Moment-based Specification Tests for Random Effects Dynamic Probit Models

Shigeki Kano[†]
Graduate School of Economics, Osaka Prefecture University
(kano@eco.osakafu-u.ac.jp)

Last revised: March 2015

Abstract

The maximum likelihood estimation of random effects dynamic probit models hinges on strong distributional assumptions. This paper offers simple diagnostic tests to detect possible specification errors for this class of models. We developed two moment-based test statistics by focusing on the marginal, period-by-period conditional mean function of binary outcomes. These statistics follow chi-square distributions asymptotically and are robust to within-individual dependences of panel data. Our Monte Carlo study revealed that they work well in a finite sample. Finally we applied the new testing procedure to an empirical analysis drawn from an existing paper.

Key words: Dynamic probit; conditional moment test; panel data.
JEL classification: C12; C23; C25.

[†]This research was supported by JSPS KAKENHI Grand Number 26380273.

1 Introduction

The identification of structural state dependence in the presence of unobserved heterogeneity has been a long-held issue in panel data analysis. Heckman (1981) advocates the random effects (RE) dynamic probit dedicated to this issue. More recently, Wooldridge (2005) shows how to conduct the maximum likelihood (ML) estimation of this class of models by commercial softwares without special programmings. Arulampalam and Stewart (2009), Akay (2012) and Rabe-Hesketh and Skrondal (2013) each demonstrate by the Monte Carlo studies that the Heckman and the Wooldridge ML estimators work equally well if the time period of panel is not too short.

As shown by Yatchew and Griliches (1985), the ML estimator for probit-type models is inconsistent if the models have non-spherical errors in latent variable regressions. Therefore, some authors regard as a limitation the distributional assumptions imposed by the RE approach. To leave the distribution of unobserved effects unspecified, Honoré and Kyriazidou (2000) offers a fixed effects dynamic logit. Alternatively, Christelis and Sanz-de Galdeano (2011) and Deza (2015), among others, adopt nonparametric, discrete mixture distributions to individual effects *à la* Heckman and Singer (1984). It is possible to run a version of heteroskedastic probit so that the second moment of individual effects depends on the regressors (Greene, 2011, pp.754).

Because the ML estimations of these generalized models are intractable, simple diagnostic tests are needed to detect specification errors without estimating them.[*1] Despite their usefulness, such tests have not drawn sufficient attentions in the literature. The aim of this paper is to propose convenient specification tests for the RE dyamic probit. Specifically, we put emphases on the cases where the unobserved effects are (i) wrongly dependent on the regressors, (ii) heteroskedastc, and (iii) non-normally distributed.

One may argue that Lagrange multiplier (LM) or score tests are natural options in the current problem because the RE dynamic probit is estimated by the ML based on the joint distribution of outcomes. However, the score function of the RE probit ML is in general formidably complicated (Greene and McKenzie, 2015). Given this difficulty, this paper employs the conditional moment (CM) tests addressed by Newey (1985), Tauchen (1985), and Pagan and Vella (1989), among others. It should be underscored that here we are interested in the conditional moment misspecification of a marginal outcome. Hence, although the correct specification of joint distribution is needed by the ML estimation, it is redundant at the stage of testing.

This paper offers two moment-based tests robust to within-individual correlations of arbitrary form, thereby taking the panel data structure into account. Importantly, one of these two

[*1] Along with the programing burden, Wooldridge (2005) points out the features of the fixed effects approach discouraging its application to empirical studies.

tests are performed by a regression package with a cluster-robust covariance matrix option. We not only derived the limiting distributions of them but also examined their small sample performances through Monte Carlo simulations. Finally our procedure was applied to test the specification of union membership dynamics model in Wooldridge (2005). We found that the non-normality of unobserved heterogeneity was not rejected in this model.

As related studies, Davidson and MacKinnon (1984) and Skeels and Vella (1999) highlight artificial regressions to conduct the LM/CM tests for the cross-sectional probit under the random sampling assumptions. Unlike these studies, we analyze panel data and so cannot take over their methodologies directly. Lechner (1995) and Bertschek and Lechner (1998) consider a robust statistical inference of static panel probit models from the standing points of quasi maximum likelihood (QML) and M-estimation. We should avoid these frameworks here because the ML estimation of RE dynamic probit needs correctly specified joint outcome distribution. Hyslop (1999) and Keane and Sauer (2010) concern the misspecified joint distribution of outcomes as a source of inconsistent estimation for the key structural parameters. In contrast to them, we pay special attentions to the specification errors at marginal distribution.

The remainder of this paper is organized as follows. Section 2 specifies the model and moment conditions to test and derives limiting distributions of the tests. Section 3 is devoted to Monte Carlo simulation verifying small sample performance of the tests and to an empirical application. Section 4 concludes the paper.

2 Specification Test for RE Dynamic Probit Model

2.1 Model specification and ML estimation

This subsection reviews a stylized model of RE dynamic probit and its ML estimation. Let y_{it} and $\boldsymbol{x}_{it}$ be a binary outcome and regressors for individual i ($i = 1, 2, \ldots, N$) in period t ($t = 0, 1, \ldots, T$) and stack them to vectors $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$ and $\boldsymbol{x}_i = (\boldsymbol{x}'_{i1}, \ldots, \boldsymbol{x}'_{iT})'$. The $(T + 1)$ periods observations, $(y_{i0}, \boldsymbol{y}_i, \boldsymbol{x}'_{i0}, \boldsymbol{x}'_i)'$, are assumed to be independent over the N individuals.[*2] We consider the case of short panel where $N \to \infty$ with T being constant.

The dynamics of binary y_{it} is modeled as latent variable regression

$$y^*_{it} = \delta y_{i,t-1} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + a^*_i + u_{it}, \quad y_{it} = 1\left[y^*_{it} > 0\right], \tag{1}$$

where $1[A]$ is an indicator function taking on unity if A is true and zero otherwise and $(\delta, \boldsymbol{\beta}')'$ are unknown parameters to be estimated. Last two terms a^*_i and u_{it} denote unobserved, time-

[*2] Regressor's initial value $\boldsymbol{x}_{i0}$ is supposed to be sampled along with y_{i0} but discarded in this paper as well as in many empirical studies. Rabe-Hesketh and Skrondal (2013) shed some light on the role of $\boldsymbol{x}_{i0}$ in the dynamic probit.

invariant characteristics of individual i (e.g., differentials in innate productivity or preference) and time-varying stochastic disturbances hitting i, respectively. Note that, because our observation starts at $t = 0$, no lagged dependent variable is available for y_{i0}'s regression and that y_{i0} may be correlated with a_i^* as well as outcome sequence $\boldsymbol{y}_i$. The endogeneity of $y_{i,t-1}$ in equation (1) arising from the possible correlation between y_{i0} and a_i^* is called the "initial condition problem" in the literature. The econometric approach originally proposed by Heckman (1981) is constructing a joint model of $(y_{i0}, \boldsymbol{y}_i)$ given $(\boldsymbol{x}_i, a_i^*)$.

