ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Fernández-Val, Iván; Vella, Francis

Working Paper

Bias corrections for two-step fixed effects panel data estimators

cemmap working paper, No. CWP04/07

Provided in Cooperation with:
The Institute for Fiscal Studies (IFS), London

Suggested Citation: Fernández-Val, Iván; Vella, Francis (2007) : Bias corrections for two-step fixed effects panel data estimators, cemmap working paper, No. CWP04/07, Centre for Microdata Methods and Practice (cemmap), London, https://doi.org/10.1920/wp.cem.2007.0407

This Version is available at:
https://hdl.handle.net/10419/79352

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

BIAS CORRECTIONS FOR TWO-STEP FIXED EFFECTS PANEL DATA ESTIMATORS

Iván Fernández-Val
Francis Vella

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP04/07

Bias Corrections for Two-Step Fixed Effects Panel Data Estimators

Iván Fernández-Val
Boston University

Francis Vella
Georgetown University

February 26, 2007

Abstract

This paper introduces bias-corrected estimators for nonlinear panel data models with both time invariant and time varying heterogeneity. These include limited dependent variable models with both unobserved individual effects and endogenous explanatory variables, and sample selection models with unobserved individual effects. Our two-step approach first estimates the reduced form by fixed effects procedures to obtain estimates of the time variant heterogeneity underlying the endogeneity/selection bias. We then estimate the primary equation by fixed effects including an appropriately constructed control function from the reduced form estimates as an additional explanatory variable. The fixed effects approach in this second step captures the time invariant heterogeneity while the control function accounts for the time varying heterogeneity. Since either or both steps might employ nonlinear fixed effects procedures it is necessary to bias adjust the estimates due to the incidental parameters problem. This problem is exacerbated by the two step nature of the procedure. As these two step approaches are not covered in the existing literature we derive the appropriate correction thereby extending the use of large-T bias adjustments to an important class of models. Simulation evidence indicates our approach works well in finite samples and an empirical example illustrates the applicability of our estimator.

JEL Classification: C23; J31; J51.

Keywords: Panel data; Two-Step Estimation; Endogenous Regressors; Fixed Effects; Bias; Union Premium.

Acknowledgment. This paper resulted from conversations with Whitney Newey. We acknowledge his valuable input. We are also grateful to Josh Angrist and the seminar participants at Boston University, MIT, and 2007 Winter Econometric Society Meetings for suggestions and comments.

1 Introduction

The incidental parameters problem arises in the estimation of nonlinear and dynamic panel models which include individual specific effects to control for unobserved time invariant heterogeneity (see, for example, Heckman, 1981, and Greene, 2002). A number of recent papers, surveyed in Arellano and Hahn (2005) and including Hahn and Kuersteiner (2002), Lancaster (2002), Woutersen (2002), Hahn and Kuersteiner (2003), Hahn and Newey (2004), and Carro (2006), provide a range of solutions, so-called large-T corrections, to reduce the incidental parameters bias in long panels. These papers derive the analytical expression for the bias (up to a certain order of T), which can be employed to adjust the biased fixed effects estimators. Numerical evidence suggests these adjustments eliminate, or significantly reduce, the bias even in short panels.

While the above papers collectively cover a large class of models, they do not handle endogeneity resulting from unobserved heterogeneity that contains a time varying component not eliminated via fixed effects style transformations. This kind of heterogeneity arises in a large class of models which are important for empirical investigations in economics. It includes, for instance, models with simultaneity, time varying omitted variables, measurement error, or sample selection. Accordingly we derive new bias corrections for models with multiple sources of endogeneity. Examples include static and dynamic models with limited endogenous regressors, panel data sample selection models, and limited dependent variable models with endogenous explanatory variables.

Below we discuss some papers which have analyzed some of the models we consider here. We differ from these existing studies in our treatment of the time invariant heterogeneity. We treat the unobserved individual effects as fixed effects (FE), potentially correlated with the explanatory variables, whereas previous investigations generally assume they are random effects (RE) which are distributed independently of the explanatory variables. RE estimation by-passes the incidental parameters problem by integrating out the individual effects. This approach, however, has three important shortcomings. First, the independence assumption is not compelling in many applications. In microeconomic studies, for instance, individual effects might capture variations in preferences or technology, and the explanatory variables are often choice variables determined optimally on the basis of this individual heterogeneity. Second, the RE estimators generally require an additional round of integration and this can complicate computation. Finally, the RE

procedures require parametric assumptions for the individual heterogeneity.

We provide a simple estimation procedure for a range of nonlinear panel data models with both time varying and time invariant endogeneity. Our two-step procedure first estimates the reduced form of the time variant heterogeneity underlying the endogeneity/selection bias by FE. We then estimate the primary equation by FE including an appropriately constructed control function. Since either or both steps might employ nonlinear FE procedures and the control function might be a nonlinear function of the first step individual effects, the incidental parameters problem arises. The existing bias corrections are not appropriate for these models, since they are not designed to account for the additional source of incidental parameters problem coming from the fixed effects estimation of the control function. We derive the appropriate bias correction.

The following section briefly describes some econometric models covered by our approach. Section 3 reviews some existing treatments of bias correction in non-linear panel data models and extends these corrections to two-step estimators. Section 4 gives asymptotic theory for the two-step bias corrected FE estimators. Section 5 provides simulation evidence and Section 6 contains an empirical example. Section 7 contains some concluding remarks.

2 Econometric Models

The leading class of econometric models we consider has the following two-index structure:

$$\begin{aligned} d_{it} &= f(I_{1it}), && \text{(Control Equation)} \\ y_{it} &= g(I_{2it}; d_{it}), && \text{(Primary Equation)} \end{aligned} \tag{1}$$

where

$$I_{1it} := \quad x_{it}'\beta_1 + z_{it}'\beta_2 + \alpha_{1i} + \epsilon_{1it}, \tag{2}$$

$$I_{2it} := \quad x_{it}'\theta_1 + \theta_2 d_{it} + \alpha_{2i} + \epsilon_{2it}, \tag{3}$$

and $f(\cdot)$ and $g(\cdot)$ are known functions. The endogenous variable of primary interest is y_{it}, and d_{it} is an endogenous explanatory variable or selection indicator. The predetermined explanatory variables are denoted by x_{it} and z_{it}; α_{1i} and α_{2i} are unobserved individual effects; and the disturbances are denoted by ϵ_{1it}, and ϵ_{2it}. The x_{it} appears in the conditional mean of each equation and $\beta_2 \neq 0$ ensures identification does not rely on distributional

assumptions. Lagged dependent variables may appear in each equation and these would be included in x_{it} or z_{it}.

Assumption 1: The idiosyncratic disturbances ϵ_{1it} and ϵ_{2it} are jointly normally distributed with variances σ_1^2 and σ_2^2, respectively; with a potentially non zero covariance σ_{12}; and

$$E[\epsilon_{jit}|x_i^t, z_i^t, \alpha_{1i}, \alpha_{2i}] = 0, \qquad \text{for } j = 1, 2;\ i = 1, ..., n;\ t = 1, ..., T, \tag{4}$$

where $x_i^t = [x_{i1}, ...x_{it}]$, and $z_i^t = [z_{i1}, ...z_{it}]$.[1] Note that we do not impose any condition on the joint distribution of α_{1i} and α_{2i}, given x_i^t and z_i^t. Assumption 1 indicates the endogeneity in the primary equation arises both through the correlation in the unobserved individual effects and also the contemporaneous correlation in the idiosyncratic errors.

To estimate the parameters from this model we propose the following strategy. We first estimate the reduced form control equation from which we construct the appropriate control function. The form of this control function depends on the type of censoring or selection, and this is generally captured by the nature of the dependent variable in the reduced form.[2] We then account for the endogeneity in the main equation by eliminating the first form, due to the $\alpha_2' s$, through the inclusion of individual fixed effects, and the second, due to the $\epsilon' s$, through the inclusion of the estimated control function. Estimation of the primary equation is based on,

$$y_{it} = g(x_{it}'\theta_1 + \theta_2 d_{it} + \rho(\lambda_{it}) + \alpha_{2i} + u_{2it}; d_{it}), \qquad \text{(Estimation Equation)} \tag{5}$$

where $\lambda_{it} := \lambda(d_{it}, x_{it}'\beta_1 + z_{it}'\beta_2 + \alpha_{1i})$ is the control function and ρ is the appropriate mapping.[3] The incidental parameters problem may arise in both steps and this is complicated by the inclusion of the control function that depends on the individual effects of the reduced form. We now briefly consider some leading examples of this general model.

