A Control Function Approach to Estimating Dynamic Probit Models with Endogenous Regressors*

John Giles† and Irina Murtazashvili‡

October 20, 2011

Abstract

This paper proposes a parametric approach to estimating a dynamic binary response panel data model that allows for endogenous contemporaneous regressors. This approach is of particular value for settings in which one wants to estimate the effects of an endogenous treatment on a binary outcome. The model is next used to examine the impact of rural-urban migration on the likelihood that households in rural China fall below the poverty line. In this application, it is shown that migration is important for reducing the likelihood that poor households remain in poverty and that non-poor households fall into poverty. Furthermore, it is demonstrated that failure to control for unobserved heterogeneity would lead the researcher to underestimate the impact of migrant labor markets on reducing the probability of falling into poverty.

JEL Codes: C13, C33, O15, P25

Key Words: Dynamic Binary Response Models; Control Function Approach; Poverty-Persistence; Migration; Rural China

*The paper has benefitted from helpful comments and conversations with Alan de Brauw, Ana Maria Herrera, Martin Ravallion, Peter Schmidt, David Tschirley, Adam Wagstaff, Jeffrey Wooldridge, from seminar participants at Ohio State University, University of South Carolina, George Washington University and conference participants at the June 2007 UNU-WIDER Conference on Fragile States held in Helsinki and the September 2009 Midwest Econometrics Group Annual Meeting. We gratefully acknowledge financial support for data collection from the National Science Foundation (SES-0214702), the Michigan State University Intramural Research Grants Program, the Ford Foundation (Beijing) and the Weatherhead Center for International Affairs (Academy Scholars Program) at Harvard University. The research discussion and conclusions presented in this paper reflect the views of the authors and should not be attributed to the World Bank or to any affiliated organization or member country.

†Development Research Group, The World Bank. Email: jgiles@worldbank.org.

‡*Corresponding Author*. Department of Economics, University of Pittsburgh, Pittsburgh, PA 15260. Tel: (412)648-1762, fax: (412)648-1793, and e-mail: irinam@pitt.edu.

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 propose a parametric approach to estimating dynamic binary response panel data models with endogenous contemporaneous regressors. Our method combines a recent approach to solving the unobserved heterogeneity and the initial conditions problems in non-linear dynamic models (Wooldridge, 2005) with a control function approach to controlling for endogeneity of contemporaneous explanatory variables in non-linear models (e.g., Smith and Blundell, 1986; Rivers and Vuong, 1988; Papke and Wooldridge, 2008).

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. Recently, Papke and Wooldridge (2008) employ this approach to deal with an endogenous regressor in a static fractional response panel data model to study the effects of school inputs on student performance. In contrast with Papke and Wooldridge, this paper develops

[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 (Devicienti and Poggi, 2007), and the identification of adverse selection in insurance markets (Chiappori and Salanie, 2000).

a control function approach for a dynamic model. To deal with the dynamic nature of the model, we consider two possibilities. We first use a "pure" random effects approach which assumes that unobserved heterogeneity is 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 (2005). 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 increasing migration and the probability that households within villages have consumption levels below the poverty line. Our empirical analysis demonstrates an economically significant causal relationship between migration and poverty reduction in rural China. In performing this exercise, we highlight the usefulness of our econometric approach to settings in which the researcher must work from binary indicators of poverty status, which is often the only information available from administrative data sources.

The paper proceeds as follows. In the Section 2 below, we first review approaches to estimation of dynamic binary response panel data models, and then propose a general approach to estimating these models when there is an endogenous regressor. In Section 3, we introduce the rural China setting, and motivate a specific implementation of the model developed in Section 2, and finally describe a strategy for identifying the effect of migration on poverty within China's villages. In Section 4, we discuss our estimation results and the performance of the model, and then in Section 5 we summarize our results and discuss the potential value of the estimator introduced in the paper.

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

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 semi-parametric 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] Semi-parametric 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 semi-parametric 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 which do not place any 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 semi-parametric approach of Honoré and Kyriazidou (2000) falls into this class of models.[3]

Due to their computational simplicity, parametric methods have received greater attention than semi-parametric methods. There are four main parametric approaches, all employing conditional maximum likelihood (CMLE) analysis, that have been employed for estimation of the dynamic nonlinear panel data models in which all covariates other than the lagged dependent variable are strictly exogenous. 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 the exogenous explanatory variables,

[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.

$\mathbf{z}_i = (\mathbf{z}_{i1}, \mathbf{z}_{i2}, ..., \mathbf{z}_{iT})$, one obtains the density of $(y_{i1}, y_{i2}, ..., y_{iT})$ given the initial conditions, y_{i0}, and $\mathbf{z}_i$, by integrating out the c_i. We refer to the relationship between the observed exogenous covariates and the unobserved heterogeneity in the first method as one of *"pure" random effects* because we assume c_i to be independent of $\mathbf{z}_i$ and y_{i0}. While this method may provide a way to obtain consistent estimates of the model parameters, nonrandomness of the initial conditions requires a very strong and often implausible assumption of independence between the initial conditions and the unobserved effects.

A second parametric approach involves treating the initial conditions as random and specifying the density for 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, $\mathbf{z}_i$. The most important drawback of this approach, however, 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 drawbacks of the second method, it is computationally challenging. Since both 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 (2005) 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 Heckman's approach. For this method to work, one must 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, similar to Heckman's approach, 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}.

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

[5]The specification of this density in Wooldridge's method is motivated by Chamberlain's (1980) approach, which models the distribution of the unobserved effect conditional on the strictly exogenous variables.

In the next section we develop an approach to consistently estimating 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, popularized by Smith and Blundell (1986) and Rivers and Vuong (1988). The main idea of our approach is to add (control) variables into the structural model to control for endogeneity. We consider a model with two possible sources of endogeneity: correlation between the unobserved heterogeneity and a regressor, and correlation between a regressor and the structural error. For this reason, we model the relationships among the unobserved effect, exogenous covariates, and the error from the reduced form equation for the endogenous explanatory variable.

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:

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

$$y_{1it} = 1[y^*_{1it} \geq 0],\ \ t = 1, ..., T, \tag{2}$$

where $\mathbf{z}_{1it}$ is an $1 \times (K-1)$ vector of exogenous covariates, which may contain a constant term, y_{2it} is an endogenous covariate[6], c_{1it} 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 $(\beta_1', \beta_2, \rho)\prime$, which is a $(K+1) \times 1$ vector of 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 $t-1$. We make two key assumptions on the conditional distribution of interest, denoted by $D(y_{1it}|\mathbf{z}_{1it}, y_{2it}, y_{1i,t-1}, c_{1i})$. First, we assume that the dynamics in model (1) are correctly specified, i.e. once $\mathbf{z}_{1it}$, y_{2it} and c_{1i} are

[6] Generally, model (1) can contain a random vector of endogenous covariates $\mathbf{y}_{2it}$. Modification of our approach discussed in this section to accommodate a random vector of endogenous regressors is straightforward assuming we have sufficient number of instrumental variables.

conditioned on, the dynamics are first order. The second assumption is that $\mathbf{z}_{1it}$ are strictly exogenous, conditional on the unobserved effect, c_{1i}. We express both of these assumptions as follows.

Assumption 1: Assume

$$D(y_{1it}|\mathbf{z}_{1it}, y_{2it}, y_{1i,t-1}, c_{1i}) = D(y_{1it}|\mathbf{z}_{1i}, y_{2it}, y_{1i,t-1}, y_{1i,t-2}, ..., y_{1i0}, c_{1i}), \tag{3}$$

where $t = 1, ..., T$, $\mathbf{z}_{1i} = (\mathbf{z}_{1i1}, \mathbf{z}_{1i2}, ..., \mathbf{z}_{1iT})$ is a set of exogenous covariates from (1) in all time periods.