Extending the correlated random effects (CRE) assumption of Chamberlain (1984, Chapter 3), Wooldridge (2005) proposes a much simpler solution than that in Heckman (1981) to circumvent the initial condition problem. Following Wooldridge, we formulate the parametric relationships among $(y_{i0}, \boldsymbol{x}_i, a_i^*)$ as

$$a_i^* = \delta_0 y_{i0} + \boldsymbol{x}_{i1}'\boldsymbol{\pi}_1 + \boldsymbol{x}_{i2}'\boldsymbol{\pi}_2 + \cdots + \boldsymbol{x}_{iT}'\boldsymbol{\pi}_T + \sigma_a a_i, \quad \sigma_a > 0. \tag{2}$$

Here last term a_i is a regression error of a_i^* given $(y_{i0}, \boldsymbol{x}_i')'$ and so independent from them by definition. Coefficient σ_a is an unknown scale parameter. Inserting equation (2) to structural form (1), we have reduced form

$$y_{it}^* = \delta y_{i,t-1} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \delta_0 y_{i0} + \boldsymbol{x}_{i1}'\boldsymbol{\pi}_1 + \boldsymbol{x}_{i2}'\boldsymbol{\pi}_2 + \cdots + \boldsymbol{x}_{iT}'\boldsymbol{\pi}_T + \sigma_a a_i + u_{it}. \tag{3}$$

Hence the heart of Chamberlain-Wooldridge CRE device boils down to include the initial outcome and regressors of every period into the set of controls as time-invariant regressors.[*3]

For later use, we adopt reparameterization

$$\boldsymbol{\pi}_1 = \frac{1}{T}\boldsymbol{\pi}, \quad \boldsymbol{\pi}_s = \frac{1}{T}\boldsymbol{\pi} + \boldsymbol{\lambda}_s, \quad s = 2, 3, \ldots, T, \tag{4}$$

to turn equation (3) into

$$\begin{aligned} y_{it}^* &= \delta y_{i,t-1} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \delta_0 y_{i0} + \bar{\boldsymbol{x}}_i'\boldsymbol{\pi} + \boldsymbol{x}_{i2}'\boldsymbol{\lambda}_2 + \cdots + \boldsymbol{x}_{iT}'\boldsymbol{\lambda}_T + \sigma_a a_i + u_{it} \\ &= \boldsymbol{z}_{it}'\boldsymbol{\theta} + \boldsymbol{w}_i'\boldsymbol{\gamma} + \sigma_a a_i + u_{it}, \end{aligned} \tag{5}$$

where $\boldsymbol{z}_{it} = (y_{i,t-1}, \boldsymbol{x}_{it}', y_{i0}, \bar{\boldsymbol{x}}_i')'$, $\boldsymbol{w}_i = (\boldsymbol{x}_{i2}', \boldsymbol{x}_{i3}', \ldots, \boldsymbol{x}_{iT}')'$, and $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ collect the corresponding sets of coefficients. Assuming identical independent normality of $(a_i, u_{i0}, u_{i1}, \ldots, u_{iT})$

[*3]Obviously, in the CRE approach, the structural coefficients of time-invariant regressors (e.g., race, sex, and education level) are not identified. This identification condition is similar to that in the within estimator of fixed effects linear regressions.

given $(\boldsymbol{z}_{it}, \boldsymbol{w}_i)$, we have the joint probability mass function of $\boldsymbol{y}_i$ given $(\boldsymbol{z}_{it}, \boldsymbol{w}_i)$ *and* a_i,

$$f(\boldsymbol{y}_i|\boldsymbol{z}_{it}, \boldsymbol{w}_i, a_i) = \prod_{t=1}^{T} \Phi\left[(2y_{it} - 1)(\boldsymbol{z}_{it}'\boldsymbol{\theta} + \boldsymbol{w}_i'\boldsymbol{\gamma} + \sigma_a a_i)\right], \tag{6}$$

which is a product of Bernelle distributions possessing time-homogeneous probabilities of occurrence. Here $\Phi(\cdot)$ is the standard normal cumulative distribution function. Let $\phi(a)$ be the normal density. Postulating the normality assumption on a_i, we close the log-likelihood function for individual i;

$$\log L_i(\boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma_a^2) = \log\left\{\int_{a=-\infty}^{\infty} \prod_{t=1}^{T} \Phi\left[(2y_{it} - 1)(\boldsymbol{z}_{it}'\boldsymbol{\theta} + \boldsymbol{w}_i'\boldsymbol{\gamma} + \sigma_a a)\right] \phi(a)\mathrm{d}a\right\}. \tag{7}$$

By maximizing $\sum_{i=1}^{N} \log L_i(\boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma_a^2)$ with respect to the parameters, we have the ML estimator of them. In running the ML estimation, statistical softwares apply Gaussian quadratures or Monte Carlo integration to evaluate the above expectation. (For example, "xtprobt" of Stata version 13.1 employs adoptive quadratures.)

2.2 Moment conditions to test

In the empirical applications of RE dynamic probit models, researchers often add within individual mean $\bar{\boldsymbol{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_{it}$ to the expanded set of control variables instead of high-dimensional $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}', \boldsymbol{x}_{i2}', \ldots, \boldsymbol{x}_{iT}')'$. See, for example, Akay (2012) and Rabe-Hesketh and Skrondal (2013) for this treatment. It is obvious from equation (5) that this common practice is equivalent to impose restriction $\boldsymbol{\gamma} = \mathbf{0}$ on the general model. Let us consider first the hypothesis testing for exclusion restriction

$$\mathrm{H}_0 : \boldsymbol{\gamma} = \mathbf{0}, \quad \mathrm{H}_1 : \boldsymbol{\gamma} \neq \mathbf{0} \tag{8}$$

or equivalently the significance of omitted variables $\boldsymbol{w}_i = (\boldsymbol{x}_{i2}', \boldsymbol{x}_{i3}', \ldots, \boldsymbol{x}_{iT}')'$. We will generalize this problem later. In the case where the dimension of time-varying variables $\boldsymbol{x}_{it}$ is high, Wooldridge (2005) procedure can blow up the number of nuisance parameters in $\boldsymbol{\gamma}$. So we may want to avoid estimating the model under the alternative hypothesis.

Because the estimation is conducted based on the full-likelihood given in equation (7), it seems natural to pursuit LM or score tests for testing the hypothesis. However, the score function of the current model, $\nabla_{\boldsymbol{\gamma}} \log L_i(\boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma_a^2)$, does not have a tractable form. See Greene and McKenzie (2015) for the case of RE static probit. For this reason, this paper offers alternative, convenient ways to test the specifications of the RE dynamic probit. The point is that, although the joint modeling of outcomes $\boldsymbol{y}_i$ is unavoidable for archiving the consistent estimation of key parameters, marginal distributions suffice to test for the misspecification concreting marginal means. By focusing on the margin, we can apply the conditional moments (CM) tests argued by Newey (1985), Tauchen (1985), and Pagan and Vella (1989) to the current problem.

The moment condition implied by the null hypothesis of equation (8), which plays the key role in the CM tests, is derived as follows. Wooldridge (2005) shows that the conditional expectation of y_{it} on $(\boldsymbol{z}_{it}, \boldsymbol{w}_i, a_i)$ is expressed as

$$p_{it} = \mathrm{E}(y_{it}|\boldsymbol{z}_{it}, \boldsymbol{w}_i, a_i) = \Phi\left[\mu_{it}(\boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma_a^2)\right], \quad \mu_{it}(\boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma_a^2) = \frac{\boldsymbol{z}_{it}'\boldsymbol{\theta} + \boldsymbol{w}_i'\boldsymbol{\gamma}}{\sqrt{1+\sigma_a^2}}. \tag{9}$$

Accordingly, the error form of binary y_{it} is written as

$$y_{it} = p_{it} + e_{it}, \quad \mathrm{E}(e_{it}|\boldsymbol{z}_{it}, \boldsymbol{w}_i, a_i) = 0, \quad \mathrm{Var}(e_{it}|\boldsymbol{z}_{it}, \boldsymbol{w}_i, a_i) = p_{it}(1-p_{it}). \tag{10}$$

Define $\Phi_{0,it} = \Phi\left[\mu_{it}(\boldsymbol{\theta}, \mathbf{0}, \sigma_a^2)\right]$ and $\phi_{0,it} = \phi\left[\mu_{it}(\boldsymbol{\theta}, \mathbf{0}, \sigma_a^2)\right]$, the former of which corresponds to the conditional mean of y_{it} when the null hypothesis is true. It follows from the first order Taylor expansion of p_{it} around $\boldsymbol{\gamma} = \mathbf{0}$ that

$$y_{it} = \Phi_{0,it} + \phi_{0,it}\boldsymbol{w}_i'\boldsymbol{\xi} + e_{it}, \quad \boldsymbol{\xi} = \frac{\boldsymbol{\gamma}}{\sqrt{1+\sigma_a^2}}. \tag{11}$$

