\ \pi^{T-1} & \pi^{T-2} & \pi^{T-3} & \cdots & \pi & 1 \end{pmatrix}, \tag{3.10}$$

where $\mathbf{j}_T$ is a $T \times 1$ vector of ones, and $\sigma^2_2 = \frac{\sigma^2_{e_2}}{1-\pi^2}$. We can obtain consistent estimates of the parameters in (3.10), and use them to transform v_{2it} to v^*_{2it}, which is a first stage error free of serial correlation. One useful method for estimating π, $\sigma^2_{a_2}$, $\sigma^2_{e_2}$, and σ^2_2 is the minimum distance estimator, described in detail by Chamberlain (1984). Cappellari (1999) has developed code that conveniently implements this method in Stata.

Once we have first stage errors free of serial correlation, we can use the transformation $u^*_{2it} = v^*_{it} - a_{2i}$ to adjust assumption (3.4). We can then assume that under joint normality of (u_{1it}, u^*_{2it}),

$$\begin{aligned} u_{1it} &= u^*_{2it}\theta + e_{1it} \\ &= \theta(v^*_{2it} - a_{2i}) + e_{1it}, \end{aligned} \tag{3.11}$$

where e_{1it} is a serially uncorrelated random term, which is independent of $\mathbf{z}_i$ and u^*_{2it}. Inclusion of u^*_{2it} instead of u_{2it} in equation (3.11) guarantees that e_{1it} will not be serially correlated. Then, we can write

$$\begin{aligned} y_{1it} &= 1[\mathbf{x}_{1it}\beta + c_{1i} + \theta v^*_{2it} - \theta a_{2i} + e_{1it} \geq 0] \\ &= 1[\mathbf{x}_{1it}\beta + \theta v^*_{2it} + (c_{1i} - \theta a_{2i}) + e_{1it} \geq 0 \\ &= 1[\mathbf{x}_{1it}\beta + \theta v^*_{2it} + c_{0i} + e_{1it} \geq 0], \end{aligned} \tag{3.12}$$

where $t = 1, ..., T$, and $c_{0i} = c_{1i} - \theta a_{2i}$ is a composite unobserved effect.

Based on equation (3.12), it is straightforward to adjust the two-step estimating procedure discussed in Section 3.2.2 to account for the presence of the serial correlation in u_{2it}. For example, under "correlated" random effects assumption (3.8), equation (3.12) can be written as

$$\begin{aligned} y_{1it} &= 1[\mathbf{x}_{1it}\beta + \theta v^*_{2it} + c_{0i} + e_{1it} \geq 0] \\ &= 1[\mathbf{x}_{1it}\beta + \theta v^*_{2it} + \mathbf{v}_{2i}\alpha_0 + \mathbf{z}_i\alpha_1 + \alpha_2 y_{1i0} + a_{1i} + e_{1it} \geq 0]. \end{aligned} \tag{3.13}$$

Then, we can estimate the parameters β, θ, α_1, and α_2 using standard random effects probit software by including $\hat{\mathbf{v}}_{2i}$, $\mathbf{z}_i$, and y_{1i0} in each time period into the list of the explanatory variables along with $\mathbf{x}_{1it}$.

3.2.4 Calculation of Average Partial Effects

To assess the magnitude of state dependence we must calculate the average partial effect (APE) of the lagged poverty status on its current value. We follow an approach

favored by Wooldridge (2002) to calculate the APEs after our two-step estimation procedure. The APEs can be calculated by taking either differences or derivatives of

$$\mathrm{E}[\Phi(\mathbf{x}_{1t}\beta + \theta v_{2it} + \mathbf{v}_{2i}\alpha_0 + \mathbf{z}_i\alpha_1 + \alpha_2 y_{1i0})], \tag{3.14}$$

where $t = 1, ..., T$ and in the argument of the expectations operator, variables with a subscript i are random and all others are fixed.

In order to obtain estimates of the parameter values in (3.14), we appeal to a standard uniform weak law of large numbers argument.[6] For any given value of $\mathbf{x}_{1t}(\mathbf{x}_1^0)$, a consistent estimator for expression (3.14) can be obtained by replacing unknown parameters by consistent estimators:

$$N^{-1}\sum_{i=1}^{N}\Phi(\mathbf{x}_1^0\hat{\beta}_* + \hat{\theta}_*\hat{v}_{2it} + \hat{\mathbf{v}}_{2i}\hat{\alpha}_{0*} + \mathbf{z}_i\hat{\alpha}_{1*} + \hat{\alpha}_{2*}y_{1i0}), \tag{3.15}$$

where $t = 1, ..., T$, the $\hat{v}_{2it}$ are the first-stage pooled OLS residuals from regressing y_{2it} on $\mathbf{z}_{it}$, $\hat{\mathbf{v}}_{2i} = (\hat{v}_{i1}, \hat{v}_{i2}, ..., \hat{v}_{iT})$, the $*$ subscript denotes multiplication by $\hat{\sigma}^2 = (\widehat{\sigma_{e_1}^2 + \sigma_{a_1}^2})^{-1/2}$, and $\hat{\beta}$, $\hat{\theta}$, $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\sigma}^2$ are the conditional MLEs. Note that $\hat{\sigma}^2$ is the usual error variance estimator from the second-stage random effects probit regression of y_{1it} on $\mathbf{x}_{1it}$, $\hat{v}_{2it}$, $\mathbf{z}_i$, and y_{1i0}. One may then employ either a mean value expansion or a bootstrapping approach to obtain asymptotic standard errors. We can compute either changes or derivatives of equation (3.15) with respect to $\mathbf{x}_{1t}$ to obtain the APEs of interest.

In common with the adjustment to our estimating procedure, one must also correct the estimated APEs when errors are serially correlated. We obtain the APEs by taking either differences or derivatives of

$$\mathrm{E}[\Phi(\mathbf{x}_{1t}\beta + \theta v_{2it}^* + \mathbf{v}_{2i}\alpha_0 + \mathbf{z}_i\alpha_1 + \alpha_2 y_{1i0})], \tag{3.16}$$

where $t = 1, ..., T$. For simplicity, consider the second approach used in Section 3.2.4 to obtain the APEs' estimates. For any given value of $\mathbf{x}_{1t}(\mathbf{x}_1^0)$, a consistent estima-

[6]See Wooldridge (2002) for details.

tor of expression (3.16) is obtained by replacing unknown parameters by consistent estimators:

$$N^{-1}\sum_{i=1}^{N}\Phi(\mathbf{x}_1^0\hat{\beta}_* + \hat{\theta}_*\hat{v}_{2it}^* + \hat{\mathbf{v}}_{2i}\hat{\alpha}_{0*} + \mathbf{z}_i\hat{\alpha}_{1*} + \hat{\alpha}_{2*}y_{1i0}), \tag{3.17}$$

where $t = 1, ..., T$, $\hat{v}_{2it}^*$ is a first stage residual cleaned of serial correlation, where the $*$ subscript denotes multiplication by $\hat{\sigma}^2 = (\widehat{\sigma_{e_1}^2 + \sigma_{a_1}^2})^{-1/2}$, and $\hat{\beta}$, $\hat{\theta}$, $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\sigma}^2$ are the conditional MLEs. We can then compute either changes or derivatives of equation (3.17) with respect to $\mathbf{x}_{1t}$ to obtain the APEs of interest.

3.3 Migrant Labor Markets and Poverty Persistence in Rural China

Before applying the dynamic binary response model discussed above to an analysis of how migrant labor market affect poverty status in rural China, we first briefly review the history of rural-urban migration in China and review other evidence on the impact of migration in the home villages of migrants. Next, we propose a specific implementation of the dynamic binary response model to an analysis of the impact of migration on the probability that a rural household is poor. We then introduce the unique panel household and village data sources used in our analysis and describe our approach to identifying the migrant networks that affect the cost of finding migrant employment for village residents.

3.3.1 Rural-Urban Migration in China

China's labor market experienced a dramatic change during the 1990s, as the volume of rural migrants moving to urban areas for employment grew rapidly. Estimates using the one percent sample from the 1990 and 2000 rounds of the Population Census

and the 1995 one percent population survey suggest that the inter-county migrant population grew from just over 20 million in 1990 to 45 million in 1995 and 79 million by 2000 (Liang and Ma, 2004). Surveys conducted by the National Bureau of Statistics (NBS) and the Ministry of Agriculture include more detailed retrospective information on past short-term migration, and suggest even higher levels of labor migration than those reported in the census (Cai, Park and Zhao, 2007).

Before labor mobility restrictions were relaxed, households in remote regions of rural China faced low returns to local economic activity, reinforcing geographic poverty traps (Jalan and Ravallion, 2002). A considerable body of descriptive evidence related to the growth of migration in China raises the possibility that migrant opportunity may be an important mechanism for poverty reduction. Studies of the impact of migration on migrant households suggest that migration is associated with higher incomes (Taylor, Rozelle and de Brauw, 2003; Du, Park, and Wang, 2006), facilitates risk-coping and risk-management (Giles, 2006; Giles and Yoo, 2006), and is associated with higher levels of local investment in productive activities (Zhao, 2003).

Institutional changes, policy signals and the high return to labor in urban areas each played a role in the expansion of migration during the 1990s. An early reform of the household registration (*hukou*) system in 1988 first established a mechanism for rural migrants to obtain legal temporary residence in China's urban areas (Mallee, 1995). In order to take advantage of this policy change, rural residents required a national identity card to obtain a legal temporary worker card (*zanzu zheng*), but not all rural counties had distributed IDs as of 1988.[7] As China recovered from its post-Tiananmen retrenchment, some credit a series of policy speeches made by Deng Xiaoping in 1992 as signals of renewed openness toward the marketization of the economy, including employment of migrant rural labor in urban areas (Chan and

[7]Legal temporary residence status does not confer access to the same set of benefits (e.g., subsidized education, health care, and housing) typically associated with permanent registration as a city resident.

Zhang, 1999). Combined with economic expansion, these institutional and policy changes led to increased demand for construction and service sector workers, and catalyzed the growth in rural-urban migration that continued throughout the 1990s.