Dynamic completeness of the model implies that the error term u_{1it} is serially uncorrelated. Allowing u_{1it} to have arbitrary serial correlation would suggest including more lags of the observed dependent variable in (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, y_{it}, actually depends on not only $y_{i,t-1}$ but also $y_{i,t-2}$. In other words, there is some logical inconsistency in studying a model with first order dynamics *and* serially correlated error. Thus, we assume the dynamics in our model are first order, and the error term has no serial correlation.

Next, we assume a linear equation for the endogenous scalar, y_{2it}, in terms of all available instrumental variables, which we assume to be strictly exogenous, conditional on c_{1it}: contemporaneous covariates, $\mathbf{z}_{1it}$, and those omitted from (1), $\mathbf{z}_{2it}$:

Assumption 2: Assume

$$y_{2it} = \mathbf{z}_{1it}\delta_1 + \mathbf{z}_{2it}\delta_2 + c_{2i} + u_{2it},\ t = 1, ..., T, \tag{4}$$

where u_{2it} is a serially uncorrelated idiosyncratic error with $\mathrm{Var}(u_{2it}) = \sigma_2^2$, and $\mathbf{z}_{it} = (\mathbf{z}_{1it}, \mathbf{z}_{2it})$ is an $1 \times L$ vector of instrumental variables, with $L \geq K$.[7]

Further, we model the heterogeneity c_{2i} as a linear function of the instrumental variables:

[7]In other words, we assume that the vector $\mathbf{z}_{2it}$ contains at least one element.

Assumption 3: Assume

$$c_{2i} = \overline{\mathbf{z}}_i \lambda + a_{2i}, \tag{5}$$

where $a_{2i}|\mathbf{z}_i \sim \text{Normal}(0, \sigma^2_{a_2})$, $\mathbf{z}_i = (\mathbf{z}_{1i}, \mathbf{z}_{2i}) = (\mathbf{z}_{i1}, \mathbf{z}_{i2}, ..., \mathbf{z}_{iT})$, and $\overline{\mathbf{z}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathbf{z}_{it}$ is the vector of time averages.

Equation (5) reflects our use of the Mundlak (1978) device for the unobserved effect, c_{2i}. We replace c_{2i} with its projection onto the time averages of all the exogenous variables. A less restrictive alternative to the Mundlak device is the specification by Chamberlain (1980), which models c_{2i} as a linear projection on $\mathbf{z}_{i1}, \mathbf{z}_{i2} ..., \mathbf{z}_{iT}$. Clearly, Mundlak's approach is a special case of Chamberlain's one: the former method is equivalent to restricting the effects of the exogenous covariates in each time period on the heterogeneity, c_{2i}, to be the same in the latter approach. While both approaches (as well as more flexible semi-parametric methods) are possible, the Mundlak device has an advantage of conserving on degrees of freedom, which is especially important when T is large. Thus, we adopt the Mundlak device.

Under Assumption 3 we can rewrite equation (4) as:

$$y_{2it} = \mathbf{z}_{it}\delta + \overline{\mathbf{z}}_i \lambda + v_{2it}, \;\; t = 1, ..., T, \tag{6}$$

where $v_{2it} = a_{2i} + u_{2it}$ is the new composite error term, and $\delta = (\delta_1', \delta_2')'$. We follow Papke and Wooldridge (2008) and refer to (6) as a reduced form equation for y_{2it}.

Next, we consider the relationship between u_{1it} and u_{2it}.

Assumption 4: Assume

$$(u_{1it}, u_{2it})|\mathbf{z}_i \sim \text{Normal}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \eta \\ \eta & \sigma_2^2 \end{pmatrix}\right). \tag{7}$$

Note that under joint normality of (u_{1it}, u_{2it}), with $\mathrm{Var}(u_{1it}) = 1$, we can write

$$\begin{aligned} u_{1it} &= \theta u_{2it} + e_{1it} \\ &= \theta(v_{2it} - a_{2i}) + e_{1it}, \end{aligned} \tag{8}$$

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

Equation (8) 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}.

Under Assumption 4, $e_{1it} \sim \mathrm{Normal}(0, \sigma_{e_1}^2)$, where $\sigma_{e_1}^2 = 1 - \xi^2$, since $\mathrm{Var}(u_{1it}) = 1$, and $\xi = \mathrm{Corr}(u_{1it}, u_{2it})$. Thus, we rewrite our main equation of interest as

$$\begin{aligned} y_{1it} &= 1[\mathbf{x}_{1it}\beta + c_{1i} + \theta(v_{2it} - a_{2i}) + e_{1it} \geq 0] \\ &= 1[\mathbf{x}_{1it}\beta + \theta v_{2it} + (c_{1i} - \theta a_{2i}) + e_{1it} \geq 0] \\ &= 1[\mathbf{x}_{1it}\beta + \theta v_{2it} + c_{0i} + e_{1it} \geq 0], \end{aligned} \tag{9}$$

where $t = 1, ..., T$, $\mathbf{x}_{1it} = (\mathbf{z}_{1it}, y_{2it}, y_{1i,t-1})$, $\beta = (\beta_1', \beta_2, \rho)'$, and $c_{0i} = c_{1i} - \theta a_{2i}$ is a composite unobserved effect. Since we assume that u_{2it} is normally distributed, we assume that $D(y_{2it}|\mathbf{z}_i, a_{2i})$ is normal and y_{2it} should have characteristics of a normal random variable. Thus, the assumption of normality of u_{2it} potentially limits the applicability of our approach, since it rules out endogenous regressors that are discrete or have severely limited support. In the application we present in Section 3 below, y_{2it} will be the share of registered long-term village residents who are employed as migrants outside the village and the support for this variables will be comfortably within the [0,1] interval. Thus, the above assumptions are plausible in our application.

Since the unobserved effect c_{0i} is present in equation (9), we should consider the relation between the unobserved effect c_{0i} and the explanatory variables in equation (9). 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 (9), we must take into account the relation between c_{0i} and v_{2it} in different time periods. Similar to our modeling of the unobserved heterogeneity in the reduced form equation for y_{2it}, here we are also flexible when choosing the exact functional form for the unobserved heterogeneity c_{0i}. In particular, we can consider the two popular options we discussed above: the Mundlak and Chamberlain specifications with respect to the sequence of the reduced form composite errors in all time periods, $\mathbf{v}_{2i}$.

First, we use a "pure" random effects approach:

Assumption 5.A: Assume

$$c_{0i}|\mathbf{z}_i, y_{1i0}, \mathbf{v}_{2i} \sim \text{Normal}(\alpha_0 \overline{v}_{2i}, \sigma^2_{a_1}), \tag{10}$$

where $\mathbf{v}_{2i} = (v_{2i1}, v_{2i2}, ..., v_{2iT})$, and $\overline{v}_{2i} = \frac{1}{T}\sum_{t=1}^{T} v_{2it}$.