So we have moment condition $\mathrm{E}\left(\sum_{t=1}^T \phi_{0,it}\boldsymbol{w}_{it}e_{it}\right) = \mathrm{E}\left[\sum_{t=1}^T \phi_{0,it}\boldsymbol{w}_{it}(1-\Phi_{0,it})\right] = \mathbf{0}$ under the null hypothesis. Since e_{it} is heteroskedastic, a more efficient one should be given by

$$\mathrm{E}\left[\sum_{t=1}^T \frac{\phi_{0,it}\boldsymbol{w}_i(y_{it}-\Phi_{0,it})}{\Phi_{0,it}(1-\Phi_{0,it})}\right] = \mathrm{E}\left(\sum_{t=1}^T r_{it}\boldsymbol{w}_i\right) = \mathbf{0}, \quad r_{it} = \frac{\phi_{0,it}(y_{it}-\Phi_{0,it})}{\Phi_{0,it}(1-\Phi_{0,it})}, \tag{12}$$

i.e., orthogonality of regressor $\boldsymbol{w}_i$ and generalized residual r_{it}. Let the ML parameter estimates under the null hypothesis be $(\hat{\boldsymbol{\theta}}, \hat{\sigma}_a^2)$ and the estimate of r_{it} based on them be $\hat{r}_{it}$. We thus obtain the sample version of the above moment, $\sum_{i=1}^N \sum_{t=1}^T \hat{r}_{it}\boldsymbol{w}_i/N$.

It is worthwhile mentioning that moment condition (12) is equivalent to the efficient score of the LM tests coming out from quasi-log-likelihood

$$\log Q_i(\boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma_a^2) = \sum_{t=1}^T \log \Phi\left[(2y_{it}-1)\mu_{it}(\boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma_a^2)\right]. \tag{13}$$

However, as mentioned before, the maximization of $\sum_{i=1}^N \log Q_i(\boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma_a^2)$ does not yield consistent parameter estimates due to the endogeneity brought by lagged dependent variable $y_{i,t-1}$.

Next we proceed to test the heteroskedastiticy and non-normality of unobserved heterogeneity a_i. For the alternative hypothesis of heteroskedastiticy, we specify a_i's scale coefficient as $\sigma_{ai}^2 = \sigma_a^2 \exp(2\boldsymbol{x}_i'\gamma)$ that has baseline parameter σ_a^2. In testing non-normality, we place an arbitrary non-normal distribution for the alternative hypothesis. The specific form of the density will be given in the Monte Carlo simulations in Section 3. To deal with all the alternative hypotheses coherently, let us define functions

$$\text{Omitted variables}: \quad \mu_{it}(\boldsymbol{\theta},\gamma,\sigma_a^2) = \frac{\boldsymbol{z}_{it}'\boldsymbol{\theta} + \boldsymbol{w}_i'\gamma}{\sqrt{1+\sigma_a^2}}, \tag{14}$$

$$\text{Heteroskedasticity}: \quad \mu_{it}(\boldsymbol{\theta},\gamma,\sigma_a^2) = \frac{\boldsymbol{z}_{it}'\boldsymbol{\theta}}{\sqrt{1+\sigma_a^2\exp(2\boldsymbol{x}_i'\gamma)}}, \tag{15}$$

$$\text{Non-normality}: \quad \mu_{it}(\boldsymbol{\theta},\gamma,\sigma_a^2) = \frac{\boldsymbol{z}_{it}'\boldsymbol{\theta} + \gamma_1(\boldsymbol{z}_{it}'\boldsymbol{\theta})^2 + \gamma_2(\boldsymbol{z}_{it}'\boldsymbol{\theta})^3}{\sqrt{1+\sigma_a^2}}. \tag{16}$$

Here the first one is the reproduction of equation (12). Invoking Ruud (1984), the last specification has a power against the general non-normality of error component $a_i + u_{it}$ Being restricted to $\gamma = \mathbf{0}$, all the three functions are reduced to

$$\text{Null model}: \quad \mu_{it}(\boldsymbol{\theta},\mathbf{0},\sigma_a^2) = \frac{\boldsymbol{z}_{it}'\boldsymbol{\theta}}{\sqrt{1+\sigma_a^2}}. \tag{17}$$

The above is obtained by the ML estimation without estimating γ.

Let the linear approximation of the outcome around the null hypothesis be

$$y_{it} = \Phi_{0,it} + \phi_{0,it}\mathring{\boldsymbol{w}}_{it}'\boldsymbol{\xi} + e_{it}, \quad \mathring{\boldsymbol{w}}_{it} = \nabla_\gamma \mu_{it}(\boldsymbol{\theta},\mathbf{0},\sigma_a^2), \tag{18}$$

where the vector of ancillary regressors $\mathring{\boldsymbol{w}}_{it}$ varies depending on which of the alternative models is considered. Specifically, it follows that

$$\text{Omitted variables}: \quad \mathring{\boldsymbol{w}}_{it} = \boldsymbol{w}_i = (\boldsymbol{x}_{i2}', \boldsymbol{x}_{i3}', \ldots, \boldsymbol{x}_{iT}')', \tag{19}$$

$$\text{Heteroskedasticity}: \quad \mathring{\boldsymbol{w}}_{it} = -\frac{1}{2}(1+\sigma_a^2)^{-\frac{3}{2}}\sigma_a^2\boldsymbol{x}_i(\boldsymbol{z}_{it}'\boldsymbol{\theta}) \propto \boldsymbol{x}_i(\boldsymbol{z}_{it}'\boldsymbol{\theta}), \tag{20}$$

$$\text{Non-normality}: \quad \mathring{\boldsymbol{w}}_{it} = \left[(\boldsymbol{z}_{it}'\boldsymbol{\theta})^2, (\boldsymbol{z}_{it}'\boldsymbol{\theta})^3\right]'. \tag{21}$$

Consequently the moment conditions to hold when the null hypothesis is true is given by

$$\mathrm{E}\left(\sum_{t=1}^{T} \frac{\phi_{0,it}\mathring{\boldsymbol{w}}_{it}(y_{it} - \Phi_{0,it})}{\Phi_{0,it}(1-\Phi_{0,it})}\right) = \mathrm{E}\left(\sum_{t=1}^{T} r_{it}\mathring{\boldsymbol{w}}_{it}\right) = \mathbf{0}, \quad r_{it} = \frac{\phi_{0,it}(y_{it} - \Phi_{0,it})}{\Phi_{0,it}(1-\Phi_{0,it})}. \tag{22}$$

The corresponding sample moment is thus

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{r}_{it}\hat{\mathring{\boldsymbol{w}}}_{it}, \tag{23}$$

computable with estimated regressors $\hat{\mathring{\boldsymbol{w}}}_{it}$ in hand. We may judge the deviation of null hypothesis from the actual data generating process through witnessing statistic (23).

2.3 Construction of robust CM test statistics

In contrast to the situations analyzed by the vast majority of existing studies (e.g., Skeels and Vella, 1999), the observations in this paper have within-individual or cluster correlation because of the panel structure. When the data is dependent it is difficult to use the auxiliary regression technique due to the generalized information equality by Newey (1985) and Tauchen (1985) to obtain the desired test statistics. So this subsection derives the limiting distribution regarding the sample moment given in equation (23).

Proposition 1 Define J dimensional vectors

$$\boldsymbol{s}_i = \sum_{t=1}^{T} r_{it}\mathring{\boldsymbol{w}}_{it}, \quad \hat{\boldsymbol{s}}_i = \sum_{t=1}^{T}\hat{r}_{it}\hat{\mathring{\boldsymbol{w}}}_{it}. \tag{24}$$

Suppose that

1. observation $(y_{i0}, \boldsymbol{y}_i, \boldsymbol{x}_i')'$, $i = 1, 2, \ldots, N$ are mutually independent,
2. T is fixed to be finite constant while $N \to \infty$, and
3. $\boldsymbol{s}_i$ is full-rank and has finite, positive definite covariance matrix

$$\boldsymbol{\Omega} = \mathrm{E}\left(\boldsymbol{s}_i\boldsymbol{s}_i'\right) = \sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{E}\left(r_{it}r_{is}\mathring{\boldsymbol{w}}_{it}\mathring{\boldsymbol{w}}_{it}'\right). \tag{25}$$

Then it follows that, under the null hypothesis,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{\boldsymbol{s}}_i \xrightarrow{\mathrm{d}} \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Omega}) \tag{26}$$

and so $\left(\sum_{i=1}^{N}\hat{\boldsymbol{s}}_i\right)'\boldsymbol{\Omega}^{-1}\left(\sum_{i=1}^{N}\hat{\boldsymbol{s}}_i\right)/N \overset{\mathrm{a}}{\sim} \mathrm{Chi}(J)$.