The use of migrant networks and employment referral in urban areas are important dimensions of China's rural-urban migration experience. Rozelle et al (1999) emphasize that villages with more migrants in 1988 experienced more rapid migration growth by 1995. Zhao (2003) shows that number of early migrants from a village is correlated with the probability that an individual with no prior migration experience will choose to participate in the migrant labor market. Meng (2000) further suggests that variation in the size of migrant flows to different destinations can be partially explained by the size of the existing migrant population in potential destinations.[8]

3.3.2 The RCRE Household Survey

The primary data sources used for our analyses are the village and household surveys conducted by the Research Center for Rural Economy at China's Ministry of Agriculture from 1986 through the 2003 survey year. We use data from 90 villages in eight provinces (Anhui, Jilin, Jiangsu, Henan, Hunan, Shanxi, Sichuan and Zhejiang) that were surveyed over the 17-year period, with an average of 6305 households surveyed per year. Depending on village size, between 40 and 120 households were randomly surveyed in each village.

The RCRE household survey collected detailed household-level information on

[8]Referral through one's social network is a common method of job search in both the developing and developed world. Carrington, Detragiache, and Vishnawath (1996) explicitly show that in a model of migration, moving costs can decline with the number of migrants over time, even if wage differentials narrow between source communities and destinations. Survey-based evidence suggests that roughly 50 percent of new jobs in the US are found through referrals facilitated by social networks (Montgomery, 1991). In a study of Mexican migrants in the US, Munshi (2003) shows that having more migrants from one's own village living in the same city increases the likelihood of employment.

incomes and expenditures, education, labor supply, asset ownership, land holdings, savings, formal and informal access to credit, and remittances.[9] In common with the National Bureau of Statistics (NBS) *Rural Household Survey,* respondent households keep daily diaries of income and expenditure, and a resident administrator living in the county seat visits with households once a month to collect information from the diaries.

Our measure of consumption includes nondurable goods expenditure plus an imputed flow of services from household durable goods and housing. In order to convert the stock of durables into a flow of consumption services, we assume that current and past investments in housing are "consumed" over a 20-year period and that investments in durable goods are consumed over a period of 7 years.[10] We also annually "inflate" the value of the stock of durables to reflect the increase in durable goods' prices over the period. Finally, we deflate all income and expenditure data to 1986 prices using the NBS rural consumer price index for each province.

There has been some debate over the representativeness of both the RCRE and NBS surveys, and concern over differences between trends in poverty and inequality in the NBS and RCRE surveys. These issues are reviewed extensively in Appendix B of Benjamin et al (2005), but it is worth summarizing some of their findings here. First, when comparing cross sections of the NBS and RCRE surveys with overlapping years from cross sectional surveys not using a diary method, it is apparent that some high and low income households are under-represented.[11]Poorer illiterate households

[9]One shortcoming of the survey is the lack of individual-level information. However, we know the numbers of working-age adults and dependents, as well as the gender composition of household members.

[10] Our approach to valuing consumption follows the suggestions of Chen and Ravallion (1996) for the NBS Rural Household Survey, and is explained in more detail in Appendix A of Benjamin et al. (2005).

[11] The cross-sections used were the rural samples of the 1993, 1997 and 2000 China Health and Nutrition Survey (CHNS) and a survey conducted in 2000 by the Center for Chinese Agricultural Policy (CCAP) with Scott Rozelle (UC Davis) and Loren Brandt (University of Toronto).

are likely to be under-represented because enumerators find it difficult to implement and monitor the diary-based survey, and refusal rates are likely to be high among affluent households who find the diary reporting method a costly use of their time. Second, much of the difference between levels and trends from the NBS and RCRE surveys can be explained by differences in the valuation of home-produced grain and treatment of taxes and fees.

3.3.3 Migration, Consumption Growth and Poverty

Trends

One of the benefits of the accompanying village survey is a question asked each year of village leaders about the number of registered village residents working and living outside the village. In our analysis, we consider all registered residents working outside their home county to be migrants.[12]Both the tremendous increase in migration from 1987 onward and heterogeneity across villages are evident in Figure C.1. In 1987 an average of 3 percent of working age laborers in RCRE villages were working outside of their home villages, which rose steadily to 23 percent by 2003. Moreover, we observe considerable variability in the share of working age laborers working as migrants. Whereas some villages still had a small share of legal village residents employed as migrants, more than 50 percent of working age adults from other villages were employed outside the village by 2003.

The relationship between migration and consumption is of central concern for our analysis. The linear fit of the relationship between annual changes in migration and average village consumption growth in the RCRE data suggest a positive relationship (Figure C.2). The lowess fit, however, suggests the presence of nonlinearities,

[12] From follow up interviews with village leaders, it is apparent that registered residents living outside the county are unlikely to be commuters and generally live and work outside the village for more than six months of the year.

particularly around zero. Indeed the prospect that out-migration may be driven by negative shocks which also depress consumption should raise concern that size of the migrant network and consumption may be endogeneous and driven in part by shocks affecting both variables.

Even if consumption grows with an increase in the number of residents earning incomes from migrant employment, it is of important policy interest to understand which residents within villages are experiencing increases in consumption. Changes in the village poverty headcount are negatively associated with the change in the number of out-migrants, suggesting that poverty declines with increased out-migration (Figure C.3). Nonlinearities in the bivariate relationship are evident again in the non-parametric lowess plot of the relationship. Whether obvious non-linearities are related to the simultaneity of shocks and increases in out-migration and poverty for some villages or the simple fact that we have not controlled for other characteristics of villages, establishing a relationship between migration and increased consumption of poorer households within villages requires an analytical framework where we eliminate bias due to simultaneity and potential sources of unobserved heterogeneity.

A Causal Relationship Between Migration and Consumption Growth

In other research using this data source, de Brauw and Giles (2007) use linear dynamic panel data methods with continuous regressors to demonstrate a robust relationship between the reduction of obstacles to rural-urban migration and household consumption growth. While one might often suspect that the non-poor, who have sufficiently high human capital and other dimensions of ability, may benefit most from reductions in barriers to migration, general equilibrium effects of out-migration may lead to greater specialization of households in villages that has benefits for the poor. In particular, de Brauw and Giles demonstrate that households at the lower end of the consumption distribution tend to expand both their investments in agriculture related

assets and the area of land that they cultivate increase more with out migration than they do for richer households. This raises the prospect that ability to migrate may be causally related to poverty reduction within rural communities as well.

In the empirical application of our discrete binary response model below, we are simply seeking to understand whether out-migration from villages is associated with reductions in the probability that household consumption falls below the poverty line in rural China. We are agnostic as to whether poverty is reduced through direct participation in the migrant labor market, or through indirect general equilibrium effects that raise the return to labor in agricultural and other local activities.

3.3.4 Estimating the Impact of Migrant Labor Markets on Poverty Persistence

The econometric approach derived in Section 3.2.2 allows us to control for household specific unobserved effects, which will include fixed effects associated with the village in which households are located. We are interested in estimating the dynamic binary choice model for the probability that a household i falls below the poverty line at time t:

$$pov_{it} = 1[\beta_1 pov_{it-1} + \beta_2(M^i_{jt} * pov_{it-1}) + \beta_3 M^i_{jt} + \mathbf{X}'_{it}\alpha_1 + \alpha_2 lpc_{it} + \mathbf{D}_t + u_i + \varepsilon_{it}], \tag{3.18}$$

where pov_{it} is a binary indicator for whether the household is in poverty in year t, which will be affected by poverty status in the prior period, pov_{it-1}, the size of the migrant network from village j through which the household i may be able to obtain a job referral, M^i_{jt}, a vector of household demographic and human capital characteristics, $\mathbf{X}_{it}$, household land per capita, lpc_{it}, and year dummies to control for macroeconomic shocks, $\mathbf{D}_t$. We will be concerned about the possibility that an unobserved household effect, u_i, may be systematically related to the size of the

household's migrant network, to other covariates, and to household poverty status, thus introduce endogeneity concerns. The error term, ε_{it}, may be serially correlated, and we may be concerned that shocks in the error term may also be systematically related to the size of the migrant network, M^i_{jt}, and to the possibility of falling into poverty, and thus contribute an additional endogeneity concern.

From the model specified in (3.18), we are particularly interested in identifying the coefficients on pov_{it-1}, M^i_{jt} and $M^i_{jt} * pov_{it-1}$. The coefficients on pov_{it-1} and $M^i_{jt} * pov_{it-1}$ allow us to gauge the importance of persistence in the probability that a household is poor, and the impact of access to migrant employment opportunities through the migrant network on poverty persistence. β_3, the coefficient on M^k_{it}, allows us to determine the impact of the migrant network on the probability that a household will fall into poverty.

The specification shown in (3.18) may have additional sources of endogeneity if we believe that household demographic and human capital, $\mathbf{X}_{it}$, or land per capita, lpc_{it}, may vary with unobserved shocks in period t or $t-1$. We address the possible concern over endogenous household composition by using household demographic and human capital variables for the legal long-term registered residents of households. While household size may vary somewhat with shocks as individuals move in and out of the household for the purpose of finding temporary work elsewhere, such variations do not show up in registered household membership. Long-term membership only changes when households split subject to such events as marraige or legal change of residence to another location. Land managed by the household may also vary with shocks. Land markets in rural China do not function well: land cannot be bought and sold, and only in the last few years have farmers gained the right to explicitly transfer land. Instead land is allocated by village leaders, and reallocated or adjusted among households within village small groups if a household is judged to have too little land to support itself. Nonetheless, there is some possibility that reallocation

may be related to shocks that occur in period t or $t-1$ that may also be systematically related to poverty status and the migrant network size. Wooldridge (2002) shows that when the assumption of strict exogeneity of the regressors fails in the context of the standard FE estimation the inconsistency of the instrument is of order T^{-1}. We thus use the period $t-2$ value of land per capita and estimate:

$$pov_{it} = 1[\beta_1 pov_{it-1} + \beta_2(M^i_{jt} * pov_{it-1}) + \beta_3 M^i_{jt} + \mathbf{X}'_{it}\alpha_1 + \alpha_2 lpc_{it-2} + \mathbf{D}_t + u_i + \varepsilon_{it}], \tag{3.19}$$

One remaining issue remains in that we do not perfectly observe the network M^i_{jt} through which household i may use for job referrals. Instead, we observe the number of registered longterm village residents who are employed as migrants outside the village in a particular year, or M_{jt}. The true migrant network may include former legal registered residents who have now changed their long-term residence status, implying that the actual potential network is larger. Alternatively, the household may not be familiar with all of the village out-migrants, and thus the actual network through which a household may seek referrals may be smaller. Thus, we will estimate:

$$pov_{it} = 1[\beta_1 pov_{it-1} + \beta_2(M_{jt} * pov_{it-1}) + \beta_3 M_{jt} + \mathbf{X}'_{it}\alpha_1 + \alpha_2 lpc_{it-2} + \mathbf{D}_t + u_i + \varepsilon_{it}], \tag{3.20}$$

In our identification strategy below, we will instrument the endogenous number of village out-migrants, M_{jt}, with village level instruments, identifying the size of the village migrant labor force, interacted with period $t-2$ lagged land per capita, lpc_{it-2}, in order to allow for differences in the effective value of the village migrant network for households with different amounts of land.