Note that under Assumption 5.A we can write $c_{0i} = \alpha_0 \overline{v}_{2i} + a_{1i}$, where $a_{1i}|\mathbf{z}_i, y_{1i0}, \mathbf{v}_{2i} \sim \text{Normal}(0, \sigma^2_{a_1})$. Under the "pure" random effects approach we assume that the composite unobserved effect, c_{0i}, is independent of the initial condition, y_{1i0} (as well as $\mathbf{z}_i$'s). Thus, we might find it reasonable to think that v_{2it}'s in different time periods have equal impacts on c_{0i}. In other words, since the initial condition y_{1i0} is independent of the unobservable c_{0i}, we might think that correlation between the unobservable c_{0i} and endogenous regressor y_{2it} is the same in all time periods. 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 1-4 and 5.A, we rewrite equation (9) as

$$y_{1it} = 1[\mathbf{x}_{1it}\beta + \theta v_{2it} + \alpha_0 \overline{v}_{2i} + a_{1i} + e_{1it} \geq 0], \tag{11}$$

which leads to s simple two-step estimation procedure for the model of our interest:

Procedure 2.1:

(i) Estimate the reduced form (6) for y_{2it} using pooled OLS. Obtain the residuals, $\hat{v}_{2it}$, for all (i, t) pairs, and calculate $\overline{\hat{v}}_{2i} = \frac{1}{T}\sum_{t=1}^{T} \hat{v}_{2it}$ for every i.

(ii) Estimate the probit y_{1it} on $\mathbf{x}_{1it}$, $\hat{v}_{2it}$, and $\overline{\hat{v}}_{2i}$ using the conditional MLE.

As Procedure 2.1 suggests, the consistent 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 explanatory variables along with $\mathbf{x}_{1it}$ and $\hat{v}_{2it}$. Since the second stage employs "generated" regressors involving the first-stage residuals $\hat{v}_{2it}$'s, the standard errors in the second stage should be adjusted for the first-stage estimation. While asymptotic standard errors can be obtained, bootstrapping provides a simple easily implementable alternative. We rely on bootstrapping in our application.

It is worth pointing out that our approach can be straightforwardly extended to models with additional endogenous covariates that are functions of y_{2it}. For example, we might be interested in a model that contains y^2_{2it} as an additional regressor. Alternatively, models with interaction terms between y_{2it} and some exogenous regressor(s) or between y_{2it} and the lagged dependent variable are also of empirical importance. These cases can be easily adopted under the above assumptions. Specifically, say, we study a model with the interaction term between y_{2it} and the lagged dependent variable. Then, we modify Procedure 2.1 in the following way. The first stage of Procedure 2.1 stays unchanged, while the second stage involves estimating the probit y_{1it} on $\mathbf{x}_{1it}$, $y_{2it} \cdot y_{1i,t-1}$, $\hat{v}_{2it}$, and $\overline{\hat{v}}_{2i}$ using the conditional MLE, where $\hat{v}_{2it}$ are the residuals from the first stage. In other words, the presence of the second endogenous regressor, which is a function of the first endogenous covariate, requires a minor modification of the second stage of our original two-step procedure when only one endogenous explanatory variable is involved. The model with the interaction term between the endogenous covariate and the lagged dependent variable is exactly the model we study in our application.

As we discussed earlier, however, the assumption of independence between the unobserved effect, 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 build on the dynamic "correlated" random effects model introduced by Wooldridge (2005). Instead of the conditional distribution of c_{0i} in Assumption 5.A, we now employ a "correlated" random effects approach:

Assumption 5.B: Assume

$$c_{0i}|\mathbf{z}_i, y_{1i0}, \mathbf{v}_{2i} \sim \text{Normal}(\mathbf{v}_{2i}\alpha_0 + \mathbf{z}_i\alpha_1 + \alpha_2 y_{1i0}, \sigma^2_{a_1}), \tag{12}$$

where $\mathbf{v}_{2i}$ and $\mathbf{z}_i$ are defined above.

Note that Assumption 5.B implies $c_{0i} = \mathbf{v}_{2i}\alpha_0 + \mathbf{z}_i\alpha_1 + \alpha_2 y_{1i0} + a_{1i}$, where $a_{1i}|\mathbf{z}_i, y_{1i0}, \mathbf{v}_{2i} \sim$ Normal$(0, \sigma^2_{a_1})$.

Since now 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 5.B extends 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 5.B 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 1-4 and 5.B, we rewrite equation (9) as

$$\begin{aligned} y_{1it} &= 1[\mathbf{x}_{1it}\beta + \theta v_{2it} + c_{0i} + e_{1it} \geq 0] \\ &= 1[\mathbf{x}_{1it}\beta + \theta v_{2it} + \mathbf{v}_{2i}\alpha_0 + \mathbf{z}_i\alpha_1 + \alpha_2 y_{1i0} + a_{1i} + e_{1it} \geq 0], \end{aligned} \tag{13}$$

which implies the following two-step estimation procedure for the model of our interest:

Procedure 2.2:

(i) Estimate the reduced form (6) for y_{2it} using pooled OLS. Obtain the residuals, $\hat{v}_{2it}$, for all (i, t) pairs.

(ii) Estimate the probit y_{1it} on $\mathbf{x}_{1it}$, $\hat{v}_{2it}$, $\hat{\mathbf{v}}_{2i}$, $\mathbf{z}_i$, and y_{0i1} using the conditional MLE.

Procedure 2.2 suggests that we can consistently 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 explanatory variables along with $\mathbf{x}_{1it}$ and $\hat{v}_{2it}$. Once again, the standard errors in the second stage should be adjusted for the first-stage estimation, and we choose bootstrapping to do so in our application.

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 $\text{Var}(e_{2it}) = \sigma^2_{e_2}$, and

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

which is more than 0, unless either $\pi = 0$ or $\theta = 0$. Clearly, equation (8) is no longer appropriate and must be modified.

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,

$$\mathbf{\Omega} \equiv \mathrm{E}(\mathbf{v}_{2i}\mathbf{v}_{2i}') = \sigma_{a_2}^2 \mathbf{j}_T \mathbf{j}_T' + \sigma_2^2 \begin{pmatrix} 1 & \pi & \pi^2 & \cdots & \pi^{T-2} & \pi^{T-1} \\ \pi & 1 & \pi & \cdots & \pi^{T-3} & \pi^{T-2} \\ \pi^2 & \pi & 1 & \cdots & \pi^{T-4} & \pi^{T-3} \\ \vdots & & & \ddots & & \vdots \\ \pi^{T-2} & \pi^{T-3} & \pi^{T-4} & \cdots & 1 & \pi \\ \pi^{T-1} & \pi^{T-2} & \pi^{T-3} & \cdots & \pi & 1 \end{pmatrix}, \tag{14}$$

where $\mathbf{j}_T$ is a $T \times 1$ vector of ones, and $\sigma_2^2 = \frac{\sigma_{e_2}^2}{1-\pi^2}$. We can obtain consistent estimates of the parameters in (14), and use them to transform v_{2it} to v_{2it}^*, which we refer to as a first-stage error free of serial correlation. In our application we obtain the transformed residuals v_{2it}^* using the Cochrane-Orcutt transformation after the OLS estimation of the slope coefficient π from an AR(1) regression of the first-stage residuals v_{2it}.[8] Once we have first-stage errors free of serial correlation, v_{2it}^*, we use the transformation $u_{2it}^* = v_{2it}^* - a_{2i}$ to adjust equation (8). We can then assume that under joint normality of (u_{1it}, u_{2it}^*),

$$\begin{aligned} u_{1it} &= u_{2it}^*\theta + e_{1it} \\ &= \theta(v_{2it}^* - a_{2i}) + e_{1it}, \end{aligned} \tag{15}$$

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 (15) guarantees that e_{1it} will not be serially correlated. We are then able to write

$$\begin{aligned} y_{1it} &= 1[\mathbf{x}_{1it}\beta + c_{1i} + \theta v_{2it}^* - \theta a_{2i} + e_{1it} \geq 0] \\ &= 1[\mathbf{x}_{1it}\beta + \theta v_{2it}^* + (c_{1i} - \theta a_{2i}) + e_{1it} \geq 0 \\ &= 1[\mathbf{x}_{1it}\beta + \theta v_{2it}^* + c_{0i} + e_{1it} \geq 0], \end{aligned} \tag{16}$$

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

[8]An alternative method for estimating π, $\sigma_{a_2}^2$, $\sigma_{e_2}^2$, and σ_2^2 is the minimum distance estimator, described in detail by Chamberlain (1984). Cappellari (1999) has developed a code that conveniently implements this method in Stata.