Proof. (i) Since the ML estimator is $\sqrt{N}$-consistent, we have $\sum_{i=1}^{N}(\hat{\boldsymbol{s}}_i - \boldsymbol{s}_i)/\sqrt{N} \xrightarrow{p} \mathbf{0}$. (ii) If the null hypothesis is true, then $\mathrm{E}(\boldsymbol{s}_i) = \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{s}_i) = \boldsymbol{\Omega}$ by assumption (25). So it follows that $\sum_{i=1}^{N} \boldsymbol{s}_i/\sqrt{N} \xrightarrow{\mathrm{d}} \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega})$ due to the Central Limit Theorem. (iii) By applying Theorem (x)-(d) of Rao (1973, pp. 123-124) to results (i) and (ii), we have equation (26). ■

Replacing J-dimensional covariance matrix $\boldsymbol{\Omega}$ in **Proposition 1** with its nonparametric estimator $\sum_{i=1}^{N} \hat{\boldsymbol{s}}_i \hat{\boldsymbol{s}}_i'/N$, we define robust CM statistic

$$\chi^2_{\mathrm{CM}} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{r}_{it}\hat{\mathring{\boldsymbol{w}}}_{it}\right)' \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \hat{r}_{it}\hat{r}_{is}\hat{\mathring{\boldsymbol{w}}}_{it}\hat{\mathring{\boldsymbol{w}}}_{is}'\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{r}_{it}\hat{\mathring{\boldsymbol{w}}}_{it}\right). \tag{27}$$

Statistic χ^2_{CM} follows chi-square distribution with degree of freedom J under the null hypothesis of $\boldsymbol{\xi} = \mathbf{0}$.

2.4 Least squares tests

In order to obtain the robust CM statistic given in equation (27), we need to make some programming efforts after the ML estimation of the null model. This post-estimation step may cast an obstacle to use it. So we offer another tractable testing procedure robust to the within-individual correlations.

Rewrite the weighted version of equation (18) as

$$\frac{y_{it} - \Phi_{0,it}}{\sqrt{(1-\Phi_{0,it})\Phi_{0,it}}} = \frac{\phi_{0,it}\mathring{\boldsymbol{w}}_{it}'\boldsymbol{\xi}}{\sqrt{(1-\Phi_{0,it})\Phi_{0,it}}} + \frac{e_{it}}{\sqrt{(1-\Phi_{0,it})\Phi_{0,it}}} \quad \Leftrightarrow \quad \tilde{y}_{it} = \tilde{\boldsymbol{w}}_{it}'\boldsymbol{\xi} + \tilde{e}_{it}. \tag{28}$$

Let

$$\hat{\boldsymbol{\xi}} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\boldsymbol{w}}_{it}\tilde{\boldsymbol{w}}_{it}'\right)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\boldsymbol{w}}_{it}\tilde{y}_{it} \tag{29}$$

be the least squares estimator of $\boldsymbol{\xi}$ in equation (18). Then its limiting distribution is given by

$$\sqrt{N}\hat{\boldsymbol{\xi}} \xrightarrow{\mathrm{d}} \mathrm{N}(\mathbf{0}, \boldsymbol{V}), \quad \boldsymbol{V} = \boldsymbol{H}\boldsymbol{\Omega}^{-1}\boldsymbol{H}, \tag{30}$$

here $\boldsymbol{H} = \mathrm{E}\left(\sum_{t=1}^{T} \tilde{\boldsymbol{w}}_{it}\tilde{\boldsymbol{w}}_{it}'\right)$ and

$$\boldsymbol{\Omega} = \sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{E}\left(r_{it}r_{is}\mathring{\boldsymbol{w}}_{it}\mathring{\boldsymbol{w}}_{it}'\right) = \sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{E}\left(\tilde{e}_{it}\tilde{e}_{is}\tilde{\boldsymbol{w}}_{it}\tilde{\boldsymbol{w}}_{is}'\right). \tag{31}$$

Thus quadratic form

$$\chi^2_{\text{LS}} = \hat{\boldsymbol{\xi}}' \left[\left(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\boldsymbol{w}}_{it} \tilde{\boldsymbol{w}}_{it}' \right)' \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{e}_{it} \tilde{e}_{is} \tilde{\boldsymbol{w}}_{it} \tilde{\boldsymbol{w}}_{is}' \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\boldsymbol{w}}_{it} \tilde{\boldsymbol{w}}_{it}' \right) \right]^{-1} \hat{\boldsymbol{\xi}} \tag{32}$$

obeys the chi-square distribution with degree of freedom J asymptotically.

The form of χ^2_{LS} in equation (32) appears to be more complicated than that of χ^2_{CM} in equation (27). Actually, however, computing χ^2_{LS} is straightforward if a statistical software with cluster-robust covariance matrix routine is available. The procedure is as follows.

1. Estimate $\boldsymbol{\theta}$ and σ^2_a under the null via dynamic probit ML in equation (7). (This step is identical to that in obtaining χ^2_{CM}.)

2. Construct regression error $\tilde{y}_{it}$ and ancillary regressor $\tilde{\boldsymbol{w}}_{it}$ from the ML estimator $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}^2_a$.

3. Regress $\tilde{y}_{it}$ on $\tilde{\boldsymbol{w}}_{it}$ (and no constant term) with "cluster" option.

4. See the F or chi-square statistics of joint significance test on $\boldsymbol{\xi}$, which is what we want.

The F or chi-square statistics at the above fourth step are often automatically generated as a goodness of fit measure after running a regression package. If $\tilde{\boldsymbol{w}}_{it}$ is scaler, the use of the cluster-robust t statistic of significance test is advised.

Two new statistics χ^2_{CM} and χ^2_{LS} share common limiting chi-square distribution Chi(J) but the latter has the advantage of simplicity. An additional attraction of the LS approach is that, as demonstrated in the empirical analysis of Section 3, it is easy to check which variable contributes to a given misspecification separately. However, their performance under the realistic numbers of observations are unclear. So the next section will be devoted to compare their finite sample performances through Monte Carlo simulations.

3 Monte Carlo Study and Empirical Application

3.1 Design of simulation

This subsection describes the details on the simulation design. The number of individuals and time periods examined were $N = \{500, 1000\}$ and $T + 1 = \{4, 8\}$. In generating individual time series we gave zero to the initial value, run the process specified below, and discarded first 50 periods to eliminate the influence of the arbitrary chosen initial value. We replicated the estimation and computation of statistics 1000 times. The simulation was conducted by Ox console version 7.00 (Doornik, 2007).

The data generating process under the null hypothesis was built as follows. For a given individual, we drew a sequence of univariate time-varying regressors from normal autoregressive process

$$x_{it} = 0.5x_{i,t-1} + \sqrt{1-0.5}x^*_{it}, \quad x^*_{it} \sim \mathrm{N}(0,1), \tag{33}$$

so that $\mathrm{E}(x_{it}) = 0$ and $\mathrm{Var}(x_{it}) = 0$.[*4] The unobserved heterogeneity was assumed to have uniform correlations with period-by-period regressors and to distribute as standard normal;

$$a^* = \gamma \sum_{s=1}^{T} x_{is} + \sigma_a a_i, \quad a_i \sim \mathrm{N}(0,1). \tag{34}$$

Outcome y_{it} was then generated from

$$\begin{aligned} y^*_{it} &= \rho y_{i,t-1} + \beta_0 + \beta_1 x_{it} + a^*_i + u_{it} \\ &= \rho y_{i,t-1} + \beta_0 + \beta_1 x_{it} + \gamma \sum_{s=1}^{T} x_{is} + \sigma_a a_i + u_{it}, \quad y_{it} = \mathbf{1}[y^*_{it} > 0], \end{aligned} \tag{DGP0}$$

where $u_{it} \sim \mathrm{N}(0,1)$ represent independent standard normal disturbances. The parameter values were fixed such that

$$(\delta, \beta_0, \beta_1, \gamma, \sigma_a) = (0.5, 1.0, 1.0, 0.5, 1.0). \tag{35}$$

Hereafter we call the setting we made so far DGP0.