[1] The distinction between the $x' s$ and the $y' s$ is somewhat superficial in that one could also supplement the model with the reduced form for x when it is correlated with the idiosyncratic disturbances.

[2] There is a large literature on the use of control functions to establish orthogonality conditions which would be otherwise violated in the presence of endogeneity or selection. In this paper we do not derive the control function for any particular model but we assume its existence and refer to the literature in which it has been developed.

[3] Our control function approach is computationally more attractive than alternative methods, like Full or Partial Maximum Likelihood of the system (1). Moreover, system estimators, although more efficient, are generally less robust to parametric assumptions than two-step procedures.

One important class of models is related to the sample selection procedure of Heckman (1979). This was extended to panels by Ridder (1990), Verbeek and Nijman (1992), and Vella and Verbeek (1999), under the assumption that the error components are RE. Wooldridge (1995) introduces a correlated RE estimator under alternative assumptions on the individual effects. A semi-parametric estimator with FE is proposed by Kyriazidou (1997). Our approach avoids the distributional assumptions for the unobservable individual effects assumed in the fully parametric approaches. We differ from Kyriazidou (1997, 2001) in that we impose less data restrictions and our estimator is easier to implement. We also allow for richer dynamics as the explanatory variables may be predetermined rather than strictly exogenous.[4] Panel data selection models under alternative selection rules, such as those considered by Vella and Verbeek (1999) with RE, can also be accommodated. Our approach encompasses models with censored endogenous regressors such as those considered by Heckman (1978) and Vella (1993) in the cross sectional context and by Vella and Verbeek (1999) in panels.

The primary equation in the above mentioned models is estimated by least squares methods. A second class of models follows the conditional MLE procedure of Smith and Blundell (1986) and Rivers and Vuong (1988), which has been extended to panels by Vella and Verbeek (1999) under the assumption of random error components. We extend this class of models by assuming FE and allowing for dynamic feedbacks in the primary and control equations.

3 Bias Corrections in Fixed Effects Two-Step Estimation

Consider a general nonlinear panel data model with a common parameter of interest θ_{10} and individual effects α_{1i0}, $i = 1, ..., n$. These parameters are the solutions to the following population optimization problem

$$(\theta_{10}, \{\alpha_{1i0}\}_{i=1}^{n}) = \arg \max_{\theta_1, \{\alpha_{1i}\}_{i=1}^{n}} E\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} g_1(w_{it}; \theta_{10}, \alpha_{1i0})\right], \tag{6}$$

[4]Gayle and Viauroux (2005) propose a semiparametric estimator for sample selection models with predetermined explanatory variables. Their estimator does not require normality in the selection equation, but imposes restrictions on the individual effects of this equation. Moreover, the parameter estimation is based on a three-step sieve method, which is computationally more difficult to implement.

where $g_1(\cdot)$ is some suitable criterion function, and w_{it} $(t = 1, ..., T; i = 1, ..., n)$ are the data observations including the covariates and endogenous variables, i.e., $w_{it} = (x_{it}, z_{it}, d_{it}, y_{it})$.[5] FE estimators for model parameters can be constructed by solving the corresponding sample analog

$$\hat{\alpha}_{1i}(\theta_1) = \arg\max_{\alpha_{1i}} \frac{1}{T}\sum_{t=1}^{T} g_1(w_{it}; \theta_1, \alpha_{1i}), \quad \hat{\theta}_1 = \arg\max_{\theta_1} \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} g_1(w_{it}; \theta_1, \hat{\alpha}_{1i}(\theta_1)), \quad (7)$$

where we first concentrate out the individual effects and then solve for θ_1.

Neyman and Scott (1948) show that nonlinear and dynamic FE estimators can be severely biased in short panels due to the incidental parameters problem. This problem arises because the unobserved individual effects are replaced by sample estimates, $\hat{\alpha}_{1i}(\theta_1)$. Since estimation of model parameters cannot be separated from the individual effects in these models, the estimation error of the individual effects contaminates the other parameter estimates. To see this, note that from the usual M-estimation properties, for $n \to \infty$ with T fixed,

$$\hat{\theta}_1 \xrightarrow{p} \theta_{1T} = \arg\min_{\theta_1} \lim_{n\to\infty} \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} g_1(w_{it}; \theta_1, \hat{\alpha}_{1i}(\theta_1)). \quad (8)$$

The probability limit $\theta_{1T} \neq \theta_{10}$ generally since $\hat{\alpha}_{1i}(\theta_{10}) \neq \alpha_{1i0}$; but $\theta_{1T} \to \theta_{10}$ as $T \to \infty$, since $\hat{\alpha}_{1i}(\theta_{10}) \to \alpha_{1i0}$. For smooth moment conditions, $\theta_{1T} = \theta_{10} + \frac{\mathcal{B}_1}{T} + O\left(\frac{1}{T^2}\right)$ for some $\mathcal{B}_1$. Then, by asymptotic normality of M-estimators, $\sqrt{nT}(\hat{\theta}_1 - \theta_{1T}) \to N(0, \Sigma_1)$ as $n \to \infty$, and therefore

$$\sqrt{nT}(\hat{\theta}_1 - \theta_{10}) = \sqrt{nT}(\hat{\theta}_1 - \theta_{1T}) + \sqrt{nT}\frac{\mathcal{B}_1}{T} + O_p\left(\sqrt{\frac{n}{T^3}}\right). \quad (9)$$

If T grows at the same rate as n the FE estimator, while consistent, has a limiting distribution which is not centered at the true parameter value. This large-T version of the incidental parameters problem invalidates inference based on the standard asymptotic distribution of the FE estimator.

Large-T bias corrections have been recently developed for nonlinear and dynamic FE estimators. After deriving an analytical expression for the leading term of the bias, $\mathcal{B}_1$, these corrections remove an estimate of $\mathcal{B}_1$ from the original FE estimator. This

[5] To simplify the exposition we start the analysis with the case where w_{it} are independent across i and t. The results are extended below to cases where there might be serial correlation across t.

correction reduces the order of the bias to $O(T^{-2})$, and produces an estimator that has an asymptotic distribution centered at the true parameter value under asymptotic sequences where $n = o(T^3)$; see, e.g., Hahn and Kuersteiner (2002) and Hahn and Newey (2004). Numerical evidence suggests that these corrections remove most of the bias even in short panels.

3.1 Two-step analytical bias correction.

We now extend the large-T bias correction methods to two-step FE estimators where the presence of endogeneity or sample selection corrupts the orthogonality of the moment conditions of the OLS or ML estimators. This orthogonality can be reestablished, however, via the inclusion of the appropriate control function. We consider models where the parameters are identified by population modified problems of the form

$$(\theta_{20}, \{\alpha_{2i0}\}_{i=1}^n) = \arg \max_{\theta_2, \{\alpha_{2i}\}_{i=1}^n} E\left[\frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T g_2(w_{it}, \lambda_{it}; \theta_{20}, \alpha_{2i0})\right], \tag{10}$$

where λ_{it} is the control function, and $g_2(\cdot)$ is some appropriate objective function, e.g., least squares, likelihood, or GMM criterion function. The control function has a parametric form, that is $\lambda_{it} := \lambda(w_{it}, \theta_{10}, \alpha_{1i0})$ where $\lambda(\cdot)$ is a known function, but generally depends on the unknown parameters θ_{10} and α_{1i0}. These parameters are estimated in a first stage from a (possibly) nonlinear panel model with parameters identified from some optimization problem as in (6).

FE estimates of the second stage parameters are obtained via the following optimization problem in the sample

$$\hat{\alpha}_{2i}(\theta_2) = \arg\max_{\alpha_{2i}} \frac{1}{T}\sum_{t=1}^T g_2(w_{it}, \tilde{\lambda}_{it}; \theta_2, \alpha_{2i}), \;\; \hat{\theta}_2 = \arg\max_{\theta_2} \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T g_2(w_{it}, \tilde{\lambda}_{it}; \theta_2, \hat{\alpha}_{2i}(\theta_2)), \tag{11}$$

where $\tilde{\lambda}_{it} = \lambda(w_{it}; \tilde{\theta}_1, \hat{\alpha}_{1i}(\tilde{\theta}_1))$. To simplify the exposition and focus on the new sources of incidental parameters bias, we assume that $\tilde{\theta}_1 = \theta_{10}$ in the following discussion. In general it suffices that $\tilde{\theta}_1 = \theta_{10} + O_p(1/\sqrt{nT})$, which holds if $\tilde{\theta}_1$ is a large-T bias corrected estimator of θ_{10} and $n = o(T^3)$. We show that the resulting FE estimators are biased if the control function is nonlinear in the first stage individual effects or the second stage is nonlinear. An additional round of bias correction is therefore needed in the second step.