Based on equation (16), it is straightforward to adjust the two-step estimating procedures discussed in Section 2.2 to account for the presence of the serial correlation in u_{2it}. In particular, under the "correlated" random effects Assumption 4.B, equation (16) can be written as

$$\begin{aligned} y_{1it} &= 1[\mathbf{x}_{1it}\beta + \theta v_{2it}^{*} + c_{0i} + e_{1it} \geq 0] \\ &= 1[\mathbf{x}_{1it}\beta + \theta v_{2it}^{*} + \mathbf{v}_{2i}\alpha_0 + \mathbf{z}_i\alpha_1 + \alpha_2 y_{1i0} + a_{1i} + e_{1it} \geq 0]. \end{aligned} \tag{17}$$

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 $\mathbf{x}_{1it}$.

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 dependent variable on its current value. We follow an approach proposed by Wooldridge (2010) to calculate the APEs after our two-step estimation procedure. The APEs, evaluated at any level of covariates $\mathbf{x}_{1t}$, can be calculated by taking either differences or derivatives of

$$\mathrm{E}[\Phi(\mathbf{x}_{1t}\beta + \theta v_{2it} + \mathbf{v}_{2i}\alpha_0 + \mathbf{z}_i\alpha_1 + \alpha_2 y_{1i0})],\ t = 1, ..., T, \tag{18}$$

with respect to the elements of $\mathbf{x}_{1t}$. Here, variables with a subscript i are random and all others are fixed.

In order to obtain estimates of the parameter values in (18), we appeal to a standard uniform weak law of large numbers argument.[9] For any given value of $\mathbf{x}_{1t}$ ($\mathbf{x}_1^0$), a consistent estimator for expression (18) can be obtained by replacing unknown parameters by consistent estimators:

$$N^{-1}\sum_{i=1}^{N}\Phi(\mathbf{x}_1^0\hat{\beta}_* + \hat{\theta}_*\hat{v}_{2it} + \hat{\mathbf{v}}_{2i}\hat{\alpha}_{0*} + \mathbf{z}_i\hat{\alpha}_{1*} + \hat{\alpha}_{2*}y_{1i0}), \tag{19}$$

[9] See Wooldridge (2010) for details.

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 $\hat{\sigma}^2 = (\widehat{\sigma^2_{e_1} + \sigma^2_{a_1}})^{-1/2}$, 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. In our application we employ bootstrapping. For each bootstrap sample, we execute the two-step procedure and compute the APEs. Then, we repeat the process to obtain a bootstrap standard error for the APEs.[10] We can compute either changes or derivatives of equation (19) 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

$$\mathrm{E}[\Phi(\mathbf{x}_{1t}\beta + \theta v^*_{2it} + \mathbf{v}_{2i}\alpha_0 + \mathbf{z}_i\alpha_1 + \alpha_2 y_{1i0})], \tag{20}$$

where $t = 1, ..., T$. For any given value of $\mathbf{x}_{1t}$ ($\mathbf{x}^0_1$), a consistent estimator of expression (20) is obtained by replacing unknown parameters by consistent estimators:

$$N^{-1}\sum_{i=1}^{N}\Phi(\mathbf{x}^0_1\hat{\beta}_* + \hat{\theta}_*\hat{v}^*_{2it} + \hat{\mathbf{v}}_{2i}\hat{\alpha}_{0*} + \mathbf{z}_i\hat{\alpha}_{1*} + \hat{\alpha}_{2*}y_{1i0}), \tag{21}$$

where $t = 1, ..., T$, $\hat{v}^*_{2it}$ is a first stage residual cleaned of serial correlation, where the $*$ subscript denotes multiplication by $\hat{\sigma}^2 = (\widehat{\sigma^2_{e_1} + \sigma^2_{a_1}})^{-1/2}$, and $\hat{\beta}$, $\hat{\theta}$, $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\sigma}^2$ are the conditional MLEs. One may then compute derivatives of equation (21) with respect to $\mathbf{x}_{1t}$ to obtain the APEs of interest.

The choice of $\mathbf{x}_{1it}$ is quite arbitrary, and ultimately depends on the empirical interests. Since we observe the covariates, we usually have a good idea of interesting values of $\mathbf{x}_{1t}$ to use for calculation of the APEs. Alternatively, we can average the APEs across the sample $\{\mathbf{x}_{1it} : i = 1, ..., N\}$ on $\mathbf{x}_{1t}$. In our application we follow Papke and Wooldridge (2008) and

[10] See Chapter 12.8.2 in Wooldridge (2010) for more details.

average the partial effects across both time and the cross section.

3 Migrant Labor Markets and Poverty Persistence in Rural China

Before applying the dynamic binary response model developed above to an analysis of how migration affects poverty status in rural China, we first briefly review the history of rural-urban migration in China and review other evidence on the impacts of migration in migrant sending communities, and introduce the data source that will be used for our analysis. 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 describe our approach to identifying the migrant networks, which affect the cost of finding migrant employment for village residents.

3.1 Rural-Urban Migration in China

Rapid growth in the volume of rural migrants moving to urban areas for work during the 1990s signalled a fundamental change in China's labor market. 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, 2008).

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, 2007), and is associated with higher levels of local investment in productive activities (Zhao, 2003).

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.[11]

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 enumerates detailed household-level information on incomes and expenditures, education, labor supply, asset ownership, land holdings, savings, formal and informal access to credit, and remittances.[12] In common with the National Bureau of Statistics (NBS) *Rural Household Survey,* respondent households keep daily diaries of

[11] 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.

[12] 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.

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.[13] 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.[14] Poorer illiterate households 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 Migration and Poverty

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

[13] 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).

[14] 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).

the village. In our analysis, we consider all registered residents working outside their home county to be migrants.[15] Both the tremendous increase in migration from 1987 onward and heterogeneity across villages are evident in Figure 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 of home villages by 2003.

In other research using this data source, de Brauw and Giles (2008) 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 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 and this may have 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 labor supply to productive activities and the land per capita cultivated by their households than do richer households when out-migration increases. This raises the prospect that migration may be causally related to poverty reduction within rural communities as well.

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 2). Nonlinearities in the bivariate relationship are evident 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 poverty within villages is likely to require an analytical approach that eliminates endogeneity bias due to simultaneity and potential sources of unobserved heterogeneity.

[15]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.

In the empirical application of our discrete binary response model below, we examine whether out-migration from villages is associated with reductions in the probability that household consumption falls below the poverty line in rural China. Researchers in the poverty literature have questioned the appropriateness of running *poverty regressions* of this type because the analyst discards richer information provided by the complete distribution of consumption in favor of a binary variable. Not only is information discarded, but one also introduces distributional assumptions associated with estimating a binary response model.[16] While recognizing these concerns, our examination of poverty persistence using a dynamic binary response model is useful for two reasons: first, it helps to highlight the strengths of our approach to estimating dynamic binary response models. When analysts only have access to administrative data on such outcomes as receipt of unemployment benefits or welfare participation, then analysis of persistence in participation or receipt of support is important and requires a binary outcome model (e.g., Adren, 2007; Bane and Ellwood, 1986). While our analysis discards some information, we do this to provide evidence on the appropriateness of our approach to estimating dynamic binary response models. Second, use of a dynamic binary response model focusses attention on whether or not a household passes a specific point in the distribution of consumption, or alternatively income (e.g., Biewen, 2009; Hansen and Wahlberg, 2009). By doing this, we address a policy-relevant question of how a treatment, in this case increased migration, affects the likelihood that poor households will remain poor and the likelihood that non-poor households will fall into poverty. 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.