As alternative hypothesis, we considered the misspecification regarding unobserved heterogeneity a^*_{it} in equation (2); omitted variables (equivalent to time variation of γ), heteroskedastictity, and non-normality. For the case of omitted variables, we replace $\gamma \sum_{s=1}^{T} x_{is}$ ($\gamma = 0.5$) of in DGP0 to

$$\begin{cases} g_{i4} = (-2.0x_{i1} - 0.5x_{i2} + 2.0x_{i3})/4 & \text{for } T+1 = 4, \\ g_{i8} = (-2.0x_{i1} - 0.5x_{i2} + 2.0x_{i3} + 2.0x_{i4} - 0.5x_{i5} - 2.0x_{i6} - 2.0x_{i7})/8 & \text{for } T+1 = 8. \end{cases} \tag{DGP1}$$

[*4] We replaced the normal with the uniform distribution and obtained similar conclusions on the Monte Carlo study.

In generating y^*_{it} with heteroskedastic a^*_{it}, parameter σ_a in DGP0 is substituted with

$$\begin{cases} \exp(g_{i4}) & \text{for } T+1=4, \\ \exp(g_{i8}) & \text{for } T+1=8. \end{cases} \tag{DGP2}$$

so that not only the mean but also variance of a^*_i is dependent upon the history of regressors. To generate non-normal individual heterogeneity, we replace $a_i \sim \mathrm{N}(0,1)$ with finite mixture of normals

$$a_i = \mathbf{1}[\kappa_i \le 0.7]b_i + \mathbf{1}[\kappa_i > 0.7]c_i, \quad b_i \sim \mathrm{N}\left(-2, \frac{1}{9}\right) \quad c_i \sim \mathrm{N}\left(2, \frac{1}{9}\right), \tag{DGP3}$$

where $\kappa_i \sim \mathrm{U}(0,1)$. Note that κ_i also brings heteroskedasticity into a_i, interpreted as latent class assignments of individual heterogeneity. Consequently, DGP3 should be detected by testing heteroskedasticity.

Each repetition in the simulation starts with the ML estimation of the null model. We estimated the simplified version of Wooldridge (2005) model,

$$y^*_{it} = \delta y_{i,t-1} + \beta_0 + \beta_1 x_{it} + \delta_0 y_{i0} + (\gamma T)\overline{x}_i + \sigma_a a_i + u_{it}, \quad t = 1,2,\dots,T, \tag{36}$$

using the panel of (x_{it}, y_{it}) generated by the aforementioned process. The log-likelihood function derived from the above model was evaluated by 20 points Gauss-Hermite quadrature. Under the null hypothesis, i.e., DGP0, this estimator is supposed to be consistent to the unknown parameter. The following points should be remarked here. First, constant term β_0 and scale parameter σ^2_a are not identified in the reduced form expression above. Second, the value of coefficient δ_0 is unknown a priori.

For constructing test statistics χ^2_{CM} and χ^2_{LS} for detecting specification errors given in DGP1, GDP2, and DGP3, we exploited the following moment conditions called TEST1, TEST2, and TEST3, respectively. (J denotes the degree of freedom.)

$$\sum_{t=1}^{T} \mathrm{E}\left[(x_{i2}, x_{i3} \dots, x_{iT})' \, r_{it}\right] = \mathbf{0}, \quad J = 2,6, \tag{TEST1}$$

$$\sum_{t=1}^{T} \mathrm{E}\left[(x_{i1}, x_{i2} \dots, x_{iT})' \, \mu_{it} r_{it}\right] = \mathbf{0}, \quad J = 3,7, \tag{TEST2}$$

$$\sum_{t=1}^{T} \mathrm{E}\left[\left(\mu^2_{it}, \mu^3_{it}\right)' \, r_{it}\right] = \mathbf{0}, \quad J = 2. \tag{TEST3}$$

Note that the degrees of freedom of TEST1 and TEST2 depend on the length of time period $T+1$

whereas fixed to two for TEST3. To verify that the asymptotic approximation works properly, we further considered a set of *placebo* moment conditions;

$$\sum_{t=1}^{T} \mathrm{E}\left[(m_{1,it}, m_{2,it})' r_{it}\right] = \mathbf{0}, \quad J = 2, \tag{PLACEBO1}$$

$$\sum_{t=1}^{T} \mathrm{E}\left[(m_{1,it}, m_{2,it})' \mu_{it} r_{it}\right] = \mathbf{0}, \quad J = 2. \tag{PLACEBO2}$$

In the above conditions, $m_{1,it} \sim \mathrm{N}(0,1)$ and $m_{2,it} \sim \mathrm{N}(0,1)$ represents pure white noises not appeared in the models. So the corresponding CM/LS tests should not have powers against GDP1, DGP2, nor DGP3 and serve as benchmarks. Consequently there are five test statistics given the data set.

Table 1 summarizes the 20 combinations of test statistics and actual data generating process. If the data follow DGP0 (i.e., the null model), the rejection frequencies of all the tests in simulation correspond to their empirical sizes. Likewise, under DGP1, DGP2, and DGP3, the rejection frequencies of TEST1, TEST2, and TEST3 mean their powers respectively. Word "possible" in the table implies that the test may have power on a given misspecification although it is not a legitimate test. For example, TEST2 is intended to detecting heteroskedastictiy (DGP2) but may have power on the omitted variables (DGP1).

3.2 Monte Carlo Results

Before moving on to the results on the testing hypothesis, we shall take a brief look at the estimated bias of RE dynamic probit due to specification errors. Table 2 shows the Monte Carol mean, mean bias (in percent), and root mean squared error (RMSE) for each of tests with $N = 500$. As mentioned in the previous subsection, constant β_0 and variance σ_a^2 are not identified in the framework of Wooldridge (2005) and so the bias on them are not defined. It is obvious from the table that misspecification concerning individual effect a_i can cause non-negligible bias on lagged outcome's and regressor's coefficients δ and β_1 when time period is short ($T + 1 = 4$). However, the bias got diminished as time period being extended. When $T + 1 = 8$, the bias was virtually zero for all DGPs. Importantly, unidentified parameters β_0 and σ_a^2 played a role of absorbing the influences of misspecification on the estimates of key parameters δ and β_1. We can see similar tendencies in the results of larger sample size $N = 1000$ in Table 3.

Table 4 through Table 7 present the mean of statistics and rejection frequencies at 10 and 5 percent nominal sizes for four patterns of data generations. (Theoretically, the mean should be close to the degree of freedom when DGP0, the null hypothesis, is true.) Each table compares the test performances of χ^2_{CM} and χ^2_{LS}. Overall, these two tests generated similar results but χ^2_{LS} had slightly higher mean and rejection frequency than χ^2_{CM} did.

Table 4 displays the simulation results when the data came from DGP0, the null hypothesis. Therefore all the rejection frequencies appeared in the table measure the sizes of tests. In shorter time period $T + 1 = 4$, TEST1 (designed for detecting DGP1) and TEST2 (for DGP2) tended to under-reject the null hypothesis. In contrast, when the time period is $T + 1 = 8$, TEST1 and TEST2 had empirical sizes close to the nominal ones but TEST3 (for DGP3) under-rejects the null. For all cases of N and $T + 1$, PLACEBO1 and PLACEB02 exhibited nearly correct sizes.