The existing correction methods for one-step procedures are generally not valid to carry out the bias correction in the second stage. Some of the explanators are FE estimates that depend on the first-stage individual effects, introducing additional incidental parameters bias. The issue here is similar to the two-step variance estimation (see, for example, Newey 1984). Thus, as $n \to \infty$, we have

$$\hat{\theta}_2 \xrightarrow{p} \theta_{2T} = \arg\min_{\theta_2} \lim_{n\to\infty} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} g_2(w_{it}, , \tilde{\lambda}_{it}; \theta_2, \hat{\alpha}_{2i}(\theta_2)). \tag{12}$$

Here $\theta_{2T} \neq \theta_{20}$, not only because $\hat{\alpha}_{2i}(\theta_{20}) \neq \alpha_{2i0}$, but also because $\tilde{\lambda}_{it} \neq \lambda_{it}$. To see the second inequality, note that $\lambda_{it}(w_{it}; \theta_{10}, \hat{\alpha}_{1i}(\theta_{10})) \neq \lambda_{it}(w_{it}; \theta_{10}, \alpha_{1i0})$ since $\hat{\alpha}_{1i}(\theta_{10}) \neq \alpha_{1i0}$. Moreover, the additional source of the bias is not related to the estimation of θ_{10}, since we are evaluating this parameter at its true value. As a result of the previous analysis, a bias expression similar to $\mathcal{B}_1$ would be not valid because it only would account for the bias coming from the estimation of the α_{2i0}'s .

We derive the general expression for the bias of the second-step FE estimator using stochastic expansions that explicitly account for the randomness introduced by the estimation of the control function. To describe these expansions we first introduce some additional notation. Let

$$u_{2it}(\theta, \alpha) := \frac{\partial g_2(w_{it}, \lambda_{it}; \theta, \alpha)}{\partial \theta}, \quad v_{2it}(\theta, \alpha) := \frac{\partial g_2(w_{it}, \lambda_{it}; \theta, \alpha)}{\partial \alpha}, \tag{13}$$

and additional subscripts denote partial derivatives, e.g., $u_{2it\theta}(\theta, \alpha) := \partial u_{2it}(\theta, \alpha)/\partial \theta'$. For notational convenience the arguments are omitted when the expressions are evaluated at the true parameter value, i.e., $v_{2it} := v_{2it}(\theta_{20}, \alpha_{2i0})$.

Using an asymptotic expansion for the two-step FE estimator (see Appendix for details of the derivation and additional notation) we have, as $n, T \to \infty$

$$T(\hat{\theta}_2 - \theta_{20}) \xrightarrow{p} T(\theta_{2T} - \theta_{20}) = -\mathcal{J}_2^{-1} b_2 := \mathcal{B}_2. \tag{14}$$

Here, $\mathcal{J}_2 = E_n \left[E_T \left[u_{2it\theta} \right] - E_T \left[u_{2it\alpha} \right] E_T \left[v_{2it\theta} \right] / E_T \left[v_{2it\alpha} \right] \right]$ is the limit of the Jacobian of the estimating equation for θ_2, where $E_T \left[f_{it} \right] := \lim_{T\to\infty} \sum_{t=1}^{T} f_{it}/T$, for any function $f_{it} := f(w_{it})$, and $E_n \left[f_i \right] := \lim_{n\to\infty} \sum_{i=1}^{n} f_i/n$, for any function $f_i = E_T[f_{it}]$; and b_2, the bias of this estimating equation, takes the form

$$\begin{aligned} b_2 \;=\; & E_n \Big\{ E_T \left[u_{2it\alpha} \psi_{2it} \right] + E_T \left[u_{2it\alpha} \right] \beta_{2i} + \frac{1}{2} \sigma_{2i}^2 E_T \left[u_{2it\alpha\alpha} \right] + E_T \left[u_{2it\lambda\alpha} \lambda_{it\alpha} \right] \sigma_{12i} \\ & + E_T \left[u_{2it\lambda} \left(\lambda_{it\alpha} (\psi_{1it} + \beta_{1i}) + \frac{1}{2} \lambda_{it\alpha\alpha} \sigma_{1i}^2 \right) \right] + \frac{1}{2} E_T \left[u_{2it\lambda\lambda} \lambda_{it\alpha}^2 \right] \sigma_{1i}^2 \Big\}. \end{aligned} \tag{15}$$

Thus, the bias of the two-step estimating equation, in addition to the three components of a one-step estimating equation (first three terms) derived by Hahn and Newey (2004), has three new components arising from the FE estimation of the control function and the nonlinearity of the second stage. Recall that the first terms arise from the randomness of $\hat{\alpha}_{2i}$ if the primary equation is nonlinear in these individual effects. The new terms come from the correlation between the estimators of the individual effects in the first and second stages σ_{12i} (fourth term), arising because both stages use the same individuals to estimate these effects; the asymptotic bias of the FE estimator of the control function $\tilde{\lambda}_{it}$ (fifth term), coming from the nonlinearity of this function in the first stage individual effects; and the nonlinearity of the second stage in the control function (last term). In the panel sample selection model, for example, the only term of the bias that does not vanish is the fifth one, since the primary equation is linear in the individual effects and control function, but the control function, which is the inverse Mills ratio, is nonlinear in the individual effects of the selection equation.

The analytical expression for $\mathcal{B}_2$ can be used to construct analytical (closed form) bias corrected estimators for the second stage parameters and other functions of parameters and individual effects, such as marginal effects. A bias corrected estimator for model parameters can be formed as

$$\tilde{\theta}_2 = \hat{\theta}_2 - \frac{\hat{\mathcal{B}}_2}{T}, \tag{16}$$

where $\hat{\mathcal{B}}_2$ is an estimator of $\mathcal{B}_2$ constructed using sample analogs of the components of $\mathcal{J}_2$ and b_2. Moreover, since $\hat{\mathcal{B}}_2$ generally depends on θ_{20}, we have that $\hat{\mathcal{B}}_2 = \hat{\mathcal{B}}_2(\hat{\theta}_2)$. Iterated bias corrections can be constructed similarly to those for one-step estimators by solving $\tilde{\theta}_2^{\infty} = \hat{\theta}_2 - \hat{\mathcal{B}}_2(\tilde{\theta}_2^{\infty})/T$. Note that the analytical expression for the bias in (14) assumes that the control function is constructed using bias corrected estimators of the first stage parameters θ_{10}.[6]

[6] A similar expression for the bias can be derived for the two-step estimator that uses a control function constructed from uncorrected first stage estimates. Our approach has the advantages that it yields bias corrected estimates of both the control and primary equations, and that the bias expressions in the second stage involve fewer terms.

3.2 Two-step Jackknife bias correction.

An alternative bias correction method that does not require the analytical expression for the bias is based on leave-one-observation-out Jackknife. This procedure obtains T different FE estimators $\hat{\theta}_2^{(t)}$, $t = 1, ..., T$; where each $\hat{\theta}_2^{(t)}$ is computed excluding the t-th observation for each individual. The Jackknife bias corrected estimator then takes the form

$$\tilde{\theta}_2^J := \hat{\theta}_2 - \frac{T-1}{T}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{\theta}_2^{(t)} - \hat{\theta}_2\right), \tag{17}$$

where $\hat{\theta}_2$ is the uncorrected (in both steps) two-step FE estimator of θ_{20}. Following Hahn and Newey (2004), we can show that the second term automatically removes the leading term of the bias of $\hat{\theta}_2$ as $n, T \to \infty$. This method, however, is not directly applicable to models with predetermined regressors, and can be computationally very intensive when the model is highly nonlinear and T is moderately large.

3.3 Dynamic Models

We now derive the analytical expression of the bias for dynamic FE estimators. We consider models where the dynamics are fully captured by lags of the endogenous variables and the explanatory variables can be serially correlated. Moreover, we only assume that the explanatory variables are predetermined, allowing for feedback from the dependent variables to the explanators. This is an important departure from previous modeling approaches in linear and nonlinear panel data estimators that typically assume the explanatory variables to be strictly exogenous.[7] These models include, for example, panel selection models with predetermined regressors and individual effects, and dynamic conditional maximum likelihood estimators with individual effects.