[16] See Ravallion (1996) for a useful exposition of these issues.

3.4 Estimating the Impact of Migrant Labor Markets on Poverty Persistence

We will estimate the dynamic binary outcome model for the likelihood that a household i from village j falls below the poverty line at time t:

$$pov_{it} = 1[\beta_1 pov_{it-1} + \beta_2(M^i_{jt} * pov_{it-1}) + \beta_3 M^i_{jt} + \mathbf{X}'_{it}\alpha_1 + \alpha_2 lpc_{it} + \mathbf{D}_t + u_i + \mathbf{v}_j * t_t + \varepsilon_{it}], \quad (22)$$

where pov_{it} is a binary indicator for whether the household is poor in year t. Current poverty status 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, and thus introduce endogeneity concerns. Since village fixed effects are at a higher level of aggregation than household fixed effects, when controlling for household fixed effects, we also effectively control for fixed effects associated with the village in which households are located. Further, we will be concerned that there may be village-specific trends, $\mathbf{v}_j * t_t$, related to underlying endowments and initial conditions that also have an impact on household poverty status. The error term, ε_{it}, may be serially correlated, and we are 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 source of endogeneity.

From the model specified in (22), 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^i_{jt}, allows us to determine the impact of the migrant network on the probability that a household will fall into poverty.

The specification shown in (22) may have additional sources of endogeneity if we believe that household demographic and human capital variables in $\mathbf{X}_{it}$, or land per capita, lpc_{it}, 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 with such events as marriage 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.[17] We thus use the period $t-2$ value of land per capita and estimate:

$$pov_{it} = 1[\beta_1 pov_{it-1} + \beta_2(M^i_{jt} * pov_{it-1}) + \beta_3 M^i_{jt} + \mathbf{X}'_{it}\alpha_1 + \alpha_2 lpc_{it-2} + \mathbf{D}_t + u_i + \mathbf{v}_j * t_t + \varepsilon_{it}] \quad (23)$$

One remaining issue is that we do not perfectly observe the network M^i_{jt} through which household i may use for job referrals. Instead, we observe the share of registered long-term 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:

$$pov_{it} = 1[\beta_1 pov_{it-1} + \beta_2(M_{jt} * pov_{it-1}) + \beta_3 M_{jt} + \mathbf{X}'_{it}\alpha_1 + \alpha_2 lpc_{it-2} + \mathbf{D}_t + u_i + \mathbf{v}_j * t_t + \varepsilon_{it}] \quad (24)$$

[17] 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}.

In our identification strategy below, we will instrument the endogenous share 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.

3.5 Identifiying the Migrant Network

To identify 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 consumption growth. 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 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 2, we show a lowess plot of the relationship between years since IDs were distributed and the 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 experimented 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. 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 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 two earlier papers, de Brauw and Giles (2008a and 2008b) 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 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 identify the village migrant network by further interacting the quartic in years-since IDs with land per capita held by households in period $t-2$. 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.[18] 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.

As period $t-2$ lagged land per capita appears as an exogenous regressor and is also interacted with the quartic in years since IDs were distributed in the first stage, our estimation approach must also eliminate bias introduced through likely serial correlation of the error term in the first stage regression. To this end, it is important to note that our two-step

[18]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.

estimation procedure developed in Section 2 above allows for serial correlation of first-stage errors.

4 Results

Before estimating equation (24), we establish that our instruments are significantly related to the migrant share of the village 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) of Table 2 for odd years from 1989 till 2001.[19] 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 share of migrants with an F-statistic of 44.62 for the 1989 to 2001 sample.

We next proceed to estimate model (20), but first treat migration as exogenous and show results for both linear probability and probit implementations in Table 3. In all four specifications, we observe a positive association between migrant share of the village and probability that a household is below the poverty line, and this reflects the response of households to short-term shocks and the simultaneity between short-term migration and consumption decisions. Given the descriptive evidence shown in Figure 2, it is unsurprising to find that short-term increases in the poverty headcount will be correlated with year to year changes in the share of migrants from the village. The positive relationship suggests that migration is truly endogenous and suggests the need for an estimation strategy that allows for identification in a dynamic binary response model where there are endogenous regressors. When we introduce the years-since IDs instrument, which is shown elsewhere to

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

be unrelated to short-term fluctuations in the local economy (de Brauw and Giles, 2008a and 2008b), we can identify the longer term relationship between growth of the migrant labor market and the probability that a household will fall below the poverty line.

In Table 4, we report the control function (CF) estimation results based on the"pure" random effects and "correlated" random effects approaches.[20] For the purposes of comparison, we also estimate model (20) using a naive linear probability model (LPM). As one might expect, the coefficients on lagged poverty status are significant and positive, indicating a strong persistence in poverty status, both in the pure random effects approach shown in columns (1) and (3) and in the correlated random effects models shown in columns (2) and (4). The decline in the value of the coefficient on lagged poverty status between pure and correlated random effects models, (1) and (2) for linear probability models and (3) and (4) for the dynamic probit models, suggests that unobserved heterogeneity associated with poverty status introduces considerable upward bias in estimates of poverty persistence. Estimates of poverty persistence using either a dynamic linear probability model or the dynamic probit would lead the researcher to overstate the importance of chronic, persistent poverty. The significant coefficient on the initial value of poverty status in the correlated random effects models suggests a substantial correlation between unobserved effects and the initial condition.

Once we instrument for migrant share of the registered village population, and thus control for simultaneity bias introduced through shocks to the local economy, we find that the migrant labor market is negatively associated with the probability of falling into poverty. Moreover, the coefficient on the interaction of village migrant share and lagged poverty status suggests that the magnitude of the effect of migration on poverty reduction is greater among households who were poor in the previous period, and thus migration reduces poverty persistence even more than it reduces the likelihood that the non-poor will fall into poverty. This result is consistent with de Brauw and Giles (2008a), who find that in a linear panel data framework, households with lower levels of prior consumption tend to experience more

[20]When applying the CF approach, we detect serial correlation in the first-stage residuals (p-value = 0.000). Thus, on the second stage, we use the first-stage residuals free of serial correlation. To obtain the latter, we first estimate the slope coefficient from an AR(1) regression of the first-stage residuals using OLS. Second, knowing the consistent slope estimate, we employ the Cochrane-Orcutt transformation to get the residuals free of serial correlation.

rapid consumption growth with increased out-migration from rural villages.[21]

In order to examine the effect of migration on poverty persistence, we calculate the average partial effects (APEs) and show the estimates in Table 5. Mostly, we report the APEs averaged across both the cross section of the covariates and time. However, due to the presence of the interaction term, the calculation of two APEs – the APE of the migrant share and the APE of the lagged poverty status – requires extra care. Specifically, we obtain the APE of the migrant share for the following three possibilities: (1) when the lagged poverty status equals zero for all households; (2) when the lagged poverty status equals one for all households; and (3) when we average the APE of the migrant share across the observed values of the lagged poverty status in our sample. Rows two-four of Table 5 contain these results. Further, the APE of the lagged poverty status reported in the first row of Table 5 is averaged across the observed values of the migrant share in the sample.