Why might we expect that interacting with lpc_{it-2} might achieve this? We believe that the land per capita managed by households will likely pick up a dimension of proximity of different households within the village. Within villages in rural China, households are separated into smaller units of roughly 20 households known as village small groups (*cun xiaozu*), which were referred to as *production teams* during

the Maoist period. These households are located in clusters and will have closer relationships with one another than with households of other small groups. Moreover, property rights to land in rural China typically reside with the small group, not with the village. Thus, when land reallocations take place they typically take place within but not across small groups. Small groups make more frequent small adjustments to household land as the land per capita available starts to become unequal with differential changes in household structure across households within the small group, but there is much less flexibility in making adjustments across small groups. As a result, much of the variability of land per capita within villages occurs across small groups.[13] Interacting a village level instrument for the migrant network with land per capita will allow the importance of M_{jt} to vary across households, and much of the difference across households occurs because of unobserved differences in the small groups in which they reside and from which migrants refer to as home.

3.3.5 Identifiying the Migrant Network

To instrument the village migrant network, we make use of two policy changes that, working together, affect the strength of migrant networks outside home counties but are plausibly unrelated to the demand for and supply of schooling. First, a new national ID card (*shenfen zheng*) was introduced in 1984. While urban residents received IDs in 1984, residents of most rural counties did not receive them immediately. In 1988, a reform of the residential registration system made it easier for migrants to gain legal temporary residence in cities, but a national ID card was necessary to

[13] We do not know village small group membership in the RCRE survey prior to 2003 when a new survey instrument was introduced. If we regress land per capita on village dummy variables in 2003, we obtain an R-Squared of 0.503, while if we run a regression of land per capita on small group dummy variables, we obtain an R-Squared of 0.616. A Lagrange Multiplier test for whether the small group effects add anything significant over the village effects, which is effectively a test of whether small group coefficients are constant within villages, yields an LM statistic of 310.67, which has a p-value of 0.0000.

obtain a temporary residence permit (Mallee, 1995). While some rural counties made national IDs available to rural residents as early as 1984, others distributed them in 1988, and still others did not issue IDs until several years later. The RCRE follow-up survey asked local officials when IDs had actually been issued to rural residents of the county. In our sample, 41 of the 90 counties issued cards in 1988, but cards were issued as early as 1984 in three counties and as late as 1997 in one county. It is important to note that IDs were not necessary for migration, and large numbers of migrants live in cities without legal temporary residence cards. However, migrants with temporary residence cards have a more secure position in the destination community, hold better jobs, and would thus plausibly make up part of a longer-term migrant network in migrant destinations. Thus, ID distribution had two effects after the 1988 residential registration (*hukou*) reform. First, the costs of migrating to a city should fall after IDs became available. Second, if the quality of the migrant network improves with the years since IDs are available, then the costs of finding migrant employment should continue to fall over time.

As a result, the size of the migrant network should be a function of both whether or not cards have been issued and the time since cards have been issued in the village. Given that the size of the potential network has an upper bound, we expect the years-since-IDs-issued to have a non-linear relationship with the size of the migrant labor force and we expect growth in the migrant network to decline after initially increasing with distribution of IDs. In Figure C.4, we show a lowess plot of the relationship between years since IDs were distributed and the change in number of migrants from the village from year $t-1$ to t. Note the sharp increase in migrants from the time that IDs are distributed and then a slowing of the increase over time (which would imply an even slower growth rate). This pattern suggests non-linearity in the relationship between ID distribution and new participants in the village migrant labor force. We thus specify our instrument as a dummy variable indicating that IDs had

been issued interacted with the years since they had been issued, and then experiment with quadratic, cubic and quartic functions of years-since-IDs-issued. We settle on the quartic function for our instruments because, as we show below, it fits the pattern of expanding migrant networks better than the quadratic or the cubic functions.

Since ID distribution was the responsibility of county level offices of the Ministry of Civil Affairs, which are distinctly separate from agencies involved in setting policies affecting land, credit, taxation and poverty alleviation (the Ministry of Agriculture and Ministry of Finance handle most decisions that affect these policies at the local level), it is plausible that ID distribution is not be systematically related to unobservable policy decisions with more direct relationship to household consumption. Still, using a function of the years since IDs were issued is not an ideal identification strategy. Ideally, a policy would exist that was randomly implemented, affecting the ability to migrate from some counties but not others. As the differential timing of the distribution of ID cards was not necessarily random, we must be concerned that counties with specific characteristics or that followed specific policies were singled out to receive ID cards earlier than other counties, or that features of counties receiving IDs earlier are systematically correlated with other policies affecting consumption growth. These counties, one might argue, were "allowed" to build up migrant networks faster than others.

In an earlier paper, de Brauw and Giles (2007) address several possible concerns with use of the years-since-IDs quartic as instruments for the size of the village migrant labor force. They first show that timing of ID distribution appears to be related to remoteness of the village, but not systematically related to village policies affecting that may affect consumption growth, with village administrative capacity, or with the demand for IDs within the village. They thus argue in favor of including a village fixed effect to control for features of the local county which may have affected timing of ID distribution, and then identify the size of the village migrant labor force off of

non-linearities in the time that it requires for migrant networks to build up. In this paper, we interact the quartic in years-since IDs with pre-determined land per capita of households in period $t-2$ to identify the size of the village migrant network.

3.4 Results

Before estimating equation (3.20), we establish that our instruments are significantly related to size of the migrant labor force. We estimate the relationship as a quadratic, cubic, and quartic function of the years since IDs were issued each interacted with period $t-2$ land per capita. These results are reported in columns (1)–(3) and columns (4)–(6) of Table C.2 for each year from 1995-2001 and odd years from 1989-2001, respectively.[14] We find a strong relationship between our instruments and the size of the migrant network for each specification. For the remainder of our estimation we favor the quartic function interacted with $t-2$ land per capita for two reasons: First, the effects of ID card distribution on the migration network can be determined more flexibly when we use the quartic specification. Secondly, the partial R^2 increases slightly from the quadratic to the quartic for the both samples we consider. After controlling for the household characteristics, the instruments have jointly significant effects on the number of migrants from the village for both samples, with F-statistics of 39.82 and 54.65 for the 1995 to 2001 and odd year 1989 to 2001 samples, respectively.

We apply the method introduced in Section 3.2.2 to estimating equation (3.20). In Table C.3, we report estimation results based on the pure random effects and correlated random effects approaches. We obtain the pure RE estimation results using the Stata "xtprobit" command, where year dummies (not shown), residuals from the first stage estimation and their time averages (not shown), number of household

[14] Since the RCRE survey was not conducted in 1992 and 1994, we estimate the dynamic model with one year spacing from 1995 to 2001, and with two-year spacing from 1989 to 2001.

members, number of prime age household laborers, second lag of land per capita, average years of education, share of females, lagged poverty status, migration network, interaction between the lagged poverty status and the migration network are included as explanatory variables. The correlated RE estimation results are obtained using the Stata "xtprobit" command, where year dummies (not shown), residuals from the first stage estimation (not shown), first-stage residuals and all the exogenous explanatory variables in each time period (not shown), number of household members, number of prime age household laborers, second lag of land per capita, average years of education, share of females, lagged poverty status, migration network, interaction between the lagged poverty status and the migration network, and the poverty status in the initial time period are included as explanatory variables. For purposes of comparison, we also estimate model (3.20) using a naive linear probability model and provide the results in Table C.4. Even after controlling for the unobserved effect using our correlated RE approach, the coefficients on the lagged poverty status are highly statistically significant for explaining the current poverty status in both datasets considered. The positive sign of the lagged poverty status suggests that being poor in a previous period significantly increases the probability of being poor in a current period. The initial value of the poverty status is also very important. It implies that there is substantial correlation between the unobserved effect and the initial condition. The coefficient on the lagged poverty status in the initial time period (1.028 for 1989-2001 dataset and 1.161 for 1995-2001 dataset) is larger than the coefficient on the lagged poverty status (0.820 for 1989-2001 dataset and 1.523 for 1995-2001 dataset). The migrant network is statistically significant for explaining the poverty status in all models but the correlated RE using the 1995-2001 dataset. Interaction between the migration network and the poverty status is also statistically significant at 0.01 level for every case we consider. The negative sign of the interaction term suggests that those households that were poor in the previous period are less likely to benefit

from the increases in the size of the migration network in the current period. In other words, in our application, we find that migration is important for reducing the likelihood that poor households remain in poverty and that non-poor households fall into poverty. Further, failure to control for unobserved heterogeneity leads to an overestimate of the impact of migrant labor markets on probability of staying poor of those who lived below the poverty lines. The coefficient on the interaction term between the migration network and the lagged poverty status for the pure RE approach (-0.188 for 1989-2001 dataset and -0.180 for 1995-2001 dataset) is larger in absolute value than the coefficient on the interaction term for the correlated RE method (-0.108 for 1989-2001 dataset and -0.137 for 1995-2001 dataset).

In Table C.5 we show the APEs for both models considered for both data samples. For example, for the sample from 1989 to 2001, the correlated random effects CF estimate of the APE of 100 more members in the migration network for those who were living above the poverty level is to reduce the probability of being poor by about 3.6 percentage points. For those who lived below the poverty line, the correlated random effects CF estimate of the APE of 100 more members in the migration network is to reduce the probability of being poor by 5.7 percentage points. Interestingly the APEs calculated using the correlated random effects dynamic probit approach are generally smaller than those calculated using the linear probability model. This suggests that using a naive LPM approach might lead us to conclude that migraton has a stronger impact on poverty reduction than found using the correlated random effects probit model.