The bias expressions are similar to the fully static case, but include additional terms due to the serial correlation of the observations (Hurwicz-type terms) that capture all the possible dynamic feedbacks from the dependent variables to the regressors. Let $\bar{E}_T[f_{it}g_{is}] := \sum_{j=-\infty}^{\infty} E_T[f_{it}g_{i,t-j}]$, for any functions $f_{it} = f(w_{it})$ and $g_{is} = g(w_{is})$. Then, standard higher-order asymptotic expansions give the following expression for the bias

[7] See Wooldridge (2001), Honoré and Lewbel (2002), and Arellano and Carrasco (2003) for examples of one-step panel data estimators with predetermined regressors.

of the estimating equation (see Appendix for details of the derivation and additional notation), as $n, T \to \infty$,

$$\begin{aligned}\bar{b}_{2i} &= \bar{E}_T\left[u_{2it\alpha}\psi_{2is}\right] + E_T\left[u_{2it\alpha}\right]\bar{\beta}_{2i} + \frac{1}{2}\bar{\sigma}_{2i}^2 E_T\left[u_{2it\alpha\alpha}\right] + E_T\left[u_{2it\lambda\alpha}\lambda_{it\alpha}\right]\bar{\sigma}_{12i} \\ &\quad + \bar{E}_T\left[u_{2it\lambda}\lambda_{it\alpha}\psi_{1is}\right] + E_T\left[u_{2it\lambda}\left(\lambda_{it\alpha}\bar{\beta}_{1i} + \frac{1}{2}\lambda_{it\alpha\alpha}\bar{\sigma}_{1i}^2\right)\right] + \frac{1}{2}E_T\left[u_{2it\lambda\lambda}\lambda_{it\alpha}^2\right]\bar{\sigma}_{1i}^2,\end{aligned} \tag{18}$$

and

$$\bar{\mathcal{B}}_2 := -\mathcal{J}_2^{-1}\bar{b}_2, \tag{19}$$

for the bias of the two-step FE estimator. Here $\bar{b}_2 := E_n\left[\bar{b}_{2i}\right]$, and the limit Jacobian $\mathcal{J}_2$ is the same as for the static estimator. This expression generalizes the bias formula for FE estimators in dynamic nonlinear panel models derived in Hahn and Kuersteiner (2003) to two-step estimators. In models defined by conditional moment restrictions, such as the examples in Section 2, the terms involving the spectral expectation $\bar{E}_T$ can be replaced for $\tilde{E}_T$ if the regressors are predetermined, where $\tilde{E}_T[f_{kit}\psi_{lis}] = \sum_{j=0}^{\infty} E_T[f_{kit}\psi_{li,t-j}]$ and $\tilde{E}_T[\psi_{kit}\psi_{lis}] = E_T[\psi_{kit}\psi_{lit}]$ for $k, l \in \{1, 2\}$; whereas $\bar{E}_T$ can be replaced by E_T everywhere if the model is fully static with exogenous regressors, see equation (15).

4 Asymptotic Theory

To guarantee the validity of the higher-order expansions used to derive the expression of the bias and to establish the validity of the bias corrections in large samples, we impose the following conditions:

Condition 1 (Sampling) *(i) $n, T \to \infty$ such that $T = O(n)$. (ii) For each i, $w_i := \{w_{it}\}_{t=1,2,...}$ is a stationary mixing sequence. Let $\mathcal{A}_t^i = \sigma(w_{it}, w_{i,t-1}, ...)$, $\mathcal{D}_t^i = \sigma(w_{it}, w_{i,t+1}, ...)$, and $a_i(m) = \sup_t \sup_{A\in\mathcal{A}_t^i, D\in\mathcal{D}_{t+m}^i} |P(A\cap B) - P(A)P(D)|$. Then, $\sup_i |a_i(m)| \leq Ca^m$ for some a such that $0 < a < 1$ and some $C > 0$. (iii) $\{w_i\}_{i=1,2,...}$ are independent across i.*

Let $\mathcal{J}_1 := E_n\left[E_T\left[u_{1it\theta}\right] - E_T\left[u_{1it\alpha}\right]E_T\left[v_{1it\theta}\right]/E_T\left[v_{1it\alpha}\right]\right]$, $\bar{\Omega}_1 := E_n\left[\bar{E}_T\left[U_{1it}'U_{1is}\right]\right]$, $U_{1it} := u_{1it} + E_T\left[u_{1it\alpha}\right]\psi_{1it}$, $\psi_{1it} := -v_{1it}/E_T\left[v_{1it\alpha}\right]$, and $\bar{\sigma}_{1i}^2 := \bar{E}_T\left[\psi_{1it}\psi_{1is}\right]$. Here $v_{1it}(\theta, \alpha) := \partial g_1(w_{it}; \theta, \alpha)/\partial\alpha$, $u_{1it}(\theta, \alpha) := \partial g_1(w_{it}; \theta, \alpha)/\partial\theta$, additional subscripts denote partial derivatives, and the arguments are omitted when the expressions are evaluated at the true parameter value.

Condition 2 (First Stage) *(i) For each i and n, $(\theta_{10}, \alpha_{1i0}) \in int\ \Gamma_1$, and the parameter space Γ_1 is a convex, compact subset of $\Re^{p_1}$. (ii) For each $\eta > 0$ and n, $\inf_i \left|g_{1i}(\theta_{10}, \alpha_{1i0}) \ - \sup_{\{(\theta_1,\alpha_1)\in\Gamma_1:|(\theta_1,\alpha_1)-(\theta_{10},\alpha_{1i0})|>\eta\}} g_{1i}(\theta_1, \alpha_1)\right| > 0$, where $g_{1i}(\theta_1, \alpha_1) := E_T\left[g_1(w_{it}; \theta_1, \alpha_1)\right]$. (iii) Let $\nu = (\nu_1, ..., \nu_{p_1})$ be a vector of non-negative integers, $|\nu| = \sum_{j=1}^{p_1} \nu_j$, and $\nabla_\nu g_1(w_{it}; \gamma) = \partial^{|\nu|} g_1(w_{it}; \gamma)/(\partial\gamma_1^{\nu_1} ... \gamma_{p_1}^{\nu_{p_1}})$, with $\gamma := (\theta_1, \alpha_1)$; then, there exists a function $M_1(w_{it})$, such that $|\nabla_\nu g_1(w_{it}; \gamma_1) - \nabla_\nu g_1(w_{it}; \gamma_2)| \leq M_1(w_{it})|\gamma_1 - \gamma_2|$ for all $\gamma_1, \gamma_2 \in \Gamma_1$; $\sup_{\gamma\in\Gamma_1} |\nabla_\nu g_1(w_{it}; \gamma)| \leq M_1(w_{it})$ for $|\nu| \leq 5$, and $\sup_i E_T[|M_1(w_{it})|^{10q_1+12+\upsilon}] < \infty$ for some integer $q_1 \geq p_1/2 + 2$ and some $\upsilon > 0$. (iv) $\mathcal{J}_1$ is negative definite and finite. (v) $\bar{\Omega}_1$ is positive definite and finite. (vii) For each n, $0 < \inf_i \bar{\sigma}_{1i}^2$ and $\sup_i \bar{\sigma}_{1i}^2 < \infty$.*

These conditions, extracted from Hahn and Kuersteiner (2003), guarantee the validity of the first stage bias correction for general one-step dynamic FE M-estimators. The stationarity assumption is restrictive as it rules out, for example, time dummies and other deterministic trend components as explanatory variables. How to extend the large-T bias corrections to allow for non-stationary variables is an open question beyond the scope of this paper. In our empirical example presented in Section 6, however, we check that the results are robust to the exclusion of the time dummies.