The APEs calculated using the correlated random effects dynamic probit approach (models 3 and 4) are generally smaller than those calculated using the linear probability model (models 1 and 2). The naive LPM approach, which is often preferred as a means of avoiding dynamic nonlinear models, will lead us to conclude that migration has a more pronounced impact on poverty reduction than one finds using the correlated random effects probit model. Again the consequences of ignoring unobserved heterogeneity in the dynamic binary response model are of considerable interest. Failure to control for unobserved heterogeneity in the pure random effects model would lead us to overstate the effects of previous period poverty status on current poverty and understate the effect of the migrant labor market in contributing to reductions in the probability that a household would fall below the poverty line. For those households living above the poverty line, the correlated random effects CF estimate of the APE (model 4) suggests that a one percentage point increase in the share of village residents working as migrants would reduce the probability of falling into poverty by about 3.2 percentage points. For those already below the poverty line, the correlated random effects

[21] We employ the Hausman test for endogeneity to formally assess the need to control for endogeneity of migration share. The t-statistic for the significance of the first-stage residuals in the pure RE probit model is 3.08 with p-value of 0.002, which suggests there is enough evidence to reject the null hypothesis that the share of village out-migrants is exogenous. For the correlated RE probit, the t-statistic for the first-stage residuals is 2.93 with p-value of 0.003. Thus, for the correlated RE model, we also reject the null hypothesis that the share of migrants is exogenous.

CF estimate of the APE shows that a one percentage point increase in the village migrant share will reduce the probability of remaining in poverty by 3.5 percentage points.

5 Conclusions

In this paper, we have developed a dynamic binary response panel data model that allows for an endogenous regressor. The control function approach which we implement is of particular value for settings in which one wants to estimate the effects of a treatment which is also endogenous. Our empirical example demonstrates that alleviating an omitted variables bias can lead to estimated effects that are larger in absolute value when we allow for the correlation between unobserved heterogeneity, initial conditions and exogenous variables.

We apply the model to examine the impact of rural-urban migration on the likelihood that households in rural China fall below the poverty line. Our application demonstrates 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. From this specific application, we show that failing to adequately control for unobserved heterogeneity in non-linear dynamic panel data models will introduce substantial bias to parameter estimates. In particular, failure to control for unobserved heterogeneity would lead us to overstate the persistence of poverty and to understate the role that migration plays in poverty reduction.

Apart from analyzing the effects of migration on a binary outcome, our application suggests that there may be many other settings in which the correlated random effects control function approach may improve an existing analytical approaches. In any analysis aiming to examine how a new program affects persistence of a state, one may be concerned that unobserved heterogeneity will lead to upward bias in estimates of the effect of the initial state. Moreover, as program participation, or take-up, may be endogenous, the analyst will need to worry about this source of bias as well. The empirical strategy developed in Section 2 offers a parametric solution to the more general problem of identifying the impact of an endogenous treatment in a dynamic binary response model.

6 References

Adren, T. 2007. “The Persistence of Welfare Partipation,” IZA Discussion Paper 3100, October 2007.

Bane, M.J. and D.T. Ellwood. 1986. “Slipping into and out of Poverty: The Dynamics of Spells,” *The 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.

Biewen, M. 2009. “Measuring State Dependence in Individual Poverty Status: Are There Feedback Effects to Employment Decisions and Household Composition?” *Journal of Applied Econometrics* 24(7): 1095-1116.

Cai, F., A. Park and Y. Zhao. 2008. “The Chinese Labor Market,” chapter prepared for *China’s Great Economic Transition*, Loren Brand and Thomas Rawski (eds), Cambridge University Press.

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

Carrington, W., E. Detragiache and T. Vishnawath. 1996. “Migration with Endogenous Moving Costs,” *American Economic Review* 86(4): 909-930.

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, Kam Wing and Li 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.

Chiappori, P., and B. Salanie. 2000. “Testing for Asymmetric Information in Insurance Markets,” *Journal of Political Economy* 108, 56-78.

de Brauw, A. and J. Giles. 2008a. “Migrant Opportunity and the Educational Attainment of Youth in Rural China,” Policy Research Working Paper 4585, The World Bank (February 2008).

de Brauw, A. and J. Giles. 2008b. “Migrant Labor Markets and the Welfare of Rural Households in the Developing World: Evidence from China,” Policy Research Working Paper 4526, The World Bank (April 2008).

Devicienti, D. and A. Poggi. 2007. “Poverty and social exclusion: two sides of the same coin or dynamically interrelated processes?,” LABORatorio R. Revelli Working Papers Series 62, LABORatorio R. Revelli, Centre for Employment Studies.

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

Giles, J. 2006. “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. 2007. “Precautionary Behavior, Migrant Networks and Household Consumption Decisions: An Empirical Analysis Using Household Panel Data from Rural

China," *The Review of Economics and Statistics*, 89(3): 534-551.

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.

Hansen, J. and R. Wahlberg. 2009. "Poverty Persistence in Sweden," *Review of the Economics of the Household*, 7(2), 105-132.

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. and R.J. Willis. 1977. "A Beta-logistic Model for the Analysis of Sequential Labor Force Participation by Married Women," *Journal of Political Economy,* 85, 27-58.

Honoré, B.E. and E. Kyriazidou. 2000. "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica* 68, 839-874.

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

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 Yaohui Zhao and Loraine West (eds) Rural Labor Flows in China, Berkeley: University of California Press, 251-277.

Montgomery, J.D. 1991. "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis," *American Economic Review* 81(5): 1407-18.

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

Munshi, K. 2003. "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market," *Quarterly Journal of Economics* 118(2): 549-99.

Papke, L.E. and J.M. Wooldridge. 2008. "Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates," *Journal of Econometrics* 145: 121-33

Ravallion, M. 1996. "Issues in Measuring and Modelling Poverty," *The Economic Journal* 106: 1328-1343.

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," *The China Quarterly* 158: 367-393.

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.

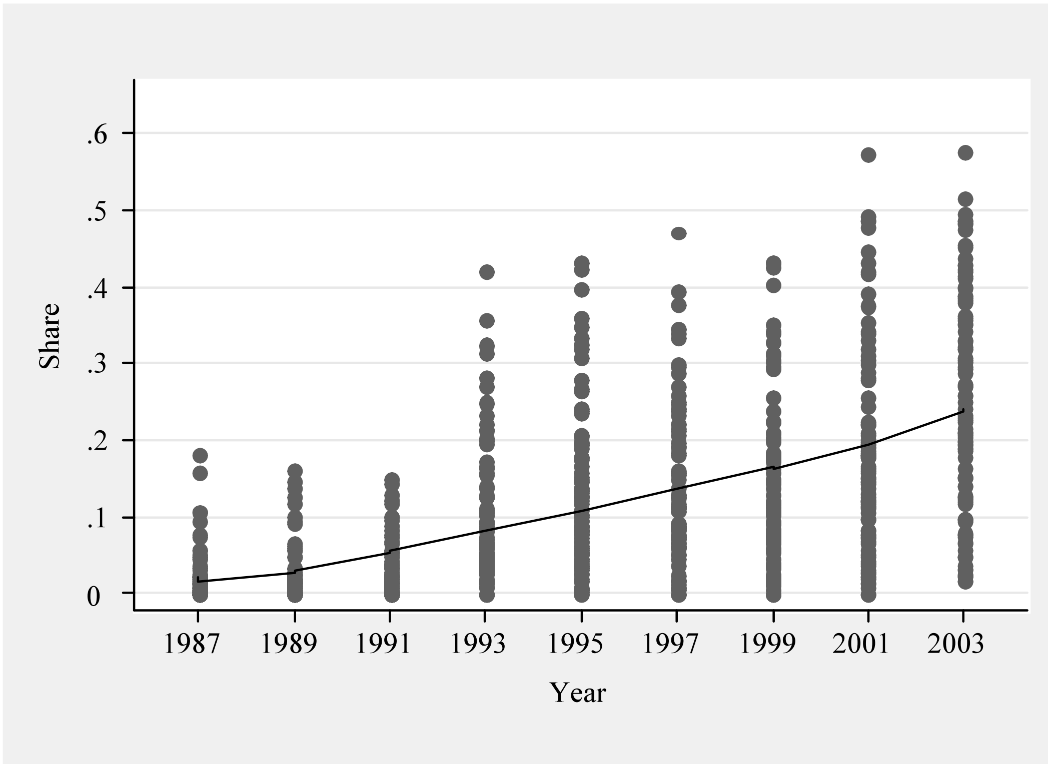
Wooldridge, J.M. 2000. “A Framework for Estimating Dynamic, Unobserved Effects Panel Data Models with Possible Feedback to Future Explanatory Variables,” *Economics Letters* 68, 245-250.