The results when the data generation process is DGP1 (omitted variables) or DGP2 (heteroskedasticity) are given in 5 and Table 6, respectively. For DGP1, TEST1 and TEST2 had great powers for detecting omitted variables (see Table 5). On the other hand, the test for heteroskedasticity, TEST2, poorly worked for its purpose in Table 6. An increase in N seem to improve its power to some extent. TEST3 responded to omitted variables (heteroskedasticity) moderately when time period was longer (shorter).

Finally, according to Table 7 designated for DGP3, TEST3 has powers against non-normality at satisfactory levels. Particularly with $N = 1000$, it correctly maintained the alternative hypothesis. However, the power was declining as T being extended. TEST1 and TEST2 may have the weak senses of fining non-normality depending upon N and/or T.

3.3 Application: Specification tests for union membership dynamics

In this subsection we test the misspecification of the model for the union membership dynamics estimated by Wooldridge (2005). The data were originally analyzed by Vella and Verbeek (1998) and available in *Journal of Econometrics Data Archive*.[*5] We performed the ML estimation of RE dynamic probit model regressing the dummy of the union membership on the previous year's membership, marital status, and individual characteristics. In this regression the time-varying regressors is only marital status and time dummies. For simplicity, we used the convenient LS approach to obtain the test statistics. All of the outputs shown up in this subsection was generated by Stata version 13.1. (Stata code is available upon request.)

Table 8 shows two sets of estimates. The first one, labeled with Model 1, imposes the exclusion restriction on the set of marital status variables (appeared in the model as controls) discussed in section 2. The second one, Model 2, represents the model under the alternative hypothesis, a variant of Wooldridge (2005) under the different parameterizations on the marital variables. The minor difference in the results of Wooldridge (2005) and of the current paper may stem from the difference in parameterizations and/or the quadrature techniques employed in the

[*5]The ULR is; `http://qed.econ.queensu.ca/jae/`

ML estimation. The conventional Wald statistic testing for the joint significance of "Married 1982" through "Married 1987" is 6.93 with degree of freedom six in the table, not rejecting the restriction.[*6]

We tested the same hypothesis as in Table 8 again using χ^2_{LS} and gave the results in the top panel of Table 9. The value of the test statistic (8.82) was slightly different from that of the conventional Wald in Table 8 but we drew the same conclusion from it. In terms of the first moment modeling of individual effects in equation (2), general coefficients of marital control variables may be redundant.

The middle panel of Table 9 shows the test statistics for heteroskedasticity. In the table we see that Model 1, which is a more parsimonious and widely employed specification in the literature, failed to reject the presence of heteroskedasticity on individual effects. In contrast, in Model 2 (with no restrictions on the marital control variables), heteroskedasticity was not statistically significant. In light of this finding, it may be advisable that, for the purpose of circumventing heteroskedasticity, one should leave the correlation patterns of unobserved effects and time-varying regressors unrestricted.

Finally, the bottom panel of Table 9 is on testing non-normality of individual effects. According to the statistics given in the table, the normality assumption is not supported in neither Model 1 nor Model 2. Even if we allowed the model to have free correlations between marital control variables and individual effects, the computed χ^2_{LS} statistic is well above its critical value. Therefore flexible distributions used in Christelis and Sanz-de Galdeano (2011) and Deza (2015) might be more appropriate to this data than the normal distribution.

4 Concluding Remark

This paper proposed two moment-based tests for detecting possible specification errors in the unobserved heterogeneity for the RE dynamic probit models. The proposed tests are computationally simple and robust to the presence of within-individual correlations in observations. We demonstrated the performances of these tests via Monte Carlo simulations and real application to the data analyzed by Vella and Verbeek (1998) and Wooldridge (2005).

Throughout this study our chief concern has been on the RE dynamic probit. However, the situation considered here encompasses general cases where the principal model consists of the joint distribution but a researcher is concerned with specification errors in the marginal distribution of an outcome. The test statistics presented here are readily applicable to, for example,

[*6]In the RE estimation of dynamic probit, the dependence of observations is handled by parametric joint modeling of the outcomes. So the z and Wald statistics appeared in Table 8 are computed based on the information matrix equality of the ML estimation.

autoregressive dynamic probit models (Hyslop, 1999).

The Monte Carlo session of this paper examined only a limited variety of alternative hypotheses. Se, to further investigate the test performances in small samples, we need an extended sets of data generating processes. We also did not compare the performances of our new tests to those of Wald tests build on the estimation of the model under the alternative hypothesis. The loss of efficiency in testing has been unclear by abandoning the parametric joint distribution used at the estimation stage. These issues remain to be considered in future studies.

References

Akay, A. (2012). Finite-sample comparison of alternative methods for estimating dynamic panel data models. *Journal of Applied Econometrics 27*(7), 1189–1204.

Arulampalam, W. and M. B. Stewart (2009). Simplified implementation of the Heckman estimator of the dynamic probit model and a comparison with alternative estimators. *Oxford Bulletin of Economics and Statistics 71*(5), 659–681.

Bertschek, I. and M. Lechner (1998). Convenient estimators for the panel probit model. *Journal of Econometrics 87*(2), 329–371.

Chamberlain, G. (1984). Panel data. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2. Elsevier.

Christelis, D. and A. Sanz-de Galdeano (2011). Smoking persistence across countries: A panel data analysis. *Journal of Health Economics 30*(5), 1077–1093.

Davidson, R. and J. G. MacKinnon (1984). Convenient specification tests for logit and probit models. *Journal of Econometrics 25*(3), 241–262.

Deza, M. (2015). Is there a stepping stone effect in drug use? separating state dependence from unobserved heterogeneity within and between illicit drugs. *Journal of Econometrics 184*(1), 193–207.

Doornik, J. (2007). *Object-Oriented Matrix Programming Using Ox* (3rd ed.). Timberlake Consultants Press and Oxford.

Greene, W. and C. McKenzie (2015). An lm test based on generalized residuals for random effects in a nonlinear model. *Economics Letters 127*, 47–50.

Greene, W. H. (2011). *Econometric Analysis* (seventh ed.). Pearson Education.

Heckman, J. and B. Singer (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica 52*(2), 271–320.

Heckman, J. J. (1981). The incidental parameters problem and the problem of initial condition in estimating a discrete time-discrete data stochastic process. In C. F. Manski and D. L. McFadden (Eds.), *Structural Analysis of Discrete Data and Econometric Applications*. The MIT Press.

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

Hyslop, D. R. (1999). State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica 67*(6), 1255–1294.

Keane, M. P. and R. M. Sauer (2010). A computationally practical simulation estimation algorithm for dynamic panel data models with unobserved endogenous state variables. *International Economic Review 51*(4), 925–958.

Lechner, M. (1995). Some specification tests for probit models estimated on panel data. *Journal of Business & Economic Statistics 13*(4), 475–88.

Newey, W. K. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica 53*(5), 1047–70.

Pagan, A. and F. Vella (1989). Diagnostic tests for models based on individual data: A survey. *Journal of Applied Econometrics 4*(S), S29–59.

Rabe-Hesketh, S. and A. Skrondal (2013). Avoiding biased versions of Wooldridge's simple solution to the initial conditions problem. *Economics Letters 120*(2), 346–349.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. John Wiley & Sons.

Ruud, P. A. (1984). Tests of specification in econometrics. *Econometric Reviews 3*(2), 211–242.

Skeels, C. L. and F. Vella (1999). A monte carlo investigation of the sampling behavior of conditional moment tests in tobit and probit models. *Journal of Econometrics 92*(2), 275–294.

Tauchen, G. (1985). Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics 30*(1-2), 415–443.

Vella, F. and M. Verbeek (1998). Whose wages do unions raise? a dynamic model of unionism and wage rate determination for young men. *Journal of Applied Econometrics 13*(2), 163–183.

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*(1), 39–54.

Yatchew, A. and Z. Griliches (1985). Specification error in probit models. *The Review of Economics and Statistics 67*(1), 134–39.