Let

$$\hat{v}_{1it}(\theta_1) := v_{1it}(\theta_1, \hat{\alpha}_{1i}(\theta_1)),\ \hat{u}_{1it}(\theta_1) := u_{1it}(\theta_1, \hat{\alpha}_{1i}(\theta_1)), \tag{20}$$

and additional subscripts denote partial derivatives, e.g., $\hat{v}_{1it\alpha}(\theta_1) := v_{1it\alpha}(\theta_1, \hat{\alpha}_{1i}(\theta_1))$. Let

$$\hat{\psi}_{1it}(\theta_1) := -\hat{v}_{1it}(\theta_1)/\hat{E}_T\left[v_{1it\alpha}(\theta_1)\right],\ \ \hat{\bar{\sigma}}_{1i}^2(\theta_1) := \hat{\bar{E}}_{T,m}\left[\hat{\psi}_{1it}(\theta_1)\hat{\psi}_{1is}(\theta_1)\right], \tag{21}$$

where $\hat{E}_T[f_{it}] := \sum_{t=1}^T f_{it}/T$ and $\hat{\bar{E}}_{T,m}[f_{it}g_{is}] := \sum_{j=-m}^m \sum_{t=\max(1,j)}^{\min(T,T+j)} f_{it}g_{i,t-j}/(T-j)$, for any functions $f_{it} = f(w_{it})$ and $g_{it} = g(w_{it})$. The parameter m is a bandwidth parameter that needs to be chosen such that $m/T^{1/2} \to \infty$ as $T \to \infty$; see Hahn and Kuersteiner (2003). $\hat{\psi}_{1it}(\theta_1)$ and $\hat{\bar{\sigma}}_{1i}^2(\theta_1)$ are estimators of the influence function and asymptotic variance of $\hat{\alpha}_{1i}(\theta_1)$ as T grows, respectively. Let

$$\begin{aligned}
&\hat{\bar{\beta}}_{1i}(\theta_1) := -\hat{E}_T\left[v_{1it\alpha}(\theta_1)\right]^{-1}\left\{\hat{\bar{E}}_{T,m}[\hat{v}_{1it\alpha}(\theta_1)\hat{\psi}_{1is}(\theta_1)] + \hat{E}_T\left[\hat{v}_{1it\alpha\alpha}(\theta_1)\right]\hat{\bar{\sigma}}_{1i}^2(\theta_1)/2\right\},\\
&\hat{\mathcal{J}}_{1i}(\theta_1) := \hat{E}_T\left[\hat{u}_{1it\theta}(\theta_1)\right] - \hat{E}_T\left[\hat{u}_{1it\alpha}(\theta_1)\right]\hat{E}_T\left[\hat{v}_{1it\theta}(\theta_1)\right]/\hat{E}_T\left[\hat{v}_{1it\alpha}(\theta_1)\right],\\
&\hat{\bar{b}}_{1i}(\theta_1) := \hat{\bar{E}}_{T,m}\left[\hat{u}_{1it\alpha}(\theta_1)\hat{\psi}_{1is}(\theta_1)\right] + \hat{E}_T\left[\hat{u}_{1it\alpha}(\theta_1)\right]\hat{\bar{\beta}}_i(\theta_1) + \hat{E}_T\left[\hat{u}_{1it\alpha\alpha}(\theta_1)\right]\hat{\bar{\sigma}}_{1i}^2(\theta_1)/2.
\end{aligned} \tag{22}$$

Here, $\hat{\bar{\beta}}_{1i}(\theta_1)$ is an estimator of the higher-order asymptotic bias of $\hat{\alpha}_{1i}(\theta_1)$ from a stochastic expansion as T grows; whereas $\hat{\mathcal{J}}_{1i}(\theta_1)$ and $\hat{\bar{b}}_{1i}(\theta_1)$ are estimators of the Jacobian and the asymptotic bias of the estimating equation of θ_1 for individual i. A bias corrected estimator of the first-step FE estimator can be formed as

$$\tilde{\theta}_1 = \hat{\theta}_1 - \hat{\bar{\mathcal{B}}}_1(\hat{\theta}_1)/T, \tag{23}$$

where $\hat{\bar{\mathcal{B}}}_1(\theta_1) := -\hat{E}_n\left[\hat{\mathcal{J}}_{1i}(\theta_1)\right]^{-1}\hat{E}_n\left[\hat{\bar{b}}_{1i}(\theta_1)\right]$ is an estimator of the bias of $\hat{\theta}_1$, where $\hat{E}_n[f_i] := \sum_{i=1}^n f_i/n$ for any function $f_i = \hat{E}_T[f_{it}]$; and $\hat{\theta}_1$ is the FE estimator of θ_{10}.

Lemma 1 *Assume that Conditions 1 and 2 hold, and $m \to \infty$ such that $m/T^{1/2} \to 0$. Then, we have*

$$\sqrt{nT}\left(\tilde{\theta}_1 - \theta_{10}\right) = \varphi_1 + \frac{1}{\sqrt{nT}}R_1 \xrightarrow{d} \mathcal{N}(0, \mathcal{J}_1^{-1}\bar{\Omega}_1\mathcal{J}_1^{-1}), \tag{24}$$

where $\varphi_1 := \sum_{i=1}^n\sum_{t=1}^T \varphi_{1it}/\sqrt{nT}$, $\varphi_{1it} := -\mathcal{J}_1^{-1}U_{1it}$, and $R_1 = o_p(\sqrt{nT})$. Let $\tilde{\alpha}_{1i} := \hat{\alpha}_{1i}(\tilde{\theta}_1)$, then

$$\tilde{\alpha}_{1i} = \alpha_{1i0} + \frac{1}{\sqrt{T}}\psi_{1i} - \frac{1}{\sqrt{nT}}E_T\left[v_{1it\alpha}\right]^{-1}E_T\left[v_{1it\theta}\right]\varphi_1 + \frac{1}{T}\bar{\beta}_{1i} + \frac{1}{T^{3/2}}R_{1i}, \tag{25}$$

where $\psi_{1i} := \sum_{t=1}^T \psi_{1it}/\sqrt{T}$, $\bar{\beta}_{1i} := -E_T\left[v_{1it\alpha}\right]^{-1}\left\{\bar{E}_T\left[v_{1it\alpha}\psi_{1is}\right] + \bar{\sigma}_{1i}^2 E_T\left[v_{1it\alpha\alpha}\right]/2\right\}$, and $\max_{1\le i\le n} R_{1i} = o_p(\sqrt{T})$.

Proof. See appendix. ■

Remark 1 *Note that for static models with exogenous regressors we can set $m = 0$. For conditional moments models with predetermined regressors we can replace $\hat{\bar{E}}_{T,m}$ for $\hat{\tilde{E}}_{T,m}$, where $\hat{\tilde{E}}_{T,m}[f_{it}\psi_{1is}] := \sum_{j=0}^m\sum_{t=j+1}^T f_{it}\psi_{1i,t-j}/(T-j)$ and $\hat{\tilde{E}}_{T,m}[\psi_{1it}\psi_{1is}] := \hat{E}_T\left[\psi_{1it}^2\right]$.*

Condition 3 (Control function) *Let $\nu = (\nu_1, ..., \nu_{p_2})$ be a vector of non-negative integers, and $\lambda_{it}(\gamma) = \lambda(w_{it};\gamma)$ be the control function , with $\gamma = (\theta_1, \alpha_1)$. There exists a function $M_\lambda(w_{it})$, such that $|\nabla_\nu\lambda_{it}(\gamma_1) - \nabla_\nu\lambda_{it}(\gamma_2)| \le M_\lambda(w_{it})|\gamma_1 - \gamma_2|$ for all $\gamma_1, \gamma_2 \in \Gamma_1$; $\sup_{\gamma\in\Gamma_1}|\nabla_\nu\lambda_{it}(\gamma)| \le M_\lambda(w_{it})$ for $|\nu| \le 5$, and $\sup_i E_T[|M_\lambda(w_{it})|^{10q_1+12+\upsilon}] < \infty$ for some integer $q_1 \ge p_1/2 + 2$ and some $\upsilon > 0$.*

This condition guarantees the existence of higher-order expansions for the fixed effects estimators of the control functions in a neighborhood of their true values, and the uniform

convergence of the remainder terms in these expansions. In most applications the control functions are generalized residuals, see Gourieroux, Monfort, Renault, and Trognon (1987), and this condition follows from Condition 2.