Wooldridge, J.M. 2010. Econometric Analysis of Cross Section and Panel Data. Second Edition. MIT Press, Cambridge, MA.

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

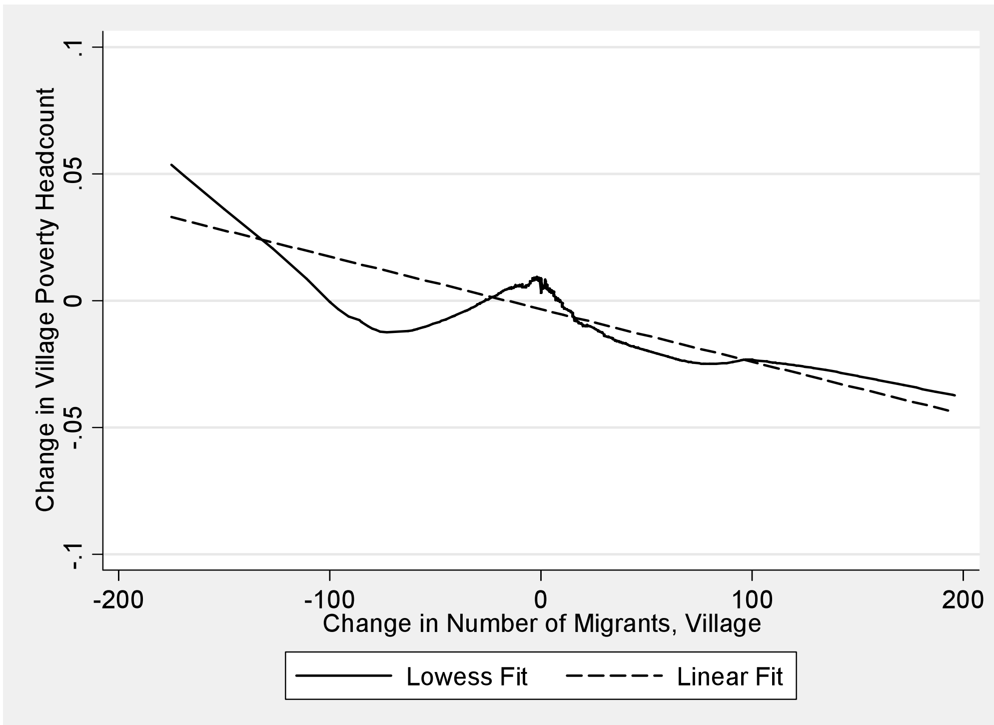
Zhao, Y. 2003. “The Role of Migrant Networks in Labor Migration: The Case of China,” *Contemporary Economic Policy* 21(4): 500-511.

Figure 1
Share of Village Labor Force Employed
as Migrants By Year



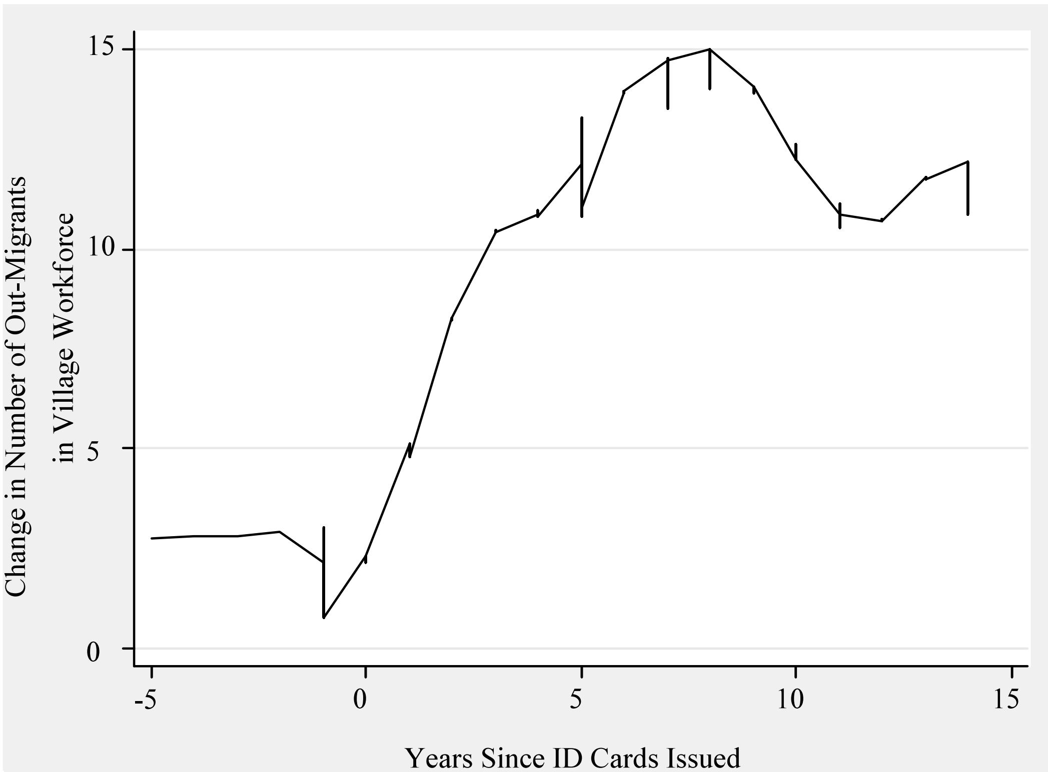
Source: RCRE Village Surveys 1987 to 2003.

Figure 2
Change in Poverty Headcount Versus Change in Number of Migrants



Source: RCRE Village and Household Surveys, 1987 to 2003.

Figure 3
Change in Out-Migrants in Village Labor Force
Versus Years-Since-IDs were Distributed



Source: 2004 RCRE Supplemental Survey on Land and Village Governance.