3.5 Conclusions

In this paper, we have developed a dynamic binary response panel data model that allows for an endogenous regressor. This estimation approach is of particular value

for settings in which one wants to estimate the effects of a treatment which is also endogenous. We next apply the model to examine the impact of rural-urban migration on the likelihood that households in rural China fall below the poverty line. In our application, we find that migration is important both for reducing the likelihood that households remain in poverty or fall into poverty if they were not poor in the previous period.

APPENDIX A

Tables for Chapter 1

Table A.1. **Usual Unobserved Effects CRC Model for $\beta = 2$ and $T = 5$**

		(1)	(2)	(3)	(4)	(5)	(6)
Estimator	Time Dummies?	Mean	SD	RMSE	LQ	Median	UQ
			$N = 100$				
POLS	no	3.363	.189	1.377	3.238	3.356	3.486
FE-OLS	no	2.616	.138	0.642	2.527	2.621	2.711
IV	no	2.752	.225	0.781	2.612	2.761	2.901
FE-IV	no	2.423	.214	0.484	2.288	2.429	2.558
FE-IV	yes	1.945	.407	0.407	1.711	1.980	2.208
			$N = 400$				
POLS	no	3.369	.091	1.372	3.299	3.366	3.434
FE-OLS	no	2.623	.067	0.635	2.575	2.626	2.667
IV	no	2.745	.110	0.760	2.666	2.740	2.818
FE-IV	no	2.428	.096	0.455	2.362	2.423	2.498
FE-IV	yes	1.988	.177	0.213	1.887	1.997	2.101
			$N = 800$				
POLS	no	3.373	.063	1.375	3.330	3.366	3.412
FE-OLS	no	2.625	.046	0.637	2.596	2.624	2.655
IV	no	2.753	.076	0.764	2.700	2.750	2.801
FE-IV	no	2.436	.068	0.458	2.389	2.437	2.480
FE-IV	yes	2.004	.131	0.182	1.919	2.009	2.091

Table A.2. **Usual Unobserved Effects CRC Model for $\beta = 2$ and $T = 10$**

		(1)	(2)	(3)	(4)	(5)	(6)
Estimator	Time Dummies?	Mean	SD	RMSE	LQ	Median	UQ
			$N = 100$				
POLS	no	3.204	.157	1.223	3.097	3.195	3.314
FE-OLS	no	2.534	.106	0.562	2.469	2.531	2.603
IV	no	2.397	.123	0.440	2.324	2.395	2.475
FE-IV	no	2.277	.115	0.331	2.208	2.276	2.351
FE-IV	yes	2.013	.283	0.313	1.841	2.020	2.210
			$N = 400$				
POLS	no	3.196	.077	1.202	3.146	3.193	3.247
FE-OLS	no	2.528	.056	0.545	2.490	2.527	2.565
IV	no	2.392	.061	0.417	2.450	2.393	2.431
FE-IV	no	2.270	.060	0.305	2.231	2.274	2.308
FE-IV	yes	1.995	.138	0.186	1.901	2.002	2.092
			$N = 800$				
POLS	no	3.194	.054	1.200	3.155	3.194	3.224
FE-OLS	no	2.525	.040	0.541	2.498	2.523	2.551
IV	no	2.388	.042	0.410	2.357	2.387	2.416
FE-IV	no	2.268	.041	0.299	2.241	2.267	2.294
FE-IV	yes	1.992	.100	0.160	1.926	1.993	2.062

Table A.3. **Random Trend CRC Model for $\beta = 2$ and $T = 5$**

		(1)	(2)	(3)	(4)	(5)	(6)
Estimator	Time Dummies?	Mean	SD	RMSE	LQ	Median	UQ
			$N = 100$				
POLS	yes	4.293	.300	2.303	4.096	4.284	4.475
FE-OLS	yes	2.673	.182	0.697	2.555	2.671	2.782
IV	yes	2.929	.850	1.247	2.444	2.941	3.496
FE-IV	yes	2.000	.626	0.642	1.635	2.057	2.383
FE-IV	no	13.414	1.411	11.422	12.464	13.221	14.225
			$N = 400$				
POLS	yes	4.308	.144	2.312	4.201	4.307	4.411
FE-OLS	yes	2.663	.085	0.679	2.607	2.666	2.721
IV	yes	3.004	.411	1.073	2.704	3.023	3.292
FE-IV	yes	2.013	.269	0.301	1.835	2.019	2.204
FE-IV	no	13.406	.665	11.406	12.915	13.340	13.878
			$N = 800$				
POLS	yes	4.296	.097	2.294	4.225	4.295	4.363
FE-OLS	yes	2.660	.060	0.671	2.617	2.658	2.700
IV	yes	2.996	.278	1.038	2.809	2.993	3.171
FE-IV	yes	1.996	.187	0.223	1.874	2.005	2.130
FE-IV	no	13.351	.478	11.328	13.049	13.318	13.654

Table A.4. **Random Trend CRC Model for $\beta = 2$ and $T = 10$**

		(1)	(2)	(3)	(4)	(5)	(6)
Estimator	Time Dummies?	Mean	SD	RMSE	LQ	Median	UQ
			$N = 100$				
POLS	yes	4.789	.407	2.820	4.522	4.814	5.051
FE-OLS	yes	2.651	.178	0.687	2.539	2.656	2.761
IV	yes	2.916	1.042	1.401	2.357	2.976	2.615
FE-IV	yes	1.968	.619	0.641	1.603	2.001	2.384
FE-IV	no	15.933	.771	13.919	15.367	15.902	16.479
			$N = 400$				
POLS	yes	4.808	.190	2.815	4.678	4.808	4.943
FE-OLS	yes	2.662	.089	0.678	2.600	2.662	2.718
IV	yes	3.000	.504	1.137	2.659	2.993	3.361
FE-IV	yes	1.981	.311	0.338	1.767	1.978	2.203
FE-IV	no	15.900	.406	13.875	15.633	15.890	16.177
			$N = 800$				
POLS	yes	4.788	.144	2.784	4.682	4.779	4.885
FE-OLS	yes	2.663	.062	0.674	2.618	2.660	2.703
IV	yes	3.000	.360	1.061	2.759	3.026	3.243
FE-IV	yes	1.997	.201	0.234	1.855	1.998	2.132
FE-IV	no	15.904	.289	13.888	15.693	15.895	16.098

APPENDIX B

Tables for Chapter 2

Table B.1. **Usual Unobserved Effect CRC Model for Continuous** y_{2it}

(1)	(2)		(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Estimator	Mean	Reg. SE	Rob. SE	SD	RMSE	LQ	Median	UQ
					$\xi \neq 0$					
					$N = 500$					
$\hat{\rho}_{y_2z} =$	.255	POLS	2.998	0.030	0.130	0.134	1.015	2.903	2.996	3.087
$\hat{\rho}_{y_2u} =$	.249	IV	3.435	0.093	0.115	0.115	1.441	3.357	3.428	3.512
$\hat{\rho}_{y_2a} =$	.247	FE-OLS	2.919	0.027	0.100	0.100	0.932	2.853	2.919	2.989
$\hat{\rho}_{y_2b} =$	.576	FE-IV	3.400	0.053	0.112	0.110	1.406	3.326	3.394	3.471
		CF	2.059	Adj. SE = 0.056		0.057	0.150	2.022	2.061	2.099
					$N = 1000$					
$\hat{\rho}_{y_2z} =$	.255	POLS	2.991	0.021	0.093	0.094	1.004	2.929	2.993	3.056
$\hat{\rho}_{y_2u} =$	.248	IV	3.441	0.066	0.082	0.081	1.447	3.383	3.435	3.492
$\hat{\rho}_{y_2a} =$	.245	FE-OLS	2.913	0.019	0.071	0.073	0.923	2.865	2.916	2.966
$\hat{\rho}_{y_2b} =$	.575	FE-IV	3.405	0.038	0.080	0.079	1.411	3.346	3.399	3.454
		CF	2.057	Adj. SE = 0.040		0.039	0.143	2.031	2.058	2.083
					$\xi = 0$					
					$N = 500$					
$\hat{\rho}_{y_2z} =$	.257	POLS	2.808	0.049	0.098	0.098	0.585	2.739	2.802	2.873
$\hat{\rho}_{y_2u} =$	.255	IV	2.235	0.121	0.123	0.128	0.291	2.146	2.235	2.324
$\hat{\rho}_{y_2a} =$	.255	FE-OLS	2.184	0.034	0.082	0.083	0.235	2.129	2.183	2.236
$\hat{\rho}_{y_2b} =$	.346	FE-IV	2.003	0.077	0.107	0.108	0.166	1.926	2.001	2.082
		CF	2.002	Adj. SE = 0.090		0.089	0.154	1.942	2.000	2.064
					$N = 1000$					
$\hat{\rho}_{y_2z} =$	.260	POLS	2.809	0.035	0.070	0.069	0.821	2.762	2.809	2.856
$\hat{\rho}_{y_2u} =$	.255	IV	2.235	0.085	0.087	0.085	0.280	2.176	2.234	2.291
$\hat{\rho}_{y_2a} =$	.256	FE-OLS	2.181	0.024	0.058	0.058	0.227	2.141	2.181	2.218
$\hat{\rho}_{y_2b} =$	.346	FE-IV	2.000	0.054	0.076	0.075	0.147	1.950	2.001	2.051
		CF	2.000	Adj. SE = 0.064		0.062	0.140	1.957	1.999	2.042

Note: $\beta = 2$ and $T = 5$. All estimates are based on the regressions without time dummies.