Lemma 2 *Assume that Conditions 1, 2, and 3 hold, and $m \to \infty$ such that $m/T^{1/2} \to 0$. Then, we have*

$$\tilde{\lambda}_{it} = \lambda_{it} + \tilde{\psi}_{it}^{\lambda}/\sqrt{T} + \bar{\beta}_{it}^{\lambda}/T + R_{it}^{\lambda}/T^{3/2}, \tag{26}$$

where $\tilde{\lambda}_{it} := \lambda(w_{it}; \tilde{\theta}_1, \hat{\alpha}_{1i}(\tilde{\theta}_1))$, $\tilde{\psi}_{it}^{\lambda} := \lambda_{it\alpha}\psi_{1i} + \mathcal{J}_{it}^{\lambda}\varphi_1/\sqrt{n} = o_{up}\left(T^{1/10+\upsilon}\right)$, $\mathcal{J}_{it}^{\lambda} := \lambda_{it\theta} - E_T\left[v_{1it\alpha}\right]^{-1} E_T\left[v_{1it\theta}\right]\lambda_{it\alpha}$, $\bar{\beta}_{it}^{\lambda} := \lambda_{it\alpha}\bar{\beta}_{1i} + \lambda_{it\alpha\alpha}\psi_{1i}^2/2 = o_{up}\left(T^{2/10+2\upsilon}\right)$, and $\max_{1\leq i\leq n} R_{it}^{\lambda} = o_p(\sqrt{T})$ P_w-almost everywhere, for some $0 < \upsilon < (100q_1 + 120)^{-1}$.

Proof. See Appendix. ■

Let $\mathcal{J}_2 := E_n\left[E_T\left[u_{2it\theta}\right] - E_T\left[u_{2it\alpha}\right]E_T\left[v_{2it\theta}\right]/E_T\left[v_{2it\alpha}\right]\right]$, $\bar{\Omega}_2 := E_n\left[\bar{E}_T\left[U_{2it}'U_{2is}\right]\right]$, $U_{2it} := u_{2it} + E_T\left[u_{2it\alpha}\right]\psi_{2it} + E_T\left[u_{2it\lambda}\lambda_{it\alpha}\right]\psi_{1it} + E_T\left[u_{2it\lambda}\mathcal{J}_{it}^{\lambda}\right]\varphi_{1it}$, $\psi_{2it} := -\left\{v_{2it} + E_T\left[v_{2it\lambda}\lambda_{it\alpha}\right]\psi_{1it} + E_T\left[v_{2it\lambda}\mathcal{J}_{it}^{\lambda}\right]\varphi_{1it}/\sqrt{n}\right\}/E_T\left[v_{2it\alpha}\right]$, and $\bar{\sigma}_{2i}^2 := \bar{E}_T\left[\psi_{2it}\psi_{2is}\right]$.

Condition 4 (Second Stage) *(i) For each i and n, $(\theta_{20}, \alpha_{2i0}) \in int\ \Gamma_2$, and the parameter space Γ_2 is a convex, compact subset of $\Re^{p_2}$. (ii) For each $\eta > 0$ and n, $\inf_i \left|g_{2i}(\theta_{20}, \alpha_{2i0}) - \sup_{\{(\theta_2,\alpha_2)\in\Gamma_2:|(\theta_2,\alpha_2)-(\theta_{20},\alpha_{2i0})|>\eta\}} g_{2i}(\theta_2, \alpha_2)\right| > 0$, where $g_{2i}(\theta_2, \alpha_2) := E_T\left[g_2(w_{it}, \lambda_{it}; \theta_2, \alpha_2)\right]$. (iii) Let $\nu = (\nu_\lambda, \nu_1, ..., \nu_{p_2})$ be a vector of non-negative integers and $\gamma := (\theta_2, \alpha_2)$; then, there exists a function $M_2(w_{it}, \lambda_{it})$, such that $|\nabla_\nu g_2(w_{it}, \lambda_{it}; \gamma_1) - \nabla_\nu g_2(w_{it}, \lambda_{it}; \gamma_2)| \leq M_2(w_{it}, \lambda_{it})|\gamma_1 - \gamma_2|$ for all $\gamma_1, \gamma_2 \in \Gamma_2$; $\sup_{\gamma\in\Gamma_2} |\nabla_\nu g_2(w_{it}, \lambda_{it}; \gamma)| \leq M_2(w_{it}, \lambda_{it})$ for $|\nu| \leq 5$, and $\sup_i E_T[|M_2(w_{it}, \lambda_{it})|^{10q_2+12+\upsilon}] < \infty$ for some integer $q_2 \geq p_2/2 + 2$ and some $\upsilon > 0$. (iv) $\mathcal{J}_2$ is negative definite and finite. (v) $\bar{\Omega}_2$ is positive definite and finite. (vii) For each n, $0 < \inf_i \bar{\sigma}_{2i}^2$ and $\sup_i \bar{\sigma}_{2i}^2 < \infty$.*

These conditions guarantee the validity of the second stage bias correction extending the conditions in Hahn and Kuersteiner (2003) to two-step dynamic FE M-estimators. Condition 4 guarantees parameter identification based on time series variation, but it does not explicitly imposes exclusion restrictions in the first and second stages. Parameter identification can be achieved, in principle, by non-linearities in the control and primary equations, or by non-linearities in the control function. To avoid such an identification scheme an exclusion restriction should be imposed in the primary equation for each source of time varying endogeneity.

Theorem 1 *Under Conditions 1, 2, 3, and 4, we have*

$$\sqrt{nT}\left(\hat{\theta}_2 - \theta_{20}\right) \xrightarrow{d} \mathcal{N}\left(\sqrt{\rho}\,\bar{\mathcal{B}}_2, \mathcal{J}_2^{-1}\bar{\Omega}_2\mathcal{J}_2^{-1}\right), \tag{27}$$

where $\rho = \lim_{n,T\to\infty} n/T$ *and* $\bar{\mathcal{B}}_2 := -\mathcal{J}_2^{-1}\bar{b}_2$. *The general expression for* $\bar{b}_2$ *is given in (19) and simplifies to the expression in (14) for static models with exogenous regressors.*

Proof. See Appendix. ■

Let

$$\begin{aligned}
&\tilde{\lambda}_{it\alpha} := \lambda_\alpha(w_{it};\tilde{\theta}_1,\hat{\alpha}_{1i}(\tilde{\theta}_1)),\ \tilde{\lambda}_{it\theta} := \lambda_\theta(w_{it};\tilde{\theta}_1,\hat{\alpha}_{1i}(\tilde{\theta}_1)),\ \tilde{v}_{2it}(\theta_2,\alpha_2) := v_2(w_{it},\tilde{\lambda}_{it};\theta_2,\alpha_2),\\
&\tilde{u}_{2it}(\theta_2,\alpha_2) := u_2(w_{it},\tilde{\lambda}_{it};\theta_2,\alpha_2),\ \hat{\tilde{v}}_{2it}(\theta_2) := \tilde{v}_{2it}(\theta_2,\hat{\alpha}_{2i}(\theta_2)),\ \hat{\tilde{u}}_{2it}(\theta_2) := \tilde{u}_{2it}(\theta_2,\hat{\alpha}_{2i}(\theta_2)),
\end{aligned} \tag{28}$$

and additional subscripts denote partial derivatives. Let

$$\begin{aligned}
&\hat{\psi}_{2it}(\theta_2) := -\left\{\hat{\tilde{v}}_{2it}(\theta_2) + \hat{E}_T\left[\hat{\tilde{v}}_{2it\lambda}(\theta_2)\tilde{\lambda}_{it\alpha}\right]\hat{\psi}_{1it}(\tilde{\theta}_1) + \hat{E}_T\left[\hat{\tilde{v}}_{2it\lambda}(\theta_2)\tilde{\mathcal{J}}_{it}^{\lambda}\right]\tilde{\varphi}_{1it}\right\}/\hat{E}_T\left[\tilde{v}_{2it\alpha}(\theta_2)\right],\\
&\tilde{\mathcal{J}}_{it}^{\lambda} := \tilde{\lambda}_{it\theta} - \hat{E}_T\left[\hat{v}_{1it\alpha}\left(\tilde{\theta}_1\right)\right]^{-1}\hat{E}_T\left[\hat{v}_{1it\theta}\left(\tilde{\theta}_1\right)\right]\tilde{\lambda}_{it\alpha},\\
&\tilde{\varphi}_{1it} := -\hat{E}_n\left[\hat{\mathcal{J}}_{1i}\left(\tilde{\theta}_1\right)\right]^{-1}\left\{\hat{u}_{1it}\left(\tilde{\theta}_1\right) + \hat{E}_T\left[\hat{u}_{1it\alpha}\left(\tilde{\theta}_1\right)\right]\hat{\psi}_{1it}\left(\tilde{\theta}_1\right)\right\},\\
&\hat{\tilde{\sigma}}_{2i}^2(\theta_2) := \hat{\tilde{E}}_{T,m}\left[\hat{\psi}_{2it}(\theta_2)\hat{\psi}_{2is}(\theta_2)\right],\ \hat{\tilde{\sigma}}_{12i}^2(\theta_2) := \hat{\tilde{E}}_{T,m}\left[\hat{\psi}_{2it}(\theta_2)\hat{\psi}_{1is}(\tilde{\theta}_1)\right].
\end{aligned} \tag{29}$$