Table 1. Household and Village Characteristics

		Odd Years from 1989 to 2001			
		Obs.	Full Sample	Obs.	Balanced Sample
Household Poverty Status	mean	42453	0.20	26159	0.20
	st. dev.		0.40		0.40
Household Income per Capita	mean	42447	721.4	26159	685.8
	st. dev.		649.3		537.5
Household Consumption per Capita	mean	42453	521.9	26159	499.1
	st. dev.		376.1		332.6
Number of Household Members	mean	42491	4.1	26159	4.2
	st. dev.		1.5		1.4
Number of Prime Age Household Laborers	mean	42491	2.5	26159	2.6
	st. dev.		1.1		1.0
Household Land per Capita	mean	42453	1.4	26159	1.4
	st. dev.		1.2		1.1
Household Average Years of Education	mean	41658	6.2	26156	6.3
	st. dev.		2.6		2.5
Household Share of Females	mean	41659	0.45	26156	0.45
	st. dev.		0.21		0.20
Share of Migrants from the Village	mean	42491	0.06	26159	0.06
	st. dev.		0.06		0.06
Year of ID Distribution in a Village	mean	41814	1988.0	26159	1988.0
	st. dev.		2.1		2.1
Years Since ID was Issued in a Village	mean	41814	6.7	26159	7.0
	st. dev.		4.5		4.5

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

Table 2. What Factors Determine the Size of the Village Migrant Network?
First-Stage Regressions

	Dependent Variable: Village Migrant Share		
	Odd Years from 1989 to 2001		
Model	(1)	(2)	(3)
Household Population	-0.0003	-0.0003	-0.0003
	(0.0003)	(0.0003)	(0.0003)
Number of Working Age Laborers in the Household	0.0002	0.0002	0.0003
	(0.0004)	(0.0004)	(0.0004)
Land Per Capita $_{t-2}$	-0.0040***	-0.0036***	-0.0017***
	(0.0004)	(0.0005)	(0.0006)
Average Years of Education	-0.0006***	-0.0006***	-0.0007***
	(0.0001)	(0.0001)	(0.0001)
Female Share of the Household	-0.0011	-0.0011	-0.0011
	(0.0015)	(0.0015)	(0.0015)
(Years-Since-IDs Available) * (Land Per Capita $_{t-2}$)	0.0008***	0.0006***	-0.0018***
	(0.0001)	(0.0002)	(0.0004)
(Years-Since-IDs Available)^2 * (Land Per Capita $_{t-2}$)	-0.0000***	-0.0000	0.0007***
	(0.0000)	(0.0000)	(0.00009)
(Years-Since-IDs Available)^3 * (Land Per Capita $_{t-2}$)		-0.0000	-0.0001***
		(0.0000)	(0.0000)
(Years-Since-IDs Available)^4 * (Land Per Capita $_{t-2}$)			0.0000***
			(0.0000)
Observations	22422	22422	22422
R-squared	0.79	0.79	0.79
F-Statistic on IVs with Averages	62.51	58.11	44.62
F-Statistic on IVs w/o Averages	46.04	31.60	32.64
Partial R^2, IVs with Averages	0.005	0.005	0.007
Partial R^2, IVs w/o Averages	0.001	0.001	0.003

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, year dummies, and interactions between village dummies and time trend.

Table 3. Estimating Determinants of Poverty Status with Migrant Share Treated as Exogenous

	Dependent Variable: Poverty Status			
	Linear Probability Model		Probit	
	Pure RE	Correlated RE	Pure RE	Correlated RE
Model	(1)	(2)	(3)	(4)
Lag Poverty Status	0.390***	0.339***	1.045***	0.794***
	(0.012)	(0.013)	(0.041)	(0.048)
Village Migrant Share Interacted with and Lag Poverty Status	-0.974***	-0.767***	-2.356***	-1.654***
	(0.134)	(0.132)	(0.465)	(0.484)
Village Migrant Share	0.285***	0.221***	1.476***	1.233**
	(0.076)	(0.074)	(0.519)	(0.543)
Number of Household Members	0.046***	0.057***	0.257***	0.371***
	(0.002)	(0.004)	(0.014)	(0.021)
Number of Prime Age Household Laborers	-0.023***	-0.024***	-0.127**	-0.147***
	(0.003)	(0.004)	(0.017)	(0.023)
Second Lag of Land per Capita	-0.002	0.000	-0.034*	-0.027
	(0.003)	(0.004)	(0.018)	(0.031)
Average Years of Education	-0.007***	0.000	-0.042***	-0.007
	(0.001)	(0.001)	(0.006)	(0.009)
Share of Females	-0.053***	-0.006	-0.293***	-0.028
	(0.012)	(0.015)	(0.068)	(0.093)
Dependent Variable in 1989		0.090***		0.508***
		(0.009)		(0.047)
Observations	22422	22422	22422	22422
Number of households	3737	3737	3737	3737
R-Squared	0.35	0.36		

Notes: In parenthesis we show fully robust standard errors [*** p<0.01, ** p<0.05, * p<0.1]. All regressions include the explanatory variables in each year, year dummies, and interactions between village dummies and time trend.

Table 4. Estimating Determinants of Poverty Status with Endogenous Share of Migrants
Second-Stage Regressions

	Dependent Variable: Poverty Status			
	Linear Probability Model		Control Function	
	(1)	(2)	(3)	(4)
Model	Pure RE	Correlated RE	Pure RE	Correlated RE
Lag Poverty Status	0.391***	0.335***	1.046***	0.792***
	(0.013)	(0.012)	(0.054)	(0.052)
Village Migrant Share Interacted with and Lag Poverty Status	-0.994***	-0.784***	-2.443***	-1.779***
	(0.128)	(0.125)	(0.512)	(0.526)
Village Migrant Share	-2.628***	-3.955***	-12.201**	-18.896**
	(0.833)	(1.039)	(5.660)	(8.191)
Number of Household Members	0.047***	0.057***	0.261***	0.368***
	(0.003)	(0.004)	(0.020)	(0.028)
Number of Prime Age Household Laborers	-0.023***	-0.023***	-0.125***	-0.143***
	(0.003)	(0.005)	(0.022)	(0.031)
Second Lag of Land per Capita	-0.006*	-0.001	-0.050*	-0.036
	(0.003)	(0.005)	(0.026)	(0.043)
Average Years of Education	-0.009***	-0.002	-0.048***	-0.018
	(0.001)	(0.002)	(0.008)	(0.013)
Share of Females	-0.058***	-0.012	-0.312***	-0.061
	(0.013)	(0.019)	(0.098)	(0.130)
Dependent Variable in 1989		0.086***		0.497***
		(0.009)		(0.053)
Observations	22422	22422	22422	22422
Number of households	3737	3737	3737	3737
R-Squared	0.29	0.32		
Replications for Bootstrap Errors	100	100	100	100

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, year dummies, and interactions between village dummies and time trend. 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 5. Average Partial Effects of Determinants of Poverty Status
(Endogenous Share of Migrants)

	LPM		Control Function	
	Pure RE	Correlated RE	Pure RE	Correlated RE
Model	(1)	(2)	(3)	(4)
Lag Poverty Status	0.324***	0.282***	0.181***	0.125***
	(0.010)	(0.009)	(0.007)	(0.006)
Share of Migrants when Lag Poverty = 0	-2.628***	-3.955***	-2.092**	-3.156**
	(0.833)	(1.039)	(1.056)	(1.413)
Share of Migrants when Lag Poverty =1	-3.621***	-4.739***	-2.511**	-3.453**
	(0.834)	(1.028)	(1.044)	(1.394)
Share of Migrants (averaged)	-2.641***	-3.965***	-2.260**	-3.273**
	(0.833)	(1.038)	(1.050)	(1.405)
Number of Household Members	0.047***	0.057***	0.045***	0.061***
	(0.003)	(0.004)	(0.004)	(0.005)
Number of Prime Age Household Laborers	-0.023***	-0.023***	-0.022***	-0.024***
	(0.003)	(0.005)	(0.004)	(0.006)
Second Lag of Land per Capita	-0.006*	-0.001	-0.009*	-0.006
	(0.003)	(0.005)	(0.005)	(0.007)
Average Years of Education	-0.009***	-0.002	-0.008***	-0.003
	(0.001)	(0.002)	(0.002)	(0.002)
Share of Females	-0.058***	-0.012	-0.053***	-0.010
	(0.013)	(0.019)	(0.018)	(0.022)
Poverty Status in 1989		0.086***		0.092***
		(0.009)		(0.009)
Replications	100	100	100	100

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