Table B.2. **Random Trend CRC Model for Continuous** y_{2it}

(1)	(2)	Estimator	(3) Mean	(4) Reg. SE	(5) Rob. SE	(6) SD	(7) RMSE	(8) LQ	(9) Median	(10) UQ
					$\xi \neq 0$					
					$N = 500$					
$\hat{\rho}_{y_2z} =$	.250	POLS	3.298	0.043	0.152	0.153	1.309	3.188	3.296	3.406
$\hat{\rho}_{y_2u} =$	.248	IV	3.875	0.211	0.233	0.236	1.894	3.712	3.865	4.005
$\hat{\rho}_{y_2a_1} =$	.248	FE-OLS	2.767	0.029	0.096	0.104	0.778	2.693	2.769	2.837
$\hat{\rho}_{y_2a_2} =$	.247	FE-IV	3.438	0.078	0.152	0.157	1.449	3.324	3.429	3.541
$\hat{\rho}_{y_2b} =$	.578	CF	2.039	Adj. SE = 0.122		0.126	0.182	1.947	2.040	2.123
					$N = 1000$					
$\hat{\rho}_{y_2z} =$	.250	POLS	3.286	0.031	0.109	0.109	1.296	3.211	3.282	3.357
$\hat{\rho}_{y_2u} =$	.250	IV	3.867	0.149	0.165	0.171	1.869	3.746	3.857	3.977
$\hat{\rho}_{y_2a_1} =$	.247	FE-OLS	2.764	0.021	0.068	0.069	0.775	2.718	2.765	2.812
$\hat{\rho}_{y_2a_2} =$	.245	FE-IV	3.435	0.054	0.107	0.111	1.438	3.356	3.427	3.503
$\hat{\rho}_{y_2b} =$	.576	CF	2.037	Adj. SE = 0.087		0.084	0.157	1.988	2.039	2.091
					$\xi = 0$					
					$N = 500$					
$\hat{\rho}_{y_2z} =$	.259	POLS	3.691	0.095	0.141	0.142	1.698	3.594	3.678	3.776
$\hat{\rho}_{y_2u} =$	.254	IV	3.413	0.370	0.393	0.389	1.464	3.131	3.307	3.663
$\hat{\rho}_{y_2a_1} =$	.257	FE-OLS	2.215	0.035	0.087	0.085	0.262	2.158	2.212	2.273
$\hat{\rho}_{y_2a_2} =$	.255	FE-IV	1.987	0.138	0.192	0.197	0.235	1.853	1.991	2.109
$\hat{\rho}_{y_2b} =$	.346	CF	1.993	Adj. SE = 0.176		0.178	0.223	1.882	1.994	2.118
					$N = 1000$					
$\hat{\rho}_{y_2z} =$	.260	POLS	3.683	0.067	0.100	0.097	1.686	3.614	3.682	3.749
$\hat{\rho}_{y_2u} =$	.255	IV	3.400	0.259	0.275	0.267	1.442	3.223	3.396	3.589
$\hat{\rho}_{y_2a_1} =$	.256	FE-OLS	2.210	0.025	0.062	0.059	0.251	2.171	2.211	2.253
$\hat{\rho}_{y_2a_2} =$	.255	FE-IV	2.005	0.096	0.135	0.136	0.184	1.909	2.006	2.094
$\hat{\rho}_{y_2b} =$	.346	CF	2.001	Adj. SE = 0.124		0.120	0.172	1.921	1.999	2.081

Note: $\beta = 2$ and $T = 5$. All the estimates (but the CF) are based on the regressions with time dummies.

Table B.3. **Usual Unobserved Effect CRC Model for $y_{2it} \in (0,1)$**

(1)	(2)	Estimator	(3) Mean	(4) Reg. SE	(5) Rob. SE	(6) SD	(7) RMSE	(8) LQ	(9) Median	(10) UQ
					$\xi \neq 0$					
					$N = 500$					
$\hat{\rho}_{y_2z} =$	.246	POLS	6.366	0.152	0.241	0.240	4.349	6.158	6.336	6.488
$\hat{\rho}_{y_2u} =$	.257	IV	3.255	0.473	0.343	0.345	1.293	3.014	3.254	3.472
$\hat{\rho}_{y_2a} =$	.249	FE-OLS	2.528	0.134	0.176	0.179	0.570	2.412	2.527	2.656
$\hat{\rho}_{y_2b} =$	.580	FE-IV	2.682	0.205	0.247	0.240	0.736	2.532	2.677	2.857
		CF	2.150	Adj. SE = 0.221		0.233	0.303	2.001	2.147	2.302
					$N = 1000$					
$\hat{\rho}_{y_2z} =$	.247	POLS	6.328	0.107	0.171	0.163	4.326	6.219	6.330	6.436
$\hat{\rho}_{y_2u} =$	.257	IV	3.284	0.333	0.242	0.236	1.313	3.120	3.282	3.432
$\hat{\rho}_{y_2a} =$	.245	FE-OLS	2.535	0.095	0.125	0.122	0.560	2.450	2.529	2.617
$\hat{\rho}_{y_2b} =$	.579	FE-IV	2.703	0.145	0.175	0.175	0.731	2.573	2.695	2.813
		CF	2.137	Adj. SE = 0.157		0.157	0.241	2.030	2.138	2.237
					$\xi = 0$					
					$N = 500$					
$\hat{\rho}_{y_2z} =$	.252	POLS	5.754	0.199	0.233	0.226	3.760	5.604	5.745	5.888
$\hat{\rho}_{y_2u} =$	.250	IV	2.507	0.506	0.368	0.370	0.627	2.246	2.520	2.742
$\hat{\rho}_{y_2a} =$	.250	FE-OLS	2.727	0.099	0.133	0.131	0.745	2.639	2.729	2.812
$\hat{\rho}_{y_2b} =$	.346	FE-IV	1.874	0.229	0.263	0.253	0.313	1.698	1.888	2.054
		CF	2.012	Adj. SE = 0.256		0.241	0.278	1.839	2.024	2.169
					$N = 1000$					
$\hat{\rho}_{y_2z} =$	.255	POLS	5.754	0.141	0.165	0.158	3.754	5.647	5.749	5.857
$\hat{\rho}_{y_2u} =$	.249	IV	2.534	0.357	0.260	0.253	0.605	2.349	2.547	2.707
$\hat{\rho}_{y_2a} =$	.250	FE-OLS	2.718	0.070	0.094	0.092	0.734	2.654	2.718	2.784
$\hat{\rho}_{y_2b} =$	.343	FE-IV	1.882	0.161	0.187	0.178	0.246	1.755	1.872	2.005
		CF	2.005	Adj. SE = 0.181		0.186	0.221	1.874	2.003	2.125

Note: $\beta = 2$ and $T = 5$. All estimates are based on the regressions without time dummies.

Table B.4. **Random Trend CRC Model for** $y_{2it} \in (0,1)$

(1)	(2)		(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Estimator	Mean	Reg. SE	Rob. SE	SD	RMSE	LQ	Median	UQ
				$\xi \neq 0$						
				$N = 500$						
$\hat{\rho}_{y_2 z} =$	.247	POLS	9.892	0.333	0.620	0.608	7.880	9.508	9.889	10.286
$\hat{\rho}_{y_2 u} =$	.255	IV	6.569	1.500	1.534	1.479	4.820	5.578	6.603	7.541
$\hat{\rho}_{y_2 a_1} =$	.249	FE-OLS	3.079	0.204	0.284	0.277	1.116	2.889	3.058	3.270
$\hat{\rho}_{y_2 a_2} =$	.246	FE-IV	2.676	0.372	0.494	0.498	0.854	2.323	2.658	3.041
$\hat{\rho}_{y_2 b} =$	.581	CF	2.229	Adj. SE = 0.693		0.723	0.750	1.712	2.229	2.731
				$N = 1000$						
$\hat{\rho}_{y_2 z} =$	.247	POLS	9.854	0.235	0.440	0.462	7.863	9.539	9.843	10.142
$\hat{\rho}_{y_2 u} =$	.256	IV	6.569	1.056	1.079	1.030	4.651	5.821	6.624	7.266
$\hat{\rho}_{y_2 a_1} =$	.245	FE-OLS	3.077	0.144	0.201	0.199	1.092	2.951	3.072	3.221
$\hat{\rho}_{y_2 a_2} =$	.245	FE-IV	2.721	0.263	0.348	0.351	0.797	2.466	2.718	2.946
$\hat{\rho}_{y_2 b} =$	.581	CF	2.182	Adj. SE = 0.499		0.494	0.560	1.878	2.171	2.489
				$\xi = 0$						
				$N = 500$						
$\hat{\rho}_{y_2 z} =$	.254	POLS	10.055	0.458	0.521	0.541	8.060	9.695	10.026	10.408
$\hat{\rho}_{y_2 u} =$	.248	IV	7.633	1.834	1.872	1.827	5.897	6.333	7.573	8.844
$\hat{\rho}_{y_2 a_1} =$	.251	FE-OLS	2.875	0.106	0.156	0.156	0.899	2.754	2.881	2.995
$\hat{\rho}_{y_2 a_2} =$	.249	FE-IV	1.818	0.434	0.565	0.547	0.599	1.426	1.827	2.202
$\hat{\rho}_{y_2 b} =$	.342	CF	1.949	Adj. SE = 0.877		0.864	0.875	1.420	1.936	2.557
				$N = 1000$						
$\hat{\rho}_{y_2 z} =$	.254	POLS	10.034	0.323	0.368	0.359	8.027	9.776	10.030	10.273
$\hat{\rho}_{y_2 u} =$	.250	IV	7.611	1.286	1.311	1.248	5.724	6.700	7.643	8.507
$\hat{\rho}_{y_2 a_1} =$	.250	FE-OLS	2.875	0.075	0.110	0.110	0.886	2.810	2.874	2.945
$\hat{\rho}_{y_2 a_2} =$	.249	FE-IV	1.883	0.302	0.393	0.397	0.444	1.605	1.870	2.154
$\hat{\rho}_{y_2 b} =$	.342	CF	1.965	Adj. SE = 0.619		0.668	0.683	1.543	1.941	2.417

Note: $\beta = 2$ and $T = 5$. All the estimates (but the CF) are based on the regressions with time dummies.