Here, $\hat{\psi}_{2it}(\theta_2)$ and $\hat{\tilde{\sigma}}_{2i}^2(\theta_2)$ are estimators of the influence function and asymptotic variance of $\hat{\alpha}_{2i}(\theta_2)$ as T grows, and $\hat{\tilde{\sigma}}_{12i}^2(\theta_2)$ is an estimator of the asymptotic covariance between $\hat{\alpha}_{1i}(\tilde{\theta}_1)$ and $\hat{\alpha}_{2i}(\theta_2)$ as T grows. Let

$$\begin{aligned}
\hat{\tilde{\beta}}_{2i}(\theta_2) \;:=\;& -\hat{E}_T\left[v_{2it\alpha}(\theta_2)\right]^{-1}\Big\{\hat{\tilde{E}}_{T,m}[\hat{\tilde{v}}_{2it\alpha}(\theta_2)\hat{\psi}_{2is}(\theta_2)] + \hat{E}_T\left[\hat{v}_{2it\alpha\alpha}(\theta_2)\right]\hat{\tilde{\sigma}}_{2i}^2(\theta_2)/2\\
&+\hat{\tilde{E}}_{T,m}\left[\hat{\tilde{v}}_{2it\lambda}(\theta_2)\tilde{\lambda}_{it\alpha}\hat{\psi}_{1is}(\tilde{\theta}_1)\right] + \hat{E}_T\left[\hat{\tilde{v}}_{2it\lambda}(\theta_2)\left(\tilde{\lambda}_{it\alpha}\hat{\tilde{\beta}}_1(\tilde{\theta}_1) + \tilde{\lambda}_{it\alpha\alpha}\hat{\tilde{\sigma}}_{1i}^2(\tilde{\theta}_1)\right)\right]\\
&+\hat{E}_T\left[\hat{\tilde{v}}_{2it\lambda\alpha}(\theta_2)\tilde{\lambda}_{it\alpha}\right]\hat{\tilde{\sigma}}_{12i}(\theta_2) + \hat{E}_T\left[\hat{\tilde{v}}_{2it\lambda\lambda}(\theta_2)\tilde{\lambda}_{it\alpha}^2\right]\hat{\tilde{\sigma}}_{1i}^2(\tilde{\theta}_1)/2\Big\},\\
\hat{\mathcal{J}}_{2i}(\theta_2) \;:=\;& \hat{E}_T\left[\hat{\tilde{u}}_{2it\theta}(\theta_2)\right] - \hat{E}_T\left[\hat{\tilde{u}}_{2it\alpha}(\theta_2)\right]\hat{E}_T\left[\hat{\tilde{v}}_{2it\theta}(\theta_2)\right]/\hat{E}_T\left[\hat{\tilde{v}}_{2it\alpha}(\theta_2)\right],\\
\hat{\tilde{b}}_{2i}(\theta_2) \;:=\;& \hat{\tilde{E}}_{T,m}\left[\hat{\tilde{u}}_{2it\alpha}(\theta_2)\hat{\psi}_{2is}(\theta_2)\right] + \hat{E}_T\left[\hat{\tilde{u}}_{2it\alpha}(\theta_2)\right]\hat{\tilde{\beta}}_{2i}(\theta_2) + \hat{E}_T\left[\hat{\tilde{u}}_{2it\alpha\alpha}(\theta_2)\right]\hat{\tilde{\sigma}}_{2i}^2(\theta_2)/2\\
&+\hat{\tilde{E}}_{T,m}\left[\hat{\tilde{u}}_{2it\lambda}(\theta_2)\tilde{\lambda}_{it\alpha}\hat{\psi}_{1is}(\tilde{\theta}_1)\right] + \hat{E}_T\left[\hat{\tilde{u}}_{2it\lambda}(\theta_2)\left(\tilde{\lambda}_{it\alpha}\hat{\tilde{\beta}}_1(\tilde{\theta}_1) + \tilde{\lambda}_{it\alpha\alpha}\hat{\tilde{\sigma}}_{1i}^2(\tilde{\theta}_1)/2\right)\right]\\
&+\hat{E}_T\left[\hat{\tilde{u}}_{2it\lambda\alpha}(\theta_2)\tilde{\lambda}_{it\alpha}\right]\hat{\tilde{\sigma}}_{12i}^2(\theta_2) + \hat{E}_T\left[\hat{\tilde{u}}_{2it\lambda\lambda}(\theta_2)\tilde{\lambda}_{it\alpha}^2\right]\hat{\tilde{\sigma}}_{1i}^2(\tilde{\theta}_1)/2.
\end{aligned} \tag{30}$$

Here, $\hat{\bar{\beta}}_{2i}(\theta_1)$ is an estimator of the higher-order asymptotic bias of $\hat{\alpha}_{2i}(\theta_2)$ from a stochastic expansion as T grows, which accounts for the estimation error of the control function; whereas $\hat{\mathcal{J}}_{2i}(\theta_2)$ and $\hat{\bar{b}}_{2i}(\theta_2)$ are estimators of the Jacobian and asymptotic bias of the estimating equation of θ_2 for individual i.

A bias corrected estimator of the two-step FE estimator can be formed as

$$\tilde{\theta}_2 = \hat{\theta}_2 - \hat{\bar{\mathcal{B}}}_2(\hat{\theta}_2)/T, \tag{31}$$

where $\hat{\bar{\mathcal{B}}}_2(\theta_2) := -\hat{E}_n\left[\hat{\mathcal{J}}_{2i}(\theta_2)\right]^{-1}\hat{E}_n\left[\hat{\bar{b}}_2(\theta_2)\right]$ is an estimator of the bias of $\hat{\theta}_2$, and $\hat{\theta}_2$ is the two-step FE estimator of θ_{20} that uses $\tilde{\lambda}_{it}$ as the control function. As for one-step estimators, iterated bias corrections can also be formed by solving $\tilde{\theta}_2^{\infty} = \hat{\theta}_2 - \hat{\bar{\mathcal{B}}}_2(\tilde{\theta}_2^{\infty})/T$, and score-corrected estimators can be obtained by solving the modified first order condition:

$$\hat{E}_n\left[\hat{E}_T\left[\hat{\tilde{u}}_{2it}(\tilde{\theta}_2^{sc})\right] - \frac{1}{T}\hat{\bar{b}}_{2i}(\tilde{\theta}_2^{sc})\right] = 0. \tag{32}$$

Theorem 2 *Assume that Conditions 1, 2, 3, and 4 hold, and* $m \to \infty$ *such that* $m/T^{1/2} \to 0$. *Then, we have*

$$\sqrt{nT}\left(\tilde{\theta}_2 - \theta_{20}\right) \xrightarrow{d} \mathcal{N}(0, \mathcal{J}_2^{-1}\bar{\Omega}_2\mathcal{J}_2^{-1}). \tag{33}$$

Proof. See Appendix. ■

Remark 2 *The sandwich form of the asymptotic variance indicates that the two-step estimator is not efficient as in the cross sectional case, see Rivers and Vuong (1988). A consistent estimator for this variance can be obtained using cross sectional sample averages of* $\hat{\mathcal{J}}_{2i}(\tilde{\theta}_2)$ *and* $\hat{\bar{\Omega}}_{2i}(\tilde{\theta}_2)$, *where*

$$\hat{\bar{\Omega}}_{2i}(\theta_2) = \hat{\bar{E}}_{T,m}\left[\hat{U}_{2it}(\theta_2)\hat{U}_{2is}(\theta_2)'\right], \tag{34}$$

with $\hat{U}_{2it}(\theta_2) = \hat{\tilde{u}}_{2it}(\theta_2) + E_T\left[\hat{\tilde{u}}_{2it\alpha}(\theta_2)\right]\hat{\psi}_{2it}(\theta_2) + E_T\left[\hat{\tilde{u}}_{2it\lambda}(\theta_2)\tilde{\lambda}_{it\alpha}\right]\tilde{\psi}_{1it} + E_T\left[\hat{\tilde{u}}_{2it\lambda}(\theta_2)\tilde{\mathcal{J}}_{it}^{\lambda}\right]\tilde{\varphi}_{1it}$ *and* $\tilde{\psi}_{1it} = \hat{\psi}_{1it}(\tilde{\theta}_1)$.