	DGP0	DGP1	DGP2	DGP3
TEST1	size	power	possible	possible
TEST2	size	possible	power	possible
TEST3	size	possible	possible	power
PLACEBO1	size	size	size	size
PLACEBO2	size	size	size	size

Table 1: Combinations of DGP and test statistics

Note: There are four data generating processes and five tests. Word “possible” implies the test may have power for the data generating process.

		$T+1=4$			$T+1=8$		
	Ture	Mean	Bias%	RMSE	Mean	Bias%	RMSE
DGP0							
δ	0.50	0.47	-6.57	0.20	0.49	-1.27	0.11
β_0	1.00	0.20	-79.99	0.82	0.22	-78.03	0.79
β_1	1.00	1.01	0.81	0.13	1.00	0.07	0.07
σ_a^2	1.00	0.80	-20.01	0.38	0.76	-23.62	0.28
DGP1							
δ	0.50	0.33	-34.44	0.24	0.51	1.48	0.09
β_0	1.00	0.06	-94.32	0.96	0.02	-98.11	0.99
β_1	1.00	0.99	-0.64	0.11	0.99	-1.34	0.06
σ_a^2	1.00	1.23	22.70	0.44	0.96	-3.65	0.16
DGP2							
δ	0.50	0.52	3.49	0.20	0.49	-1.38	0.11
β_0	1.00	-0.18	-117.88	1.19	-0.10	-109.92	1.11
β_1	1.00	0.94	-5.97	0.14	0.99	-0.90	0.07
σ_a^2	1.00	1.06	5.67	0.42	1.13	13.17	0.27
DGP3							
δ	0.50	0.47	-6.41	0.21	0.50	-0.88	0.11
β_0	1.00	-1.31	-231.10	2.32	-1.30	-230.13	2.30
β_1	1.00	1.00	-0.17	0.13	1.00	0.00	0.07
σ_a^2	1.00	1.78	78.12	0.97	1.81	81.13	0.88

Table 2: Bias of RE dynamic probit ML estimation under misspecification ($N = 500$)

Note: The number of replications in the Monte Carlo simulation is 1000. DGP0, DGP1, DGP2, and GDP3 correspond to the null hypothesis, omitted variables, heteroskedasticity, and non-normality (normal mixture), respectively. Parameter β_0 and σ_a^2 are not identified.

	Ture	$T+1=4$			$T+1=8$		
		Mean	Bias%	RMSE	Mean	Bias%	RMSE
DGP0							
δ	0.50	0.47	-6.98	0.14	0.50	-0.93	0.08
β_0	1.00	0.20	-80.39	0.81	0.22	-77.93	0.79
β_1	1.00	1.01	0.45	0.08	1.00	-0.19	0.05
σ_a^2	1.00	0.79	-21.20	0.30	0.76	-23.59	0.26
DGP1							
δ	0.50	0.32	-36.40	0.22	0.51	1.66	0.06
β_0	1.00	0.06	-93.77	0.94	0.02	-98.22	0.99
β_1	1.00	0.99	-0.76	0.08	0.98	-1.73	0.04
σ_a^2	1.00	1.23	22.74	0.35	0.96	-3.68	0.12
DGP2							
δ	0.50	0.52	3.82	0.14	0.49	-1.51	0.08
β_0	1.00	-0.18	-117.98	1.19	-0.10	-110.02	1.11
β_1	1.00	0.94	-6.10	0.10	0.99	-1.25	0.05
σ_a^2	1.00	1.05	4.72	0.28	1.13	13.08	0.21
DGP3							
δ	0.50	0.46	-8.08	0.15	0.50	-0.78	0.08
β_0	1.00	-1.31	-231.12	2.31	-1.29	-229.32	2.30
β_1	1.00	1.00	-0.14	0.09	1.00	-0.50	0.05
σ_a^2	1.00	1.79	79.15	0.89	1.81	80.78	0.84

Table 3: Bias of RE dynamic probit ML estimation under misspecification ($N = 1000$)

Note: The number of replications in the Monte Carlo simulation is 1000. DGP0, DGP1, DGP2, and GDP3 correspond to the null hypothesis, omitted variables, heteroskedasticity, and non-normality (normal mixture), respectively. Parameter β_0 and σ_a^2 are not identified.

	J	CM			LS		
		Mean	10%	5%	Mean	10%	5%
$N = 500, T+1 = 4$							
TEST1	2	1.4	5.6	2.2	1.4	5.8	2.6
TEST2	3	2.7	6.0	2.3	2.8	7.0	2.5
TEST3	2	1.8	9.6	5.6	1.8	10.3	6.0
PLACEBO1	2	2.0	9.8	4.8	2.0	10.6	5.0
PLACEBO2	2	2.0	9.1	4.5	2.0	9.4	5.3
$N = 500, T+1 = 8$							
TEST1	6	5.6	9.2	4.2	6.0	13.6	7.0
TEST2	7	6.7	7.0	3.8	7.0	9.4	5.1
TEST3	2	1.5	6.0	3.2	1.6	6.9	3.4
PLACEBO1	2	2.0	10.2	4.3	2.0	10.6	4.6
PLACEBO2	2	2.2	10.7	5.0	2.2	11.7	5.3
$N = 1000, T+1 = 4$							
TEST1	2	1.6	5.5	2.4	1.6	5.6	2.4
TEST2	3	2.9	9.0	4.2	3.0	9.5	4.9
TEST3	2	1.6	7.2	3.7	1.6	7.4	4.0
PLACEBO1	2	2.0	9.6	4.8	2.0	9.7	4.9
PLACEBO2	2	2.0	10.4	4.5	2.0	10.5	4.9
$N = 1000, T+1 = 8$							
TEST1	6	6.0	10.5	5.5	6.2	12.1	7.1
TEST2	7	6.8	7.9	2.8	7.0	9.4	3.8
TEST3	2	1.6	6.8	3.7	1.6	6.8	3.7
PLACEBO1	2	2.0	10.8	5.0	2.0	11.1	5.1
PLACEBO2	2	2.0	10.4	5.1	2.0	10.4	5.2

Table 4: Mean of statistics and empirical size for DGP0

Note: The number of replications in the Monte Carlo simulation is 1000. The means of statistics and rejection frequencies are given here. J denotes the degree of freedom of the test. DGP0 implies that the data is generated from the model under the null hypothesis. Theoretically no tests have powers and so all the rejection frequencies measure the size.

		CM			LS		
	DF	Mean	10%	5%	Mean	10%	5%
$N = 500, T + 1 = 4$							
TEST1	2	26.2	100.0	99.9	31.4	100.0	99.9
TEST2	3	22.7	99.6	98.7	25.4	99.6	99.2
TEST3	2	1.5	7.3	4.7	1.5	7.9	5.2
PLACEBO1	2	1.9	8.6	4.5	1.9	8.9	4.7
PLACEBO2	2	2.0	9.8	4.2	2.0	10.2	4.2
$N = 500, T + 1 = 8$							
TEST1	6	44.2	100.0	100.0	60.5	100.0	100.0
TEST2	7	32.3	99.6	99.1	38.6	99.7	99.2
TEST3	2	2.7	17.7	10.6	2.8	18.6	11.8
PLACEBO1	2	2.0	8.9	4.0	2.0	9.2	4.1
PLACEBO2	2	2.1	10.0	5.3	2.1	10.3	5.7
$N = 1000, T + 1 = 4$							
TEST1	2	51.9	100.0	100.0	61.6	100.0	100.0
TEST2	3	42.2	100.0	100.0	46.9	100.0	100.0
TEST3	2	1.4	7.3	3.6	1.5	7.8	3.9
PLACEBO1	2	2.0	9.4	4.6	2.0	9.8	4.9
PLACEBO2	2	1.9	10.5	4.9	2.0	10.7	5.2
$N = 1000, T + 1 = 8$							
TEST1	6	85.2	100.0	100.0	114.6	100.0	100.0
TEST2	7	58.8	100.0	100.0	69.1	100.0	100.0
TEST3	2	3.7	28.7	19.4	3.9	28.9	20.7
PLACEBO1	2	1.9	8.0	4.6	1.9	8.4	4.6
PLACEBO2	2	2.1	10.9	5.3	2.1	10.9	5.3

Table 5: Mean and size/power for DGP1

Note: The number of replications in the Monte Carlo simulation is 1000. The means of statistics and rejection frequencies are given here. J denotes the degree of freedom of the test. DGP1 implies that the data is generated from the model in the presence of omitted variables. Theoretically TEST1 has power against DGP1.