Table B.5. **Standard Errors for the Control Function Approach**

(1)	(2)	(3)	(4)	(5)	(6)	(7)
ξ	N	Mean	Reg. SE	Rob. SE	Adj. SE	SD
y_{2it} on a large support set						
Usual Unobserved Effect CRC Model						
$\neq 0$	500	2.059	0.040	0.035	0.056	0.057
	1000	2.057	0.029	0.025	0.040	0.039
$= 0$	500	2.002	0.080	0.075	0.090	0.089
	1000	2.000	0.057	0.053	0.064	0.062
Random Trend CRC Model						
$\neq 0$	500	2.039	0.194	0.113	0.122	0.126
	1000	2.037	0.137	0.081	0.087	0.084
$= 0$	500	1.993	0.278	0.169	0.176	0.178
	1000	2.001	0.196	0.120	0.124	0.120
$y_{2it} \in (0, 1)$						
Usual Unobserved Effect CRC Model						
$\neq 0$	500	2.150	0.204	0.211	0.221	0.233
	1000	2.137	0.143	0.150	0.157	0.157
$= 0$	500	2.012	0.319	0.246	0.256	0.241
	1000	2.005	0.225	0.174	0.181	0.186
Random Trend CRC Model						
$\neq 0$	500	2.229	0.615	0.690	0.693	0.723
	1000	2.182	0.433	0.497	0.499	0.494
$= 0$	500	1.949	1.077	0.875	0.877	0.864
	1000	1.965	0.760	0.617	0.619	0.668

Table B.6. **Summary Statistics from Unbalanced and Balanced Datasets**

Variable	All Firms		Firms That Received Grant		Firms That Did Not Receive Grant	
	Unbalanced Dataset					
Scrap Rate (per 100 Items)	3.84	(6.01)	3.63	(5.02)	4.09	(7.01)
Hours of Training per Worker	14.97	(25.71)	19.39	(29.98)	10.98	(20.40)
Number of Firms	157		66		91	
	Balanced Dataset					
Scrap Rate (per 100 Items)	3.79	(5.64)	3.81	(5.15)	3.77	(6.37)
Hours of Training per Worker	15.60	(27.03)	19.30	(29.58)	10.06	(21.89)
Number of Firms	45		27		18	

Notes: (i) Quantities in parentheses are standard deviations. (ii) The means and standard deviations are calculated over the three-year period for each variable.

Table B.7. **POLS Estimates of the First Stage Regressions**

Variable	(1)	(2)	(3)	(4)	(5)	(6)
$grant_{it}$	35.410*	36.198*	28.697*	32.504*	29.185*	32.736*
	[5.934]	[5.515]	[5.364]	[5.024]	[5.294]	[5.081]
$\overline{grant_i}$	-7.693	-8.482	6.788	2.981	-2.352	-5.903
	[11.885]	[12.530]	[16.474]	[17.456]	[12.982]	[14.064]
dt_t	No	Yes	No	Yes	No	Yes
Control Variable(s)	No	No	Yes	Yes	Yes	Yes
R^2	.258	.291	.314	.332	.284	.304
Number of Observations	135	135	117	117	120	120

Notes: (i) Dependent variable is $hrsemp_{it}$; $grant_{it}$ is a dummy indicator for whether a grant was received, $\overline{grant_i}$ is a time average of $grant_{it}$. (ii) Quantities in square brackets are fully robust standard errors. (iii) Row called "dt_t" indicates whether a regression includes separate year intercepts for 1988 and 1989 on the first and on the second stages of estimation. (iv) Control variables in (3) and (4) include $\log(employ_{it})$ – the log of number of employees, $\log(sales_{it})$ – the log of annual sales, and $\log(avgsal_{it})$ – the log of average employee salary. Controls in (5) and (6) include $\log(avgsal_{it})$ only.

*Statistically significant at the .01 level.

Table B.8. **FE-IV and CF Estimates of the Second Stage Regressions**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variable	FE-IV	FE-IV	FE-IV	FE-IV	CF	CF	CF	CF	CF	CF
$hrsemp_{it}$	-.005**	-.002	-.002	-.001	-.040*	-.037*	-.035*	-.035*	-.038*	-.036*
	(.003)	(.003)	(.004)	(.004)	(.013)	(.014)	(.015)	(.015)	(.016)	(.016)
	[.002]	[.002]	[.003]	[.002]	[.013]	[.013]	[.013]	[.013]	[.017]	[.017]
					{.012}	{.012}	{.013}	{.013}	{.016}	{.016}
Control	No	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes
dt_t	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	135	135	120	120	135	135	120	120	120	120

Notes: (i) The dependent variable is the log of the output scrap rate per 100 items. (ii) Quantities in parentheses are regular standard errors, quantities in square brackets are robust standard errors, i.e., standard errors robust to general second moment misspecification (conditional variance and serial correlation), and quantities in curly brackets are standard errors adjusted for the first stage estimation. (iii) The control variable is the log of the average annual salary per worker. (iv) Row called "dt_t" indicates whether a regression includes separate year intercepts for 1988 and 1989 on the first and on the second stages of estimation. (v) For regressions (1) through (10), the instrumental variable is a dummy variable indicating whether or not a grant was received.

*Statistically significant at the .01 level; **at the .05 level (two-tailed tests using fully robust s.e. for the FE-IV estimates, and adjusted s.e. for the CF estimates).

Table B.9. **Summary Statistics for the Control Variables**

	All Firms		Firms That Received Grant		Firms That Did Not Receive Grant	
Employment	69.75	(70.88)	74.28	(79.91)	62.94	(54.66)
Sales	7,765,778	(8,118,983)	7,313,147	(8,415,131)	8,444,723	(7,683,374)
Average Annual Salary	18,247.73	(5,936.98)	18,625.79	(6,826.15)	17,606.67	(3,993.06)
Number of Firms	45		27		18	

Notes: (i) Quantities in parentheses are standard deviations. (ii) The means and standard deviations are calculated over the three-year period for each variable.

APPENDIX C

Tables and Figures for Chapter 3

Figure C.1. **Share of Village Labor Force Employed as Migrants by Year**

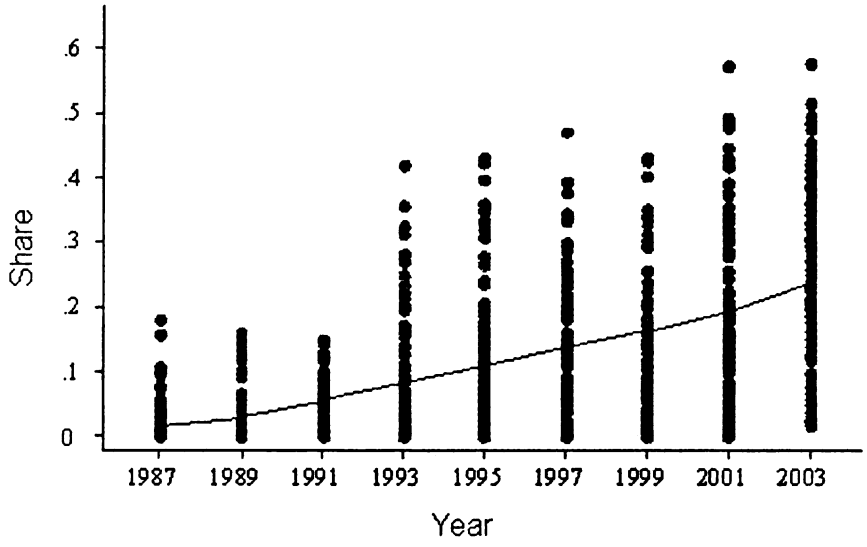

Figure C.2. **Village Consumption Growth**

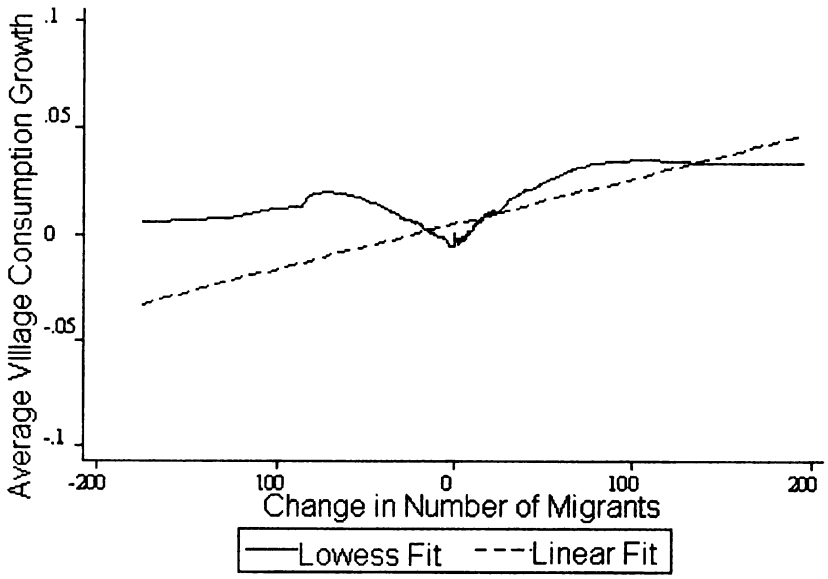

Figure C.3. **Change in Poverty Headcount**

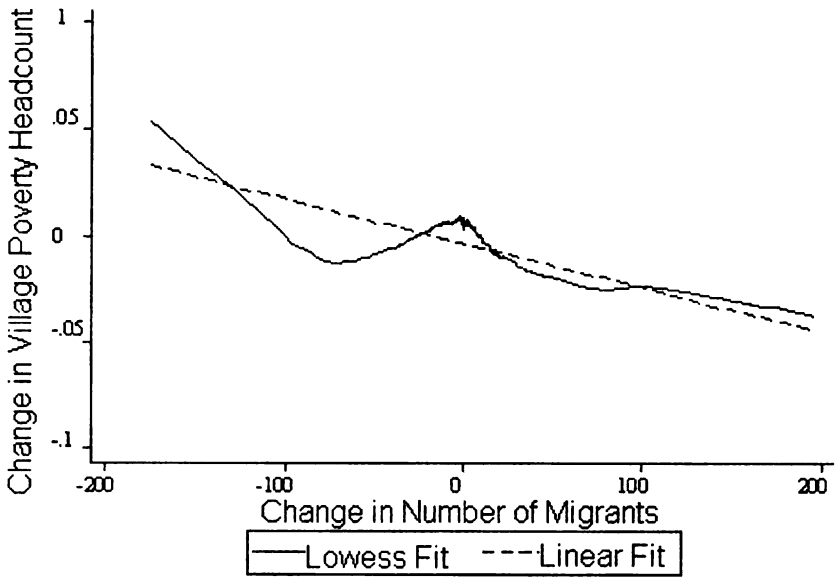

Figure C.4. **Change in Out-Migrants in Village Labor Force**

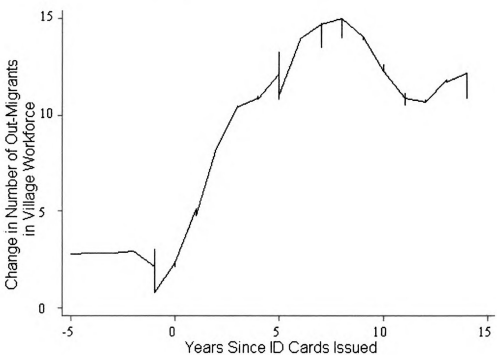

Table C.1. **Household and Village Characteristics**

		Years from 1995 to 2001				Odd Years from 1989 to 2001			
		Obs.	Full Sample	Obs.	Balanced Sample	Obs.	Full Sample	Obs.	Balanced Sample
Household Poverty Status	mean	41801	0.15	29974	0.15	44063	0.19	26614	0.203
	st. dev.		0.35		0.36		0.40		0.401
Household Income per Capita	mean	41856	824.3	29974	793.0	44057	732.9	26614	691.2
	st. dev.		785.6		686.4		656.6		554.6
Household Consumption per Capita	mean	41801	589.9	29974	570.6	44063	527.9	26614	502.1
	st. dev.		451.4		413.1		381.3		340.4
Number of Household Members	mean	41924	3.9	29974	4.1	44101	4.1	26614	4.2
	st. dev.		1.5		1.4		1.5		1.4
Number of Prime Age Household Laborers	mean	41924	2.5	29974	2.6	44101	2.5	26614	2.579
	st. dev.		1.1		1.0		1.1		1.012
Household Land per Capita	mean	41860	1.3	29974	1.2	44063	1.4	26614	1.4
	st. dev.		1.2		1.0		1.2		1.1
Household Average Years of Education	mean	40966	6.5	29974	6.5	43241	6.3	26614	6.3
	st. dev.		2.6		2.5		2.6		2.6
Household Share of Females	mean	40967	0.44	29974	0.45	43242	0.45	26614	0.45
	st. dev.		0.21		0.20		0.21		0.20
Number of Migrants from the Village	mean	41924	112.9	29974	116.9	44101	77.4	26614	78.0
	st. dev.		107.5		105.9		97.1		94.6
Year of ID Distribution in a Village	mean	41379	1988.05	29974	1987.96	43424	1988.01	26614	1988.00
	st. dev.		2.111		2.076		2.056		2.083
Years Since ID was Issued in a Village	mean	41379	9.980	29774	10.047	43424	6.849	26614	7.095
	st. dev.		2.882		2.846		4.327		4.311

Notes: Consumption and income per capita are reported in 1986 RMB Yuan.

Table C.2. **Factors Determining the Size of the Village Migrant Network**
First-Stage Regressions

	Dependent Variable: Number of Migrants					
	Years from 1995 to 2001			Odd Years from 1989 to 2001		
Model	(1)	(2)	(3)	(4)	(5)	(6)
Household Population	-1.709**	-1.763**	-1.785**	-2.335***	-2.338***	-2.341***
	(0.867)	(0.862)	(0.862)	(0.656)	(0.655)	(0.655)
Number of Working Age Laborers in the Household	2.516**	2.300**	2.370**	4.348***	4.269***	4.284***
	(1.046)	(1.037)	(1.037)	(0.773)	(0.771)	(0.770)
Land Per Capita $_{t-2}$	-9.363***	-3.719***	-4.752***	10.532***	14.432***	15.540***
	(2.666)	(1.337)	(1.414)	(1.503)	(1.825)	(2.067)
Average Years of Education	-0.293	-0.267	-0.272	-0.190	-0.193	-0.206
	(0.345)	(0.343)	(0.343)	(0.270)	(0.270)	(0.270)
Female Share of the Household	-0.410	-0.564	-0.707	1.805	2.023	1.999
	(3.139)	(3.123)	(3.126)	(2.925)	(2.923)	(2.924)
(Years-Since-ID-Issued) * (Land Per Capita $_{t-2}$)	2.020***	-3.899***	-7.396***	-0.169	-2.513***	-3.925***
	(0.421)	(0.462)	(0.823)	(0.273)	(0.535)	(0.802)
(Years-Since-ID-Issued)^2 * (Land Per Capita $_{t-2}$)	-0.100***	0.772***	1.779***	-0.093***	0.246***	0.633***
	(0.020)	(0.081)	(0.241)	(0.015)	(0.071)	(0.167)
(Years-Since-ID-Issued)^3 * (Land Per Capita $_{t-2}$)		-0.034***	-0.123***		-0.014***	-0.050***
		(0.003)	(0.022)		(0.003)	(0.015)
(Years-Since-ID-Issued)^4 * (Land Per Capita $_{t-2}$)			0.003***			0.001**
			(0.001)			(0.000)
Observations	25692	25692	25692	22812	22812	22812
R-squared	0.09	0.10	0.10	0.22	0.22	0.22
F-Statistic on IVs with Averages	11.80	23.10	34.21	55.84	37.59	29.68
F-Statistic on IVs w/o Averages	12.63	40.28	39.82	109.40	71.81	54.65
Partial R^2, IVs with Averages	0.003	0.009	0.015	0.007	0.007	0.008
Partial R^2, IVs w/o Averages	0.000	0.001	0.002	0.006	0.006	0.006

Notes: In parenthesis we show fully robust standard errors [*** p<0.01, ** p<0.05, * p<0.1]. All regressions include time averages of the explanatory variables and year dummies.

Table C.3. **CF Approach to Estimating Determinants of Poverty Status**
Second-Stage Regressions

	Dependent Variable: Poverty Status			
	Years from 1995 to 2001		Odd Years from 1989 to 2001	
	(1)	(2)	(3)	(4)
Model	Pure RE	Correlated RE	Pure RE	Correlated RE
Lagged Dependent Variable	1.614***	1.161***	1.252***	0.820***
	(0.054)	(0.055)	(0.050)	(0.048)
Interaction of (Migrant Network)/100 and Lagged	-0.180***	-0.137***	-0.188***	-0.108***
Dependent Variable	(0.034)	(0.035)	(0.034)	(0.034)
Mig Network/100	-0.090**	-0.016	-0.154***	-0.188***
	(0.045)	(0.051)	(0.038)	(0.040)
Number of Household Members	0.271***	0.424***	0.321***	0.354***
	(0.022)	(0.038)	(0.020)	(0.027)
Number of Prime Age Household Laborers	-0.102***	-0.081*	-0.155***	-0.137***
	(0.028)	(0.041)	(0.023)	(0.028)
Second Lag of Land per Capita	0.041*	-0.010	-0.015	-0.031
	(0.023)	(0.042)	(0.020)	(0.036)
Average Years of Education	-0.052***	-0.001	-0.044***	-0.011
	(0.008)	(0.013)	(0.007)	(0.010)
Share of Females	-0.271***	-0.005	-0.258***	0.012
	(0.102)	(0.129)	(0.092)	(0.113)
Dependent Variable in 1995/1989		1.523***		1.028***
		(0.070)		(0.051)
Observations	24692	25692	22812	22812
Number of households	4282	4282	3802	3802
Replications for Bootstrap Errors	250	250	250	250

Notes: In parenthesis we show bootstrapped standard errors [*** p<0.01, ** p<0.05, * p<0.1]. All regressions include the explanatory variables in each year, and year dummies. Regressions (1) and (3) include first stage residuals free of serial-correlation and their time averages. Regressions (2) and (4) include first stage residuals free of serial-correlation and residuals from the first stage in each year. For regressions (1) through (4) the instrumental variables are quartic polynomial of years-since-ID-was-issued with each term interacted with second lag of land per capita.

Table C.4. **Linear Probability Model for Determinants of Poverty Status**
Second-Stage Regressions

	Dependent Variable: Poverty Status			
	Years from 1995 to 2001		Odd Years from 1989 to 2001	
	Pure RE	Correlated RE	Pure RE	Correlated RE
Model	(1)	(2)	(3)	(4)
Lagged Dependent Variable	0.659***	0.542***	0.495***	0.399***
	(0.013)	(0.015)	(0.012)	(0.012)
Interaction of Mig Network/100 and Lagged Dependent Variable	-0.059***	-0.048***	-0.063***	-0.047***
	(0.010)	(0.009)	(0.009)	(0.008)
Mig Network/100	-0.010**	0.003	-0.028***	-0.036***
	(0.005)	(0.005)	(0.006)	(0.006)
Number of Household Members	0.031***	0.058***	0.048***	0.056***
	(0.002)	(0.005)	(0.003)	(0.004)
Number of Prime Age Household Laborers	-0.013***	-0.014***	-0.025***	-0.021***
	(0.003)	(0.005)	(0.003)	(0.004)
Second Lag of Land per Capita	0.005**	0.001	-0.003	0.001
	(0.002)	(0.005)	(0.002)	(0.004)
Average Years of Education	-0.007***	0.000	-0.008***	-0.001
	(0.001)	(0.002)	(0.001)	(0.001)
Share of Females	-0.035***	-0.003	-0.054***	-0.001
	(0.011)	(0.016)	(0.012)	(0.016)
Dependent Variable in 1995/1989		0.212***		0.135***
		(0.013)		(0.008)
Observations	25692	25692	22812	22812

Notes: In parenthesis we show bootstrapped standard errors [*** p<0.01, ** p<0.05, * p<0.1]. All regressions include the explanatory variables in each year, and year dummies. Regressions (1) and (3) include first stage residuals free of serial-correlation and their time averages. Regressions (2) and (4) include first stage residuals free of serial-correlation and residuals from the first stage in each year. For regressions (1) through (4) the instrumental variables are quartic polynomial of years-since-ID-was-issued with each term interacted with second lag of land per capita.