	DF	CM			LS		
		Mean	10%	5%	Mean	10%	5%
$N = 500, T + 1 = 4$							
TEST1	2	1.4	5.2	1.9	1.4	5.8	2.5
TEST2	3	3.8	15.9	8.4	3.9	18.1	9.4
TEST3	2	3.1	22.0	15.1	3.2	23.4	15.6
PLACEBO1	2	2.0	10.0	5.4	2.0	10.4	5.8
PLACEBO2	2	2.0	10.5	5.1	2.0	10.9	5.5
$N = 500, T + 1 = 8$							
TEST1	6	6.0	10.2	5.7	6.5	14.6	8.5
TEST2	7	8.2	13.5	5.4	8.6	16.9	8.4
TEST3	2	2.1	11.7	7.0	2.2	12.5	7.8
PLACEBO1	2	2.0	8.6	4.3	2.0	9.3	4.7
PLACEBO2	2	2.0	9.7	5.0	2.0	9.8	5.3
$N = 1000, T + 1 = 4$							
TEST1	2	1.7	9.3	4.0	1.8	9.8	4.4
TEST2	3	5.2	32.2	18.5	5.4	33.7	20.3
TEST3	2	3.3	25.3	15.6	3.4	25.9	16.4
PLACEBO1	2	2.0	10.4	5.3	2.0	10.5	5.4
PLACEBO2	2	1.9	8.3	3.5	1.9	8.6	3.5
$N = 1000, T + 1 = 8$							
TEST1	6	6.6	15.1	9.5	6.9	17.1	11.3
TEST2	7	11.0	36.9	21.9	11.5	41.7	25.4
TEST3	2	2.5	14.2	9.3	2.5	14.2	9.6
PLACEBO1	2	2.0	9.9	5.4	2.1	10.5	5.5
PLACEBO2	2	2.0	11.3	5.2	2.0	11.6	5.2

Table 6: Mean and size/power for DGP2

Note: The number of replications in the Monte Carlo simulation is 1000. The means of statistics and rejection frequencies are given here. J denotes the degree of freedom of the test. DGP1 implies that the data is generated from the model in the presence of parametric heteroskedasticity. Theoretically TEST2 has power against DGP2.

	DF	CM			LS		
		Mean	10%	5%	Mean	10%	5%
$N = 500, T + 1 = 4$							
TEST1	2	1.7	7.2	3.3	1.8	7.6	3.6
TEST2	3	4.9	28.6	19.9	5.0	29.2	20.6
TEST3	2	5.8	40.2	30.0	6.2	40.5	30.4
PLACEBO1	2	2.0	9.2	5.1	2.0	9.8	5.5
PLACEBO2	2	2.1	10.0	4.4	2.1	10.3	4.9
$N = 500, T + 1 = 8$							
TEST1	6	6.3	12.3	5.9	6.7	14.7	8.1
TEST2	7	7.8	15.4	8.4	8.1	18.3	10.5
TEST3	2	2.1	10.2	6.1	2.1	10.4	6.4
PLACEBO1	2	2.0	10.2	4.7	2.0	10.6	5.2
PLACEBO2	2	2.0	9.7	4.9	2.0	9.9	5.0
$N = 1000, T + 1 = 4$							
TEST1	2	1.9	8.7	3.9	2.0	8.8	4.0
TEST2	3	5.8	36.5	26.5	5.8	36.4	26.7
TEST3	2	8.0	64.9	50.5	8.1	64.8	50.8
PLACEBO1	2	2.0	11.0	5.4	2.1	11.1	5.5
PLACEBO2	2	2.0	9.4	6.1	2.0	9.6	6.1
$N = 1000, T + 1 = 8$							
TEST1	6	7.1	18.4	11.9	7.4	19.7	13.5
TEST2	7	7.9	15.6	8.5	8.1	16.6	9.7
TEST3	2	2.5	14.9	8.4	2.6	14.8	8.4
PLACEBO1	2	2.1	9.7	5.0	2.1	9.8	5.2
PLACEBO2	2	2.0	8.7	4.8	2.0	9.1	5.0

Table 7: Mean and size/power for DGP3

Note: The number of replications in the Monte Carlo simulation is 1000. The means of statistics and rejection frequencies are given here. J denotes the degree of freedom of the test. DGP1 implies that the data is generated from the model when the unobserved individual effects follow the mixture of normal distributions. Theoretically TEST3 has power against DGP3.

	Model 1		Model 2	
	Est.	z-value	Coef	z-value
Married	0.17	1.56	0.17	1.57
Lagged union	0.90	9.69	0.90	9.69
Union 1980	1.42	8.71	1.45	8.80
Avg. married	0.11	0.52	-0.32	-0.34
Married 1982			0.03	0.10
Married 1983			-0.05	-0.18
Married 1984			0.07	0.23
Married 1985			0.44	1.52
Married 1986			0.15	0.55
Married 1987			-0.34	-1.43
Education	-0.02	-0.44	-0.02	-0.57
Black	0.54	2.92	0.53	2.89
Year 1982	0.03	0.25	0.03	0.24
Year 1983	-0.09	-0.77	-0.09	-0.77
Year 1984	-0.05	-0.43	-0.05	-0.43
Year 1985	-0.27	-2.19	-0.27	-2.19
Year 1986	-0.32	-2.55	-0.32	-2.55
Year 1987	0.07	0.60	0.07	0.60
Constant	-1.75	-3.93	-1.65	-3.74
σ_a^2	1.09		1.08	
Joint Wald $\sim$ Chi(6)			6.93	
Log likelihood	-1283.74		-1283.70	
N	545		545	
T	7		7	

Table 8: RE dynamic probit estimates for union status by Wooldridge (2005) data

Note: The estimates and z-values for a significance test are given in the table. “Joint Wald” implies the joint significance test on “Married 1982” through M“arried 1987”.

	Model 1		Model 2	
	Est.	z-value	Est.	z-value
Omitted variables				
Married 1982	-0.01	-0.06		
Married 1983	-0.11	-0.58		
Married 1984	0.06	0.38		
Married 1985	0.30	1.79		
Married 1986	0.00	-0.01		
Married 1987	-0.22	-1.90		
Joint $\chi^2_{\mathrm{LS}} \sim$ Chi(6)		8.82		
Heteroskedasticity				
$\phi_{0it}\mu_{0it} \times$ Married 1981	-0.01	-0.04	-0.09	-0.73
$\phi_{0it}\mu_{0it} \times$ Married 1982	0.13	0.84	0.11	0.81
$\phi_{0it}\mu_{0it} \times$ Married 1983	0.01	0.08	0.00	0.01
$\phi_{0it}\mu_{0it} \times$ Married 1984	-0.03	-0.22	-0.01	-0.07
$\phi_{0it}\mu_{0it} \times$ Married 1985	-0.26	-1.63	-0.01	-0.06
$\phi_{0it}\mu_{0it} \times$ Married 1986	0.05	0.27	0.03	0.16
$\phi_{0it}\mu_{0it} \times$ Married 1987	0.25	2.51	0.07	0.63
Joint $\chi^2_{\mathrm{LS}} \sim$ Chi(7)		16.80		6.23
Non-normality				
$\phi_{0it}\mu^2_{0it}$	0.70	3.67	0.38	2.37
$\phi_{0it}\mu^3_{0it}$	0.61	4.33	0.35	3.03
Joint $\chi^2_{\mathrm{LS}} \sim$ Chi(2)		49.66		30.46

Table 9: Misspecification tests results on Wooldridge (2005) data

Note: The estimates and z-values for a significance test are given in the table. "Joint χ^2_{LS}" of different degrees of freedom tests the omitted variables, heteroskedasticity, and non-normality of the model.