Table C.5. **Average Partial Effects of Determinants of Poverty Status**

	Years from 1995 to 2001				Odd Years from 1989 to 2001			
	LPM		Control Function		LPM		Control Function	
	Pure RE	Correlated RE	Pure RE	Correlated RE	Pure RE	Correlated RE	Pure RE	Correlated RE
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lagged Poverty Status	0.58866***	0.48379***	0.32241***	0.18687***	0.43793***	0.35639***	0.24716***	0.14038***
	(0.01026)	(0.01201)	(0.00630)	(0.00535)	(0.00991)	(0.01050)	(0.00712)	(0.00611)
Migration Network when $y1_{t-1}=0$	-0.00010**	0.00003	-0.00017*	-0.00003	-0.00028***	-0.00036***	-0.00033***	-0.00036***
	(0.00005)	(0.00005)	(0.00010)	(0.00009)	(0.00006)	(0.00006)	(0.00009)	(0.00008)
Migration Network when $y1_{t-1}=1$	-0.00069***	-0.00046***	-0.00050***	-0.00025***	-0.00091***	-0.00083***	-0.00072***	-0.00057***
	(0.00010)	(0.00010)	(0.00010)	(0.00009)	(0.00011)	(0.00010)	(0.00011)	(0.00010)
Number of Household Members	0.03123***	0.05809***	0.05018***	0.06878***	0.04819***	0.05633***	0.06761***	0.06814***
	(0.00219)	(0.00456)	(0.00457)	(0.00648)	(0.00253)	(0.00367)	(0.00456)	(0.00537)
Number of Prime Age Household Laborers	-0.01259***	-0.01354***	-0.01896***	-0.01310*	-0.02445***	-0.02106***	-0.03267***	-0.02629***
	(0.00284)	(0.00518)	(0.00597)	(0.00705)	(0.00304)	(0.00395)	(0.00535)	(0.00559)
Second Lag of Land per Capita	0.00499**	0.00107	0.00764	-0.00161	-0.00328	0.00089	-0.00310	-0.00602
	(0.00213)	(0.00504)	(0.00487)	(0.00721)	(0.00243)	(0.00442)	(0.00459)	(0.00710)
Average Years of Education	-0.00686***	0.00042	-0.00968***	-0.00008	-0.00768***	-0.00083	-0.00929***	-0.00218
	(0.00078)	(0.00151)	(0.00178)	(0.00226)	(0.00097)	(0.00138)	(0.00171)	(0.00198)
Share of Females	-0.03520***	-0.00292	-0.05009**	-0.00083	-0.05370***	-0.00094	-0.05447***	0.00223
	(0.01094)	(0.01563)	(0.02151)	(0.02200)	(0.01245)	(0.01557)	(0.02113)	(0.02229)
Poverty Status in 1995/1989		0.21219***		0.30778***		0.13467***		0.22314***
		(0.01330)		(0.01360)		(0.00804)		(0.00989)
Replications	250	250	250	250	250	250	250	250

Notes: In parenthesis we show bootstrapped standard errors [*** p<0.01, ** p<0.05, * p<0.1].

BIBLIOGRAPHY

Angrist, J. D. (1991). Instrumental Variables Estimation of Average Treatment Effects in Econometrics and Epidemiology. *National Bureau of Economics Research Technical Working Paper Number 115.*

Ahn, S. C., Y. H. Lee, and P. Schmidt. (2001). GMM Estimation of Linear Panel Data Models with Time-varying Individual Effects. *Journal of Econometrics* 101, 219-255.

Bane, M. J. and D. T. Ellwood. (1986). Slipping into and out of Poverty: The Dynamics of Spells. *Journal of Human Resources*, 21(1), 1-23.

Benjamin, D., L. Brandt, and J. Giles. (2005). The Evolution of Income Inequality in Rural China. *Economic Development and Cultural Change* 53(4), 769-824.

Cai, F., A. Park, and Y. Zhao. (2007). The Chinese Labor Market, chapter prepared for China's Great Economic Transition, Loren Brand and Thomas Rawski (eds), Cambridge University Press (in press).

Cappellari, L. (1999). Minimum Distance Estimation of Covariance Structures, 5th UK Meeting of Stata Users.

Card, D. (2001). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica* 52, 1199-1218.

Chamberlain, G. (1980). Analysis of Covariance with Qualitative Data. *Review of Economic Studies* 47, 225-238.

Chamberlain, G. (1984). Panel Data, in Handbook of Econometrics Volume 2, Z. Griliches and M. D. Intriligator, (Eds.). Amsterdam: North Holland, 1247-1318.

Chan, K. W. and L. Zhang. (1999). The Hukou System and Rural-Urban Migration in China: Processes and Changes. *China Quarterly* 160, 818-55.

Chay, K. Y. and D. R. Hyslop. (2000). Identification and Estimation of Dynamic Binary Response Models: Empirical Evidence Using Alternative Approaches, mimeo.

Chen, S. and M. Ravallion. (1996). Data in Transition: Assessing Rural Living Standards in Southern China. *China Economic Review* 7(1), 23-56.

Cornwell, C., P. Schmidt, and R. C. Sickles. (1990). Production Frontiers with Cross-sectional and Time-series Variation in Efficiency Levels. *Journal of Econometrics* 46, 185-200.

de Brauw, A. and J. Giles. (2007). Migrant Labor Markets and the Welfare of Rural Households in the Developing World: Evidence from China. Michigan State University, Department of Economics, Mimeo.

Du, Y., A. Park, and S. Wang. (2005). Is Migration Helping China's Poor? *Journal of Comparative Economics* 33(4), 688-709.

Garen, J. (1984), The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable. *Econometrica* 52, 1199-1218.

Giles, J. (2006). Is Life More Risky in the Open? Household Risk-Coping and the Opening of China's Labor Markets. *Journal of Development Economics* 81(1), 25-60.

Giles, J. and K. Yoo. (2006). Precautionary Behavior, Migrant Networks and Household Consumption Decisions: An Empirical Analysis Using Household Panel Data from Rural China. *Review of Economics and Statistics* (in press).

Hahn, J. and G. Kuersteiner. (2002). Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T Are Large. *Econometrica* 70, 1639-1657.

Hall, R. E. and C. I. Jones. (1999), Why Do Some Countries Produce So Much More Output per Worker than Others?. *Quarterly Journal of Economics* 114, 83-116.

Heckman, J. J. (1981). The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time - Discrete Data Stochastic Process, in C.F. Manski and D. McFadden, (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA, 179-195.

Heckman, J. J. (1997). Instrumental variables: A study of Implicit Behavioral Assumptions Used in Making Program Evaluations. *Journal of Human Resources* 32, 441-462.

Heckman, J. J. and E. Vytlacil. (1998), Instrumental Variables Methods for the Correlated Random Coefficient Model. *Journal of Human Resources* 33, 974-987.

Heckman, J. J. and E. Vytlacil. (2005). Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica* 73, 669-738.

Holzer, H., R. Block, M. Cheatham, and J. Knott. (1993), Are Training Subsidies for Firms Effective? The Michigan Experience. *Industrial and Labor Relations Review* 46, 625-636.

Honor'e, B. E. and E. Kyriazidou. (2000). Panel Data Discrete Choice Models with Lagged Dependent Variables. *Econometrica* 68, 839-874.

Hyslop, D. R. (1999). State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women. *Econometrica* 67(6), 1255-94.

Imbens, G. and J. D. Angrist. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica* 62, 467-476.

Jalan, J. and M. Ravallion. (1998). Transient Poverty in Post-Reform Rural China. *Journal of Comparative Economics* 26(2), 338-357.

Jalan, J. and M. Ravallion. (2002). Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China. *Journal of Applied Econometrics* 17(4), 329-46.

Liang, Z. and Z. Ma. (2004). China's Floating Population: New Evidence from the 2000 Census. *Population and Development Review* 30(3), 467-488.

Mallee, H. (1995). China's Household Registration System Under Reform, *Development and Change* 26(1), 1-29.

Meng, X. (2000). Regional Wage Gap, Information Flow, and Rural-urban Migration in Y. Zhao and L. West (eds) Rural Labor Flows in China, Berkeley: University of California Press, 251-277.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica* 46, 69-85.

Murtazashvili, I. (2006). A Control Function Approach to Estimation of Correlated Random Coefficient Panel Data Models. Michigan State University Department of Economics, Mimeo.

Murtazashvili, I. and J. M. Wooldridge. (2005), Fixed Effects Instrumental Variables Estimation in Correlated Random Coefficient Panel Data Models, Mimeo, Michigan State University Department of Economics, Mimeo.

Ravallion, M. and S. Chen. 2007. China's (Uneven) Progress Against Poverty. *Journal of Development Economics* 82(1), 1-42.

Rivers, D. and Q. H. Vuong. (1988), Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models. *Journal of Econometrics* 39, 347-366.

Rozelle, S., L. Guo, M. Shen, A. Hughart and J. Giles. (1999). Leaving China's Farms: Survey Results of New Paths and Remaining Hurdles to Rural Migration. *China Quarterly* 158, 367-393.

Semykina, A. and J. M. Wooldridge. (2005). Estimating Panel Data Models in the Presence of Endogeneity and Selection: Theory and Application. Michigan State University Department of Economics, Mineo.

Smith, R. and R. Blundell. (1986). An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply. *Econometrica* 54, 679-685.

Taylor, J. E., S. Rozelle, and A. de Brauw. (2003). Migration and Incomes in Source Communities: A New Economics of Migration Perspective from China. *Economic Development and Cultural Change* 52(1), 75-101.

Wooldridge, J. M. (1997). On Two Stage Least Squares Estimation of the Average Treatment Effect in a Random Coefficient Model. *Economics Letters* 56, 129-133.

Wooldridge, J. M. (2000). A Framework for Estimating Dynamic, Unobserved Effects Panel Data Models with Possible Feedback to Future Explanatory Variables. *Economics Letters* 68, 245-250.

Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data (MIT Press, Cambridge, MA).

Wooldridge, J. M. (2003). Further Results on Instrumental Variables Estimation of the Average Treatment Effect in the Correlated Random Coefficient Model. *Econometric Theory* 79, 185-191.

Wooldridge, J. M. (2005a). Fixed Effects and Related Estimators in Correlated Random Coefficient and Treatment Effect Panel Data Models. *Review of Economics and Statistics* 87, 385-390.

Wooldridge, J. M. (2005b). Unobserved Heterogeneity and Estimation of Average Partial Effects, in Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas J. Rothenberg}. Donald W.K. Andrews and James H. Stock (eds.). Cambridge: Cambridge University Press), 27-55.

Wooldridge, J. M. (2005c). Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity. *Journal of Applied Econometrics* 20, 39-54.

Zhao, Y. (2003). The Role of Migrant Networks in Labor Migration: The Case of China. *Contemporary Economic Policy* 21(4), 500-511.

MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 02956 0251