Remark 3 *Note that, as in the first stage, if the model is fully static we can set* $m = 0$. *For conditional moments models with predetermined regressors we can replace* $\hat{\bar{E}}_{T,m}$ *for* $\hat{\tilde{E}}_{T,m}$, *where* $\hat{\tilde{E}}_{T,m}[f_{kit}\psi_{lis}] := \sum_{j=0}^{m}\sum_{t=j+1}^{T} f_{kit}\psi_{li,t-j}/(T-j)$ *and* $\hat{\tilde{E}}_{T,m}[\psi_{kit}\psi_{lis}] := \hat{E}_T\left[\psi_{kit}\psi_{lit}\right]$ *for* $k, l \in \{1, 2\}$.

5 Monte Carlo Experiments

This section reports evidence on the finite sample behavior of two-step FE estimators for static and dynamic models. We examine the finite sample properties of uncorrected and bias-corrected estimators in terms of bias and inference accuracy of their asymptotic distributions. The results are based on 1000 replications, and the designs correspond to a static panel sample selection model with probit selection rule and exogenous regressors, and a dynamic Tobit model with an endogenous explanatory variable.

5.1 Static Panel Sample Selection Model

The model design is

$$d_{it} = \mathbf{1}\left\{x_{1it}\beta_1 + x_{2it}\beta_2 + \alpha_{1i} - \epsilon_{1it} > 0\right\}, \tag{35}$$

$$y_{it} = d_{it} \times (x_{1it}\theta + \alpha_{2i} + \epsilon_{2it}), \qquad (i = 1, ..., n; t = 1, ..., T) \tag{36}$$

where $\theta = 1$; $\beta_1 = \beta_2 = 1$; x_{1it} and x_{2it} are independent $\mathcal{N}(-1, .5)$ variables; $\alpha_{1i} = \alpha_{2i} = 2 + \sum_{t=1}^{T}(x_{1it} + 1)/\sqrt{T} + \xi_i/\sqrt{2}$, with ξ_i an independent $\mathcal{N}(0, 1)$ variable; ϵ_{1it} and ϵ_{2it} are jointly distributed as a standard bivariate normal with correlation $\rho = .6$. All data are generated *i.i.d.* across individuals and over time. This design implies that $\Pr\{d_{it} = 1\} \approx .5$, so that approximately 50% of the sample is used to estimate θ in the second step. We generate panel data sets with $n = 100$ individuals and three different numbers of time periods T: 6, 8 and 12.

In this panel version of Heckman sample selection model the control function corresponds to the inverse mills ratio, that is $\lambda_{it} = \phi(x_{1it}\beta_1 + x_{2it}\beta_2 + \alpha_{1i})/\Phi(x_{1it}\beta_1 + x_{2it}\beta_2 + \alpha_{1i})$, where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and cdf of the standard normal distribution, respectively. The second stage is estimated by OLS including an estimate of the control function $\tilde{\lambda}_{it}$. Concentrating-out the second stage individual effects and the control function, the uncorrected estimator of θ takes the form

$$\hat{\theta} = \frac{\hat{\mu}_{x_1 y}\hat{\mu}_{\tilde{\lambda}\tilde{\lambda}} - \hat{\mu}_{x_1\tilde{\lambda}}\hat{\mu}_{\tilde{\lambda}y}}{\hat{\mu}_{x_1 x_1}\hat{\mu}_{\tilde{\lambda}\tilde{\lambda}} - \hat{\mu}^2_{x_1\tilde{\lambda}}}, \tag{37}$$

where $\hat{\mu}_{wv} = \sum_{i=1}^{n}\sum_{t=1}^{T} d_{it}\bar{w}_{it}\bar{v}_{it}$, and the bars denote that the variables are in deviations with respect to their individual mean in the observed sample; e.g., $\bar{\tilde{\lambda}}_{it} := \tilde{\lambda}_{it} - \sum_{t=1}^{T} d_{it}\tilde{\lambda}_{it}/\sum_{t=1}^{T} d_{it}$. From this expression we see that, after the fixed effects transformation, the only source of bias is the nonlinearity of the estimated control function in the

first stage individual effects. Thus, individual differences remove the individual effects of the primary equation α_{2i}; but they do not completely eliminate the individual effects of the selection equation that enter the primary equation through the control function since $\tilde{\lambda}_{it}$ is a nonlinear function of these individual effects, i.e., $\bar{\tilde{\lambda}}_{it}$ still depends on α_{1i}. Our general bias correction procedure in this case is equivalent to removing the bias from each of the FE averages involving the control function, that is from $\hat{\mu}_{\tilde{\lambda}\tilde{\lambda}}$, $\hat{\mu}_{x_1\tilde{\lambda}}$, and $\hat{\mu}_{\tilde{\lambda}y}$.

Throughout the tables, SD is the standard deviation of the estimator; $\hat{p}; \#$ denotes a rejection frequency with # specifying the nominal value; SE/SD is the ratio of the average standard error to standard deviation; and MAE denotes median absolute error.[8] $BC1$ and $BC2$ correspond to the two-step version of the analytical bias-corrected estimators of Hahn and Newey (2004) based on maximum likelihood setting and general estimating equations, respectively. $BC3$ is the two-step version of the bias-corrected estimator proposed in Fernandez-Val (2005), which replaces observed quantities for expected quantities in the expression of the bias. $JACK$ is the leave-one-period-out Jackknife-type estimator.

Note that due to the binary nature of the dependent variable in the selection equation the observations which have the same value for the dependent variable for each period are automatically removed when estimating the first step. In the second step we retain the observations for which the dependent variable in the selection equation is always one, and we assign a value of zero for their correction terms, i.e., their ML estimate.

Table 1 gives the results for the estimators of the probit parameters of the first stage, β_1 and β_2. These results are qualitatively similar to previous numerical studies; see, e.g., Hahn and Newey (2004) and Fernandez-Val (2005).[9] The uncorrected FE estimator MLE is severely biased, and the large-T bias corrections remove most of the incidental parameters bias for panels with even $T = 6$. This is especially true with the $BC3$ refinement.

Table 2 presents the finite sample properties for the FE estimators of θ in the second stage. OLS denotes a least squares estimator in the observed sample that ignores sample selection and is therefore inconsistent. $H - 1$ denotes the unfeasible OLS estimator that controls for selection by using the true (unobserved) inverse mills ratio, whereas $H - MLE$ is the feasible version of $H - 1$ that uses an estimate of the inverse mills ratio

[8] We use median absolute error instead of root mean squared error as an overall measure of goodness of fit because it is less sensitive to outliers.

[9] These studies, however, use a different design and include only one regressor.

(evaluated at uncorrected estimates of the probit parameters). $H - BC1$, $H - BC2$, and $H - BC3$ in addition to using estimates of the control function evaluated at bias-corrected estimates of the probit parameters, perform another round of bias-correction in the second stage. Uncorrected FE estimators have small biases, about 7%, 4% and 2% for 6, 8, and 12 time periods, respectively, which are reduced by the jackknife and analytical large-T bias corrections. Rejection frequencies are higher than their nominal levels due to underestimation of dispersion.[10] The corrected estimators have similar MAE to the uncorrected estimators because the corrections in this case increase dispersion.

Table 3 reports the ratio of estimators to the truth for the coefficient of the control function in the second stage. This is an important parameter as a significance test for this coefficient can be used to assess if there is endogenous sample selection. The results here show important biases towards zero in uncorrected FE estimators. Jackknife and analytical bias corrected estimators remove most of the bias and bring down the rejection frequencies closer to their nominal values, although the tests are still oversized due to the underestimation of the dispersion.

Some intuition for these numerical results can be obtained through a simple example. Specifically, suppose that $\alpha_i = \alpha \ \ \forall i$, that is the individual effects are the same for all the individuals. Fernandez-Val (2005) finds that in this case the biases for all the parameters in the first stage probit are scalar multiples of the true value of the parameters, and the limit probit index is also proportional to the true index. Since the inverse mills ratio is either close to zero or close to linear in the selected sample, the estimated control function is approximately proportional to the true inverse mills ratios. This is consistent with the small bias found for θ and the significant bias for ρ.

[10]The expressions used to compute the standard errors are robust to heteroskedasticity and account for estimated regressors using the method in Lee, Maddala, and Trost (1980). In results not reported, we find that the finite sample adjustments of MacKinnon and White (1985) to the heteroskedasticity corrections give rise to conservative standard errors.

5.2 Dynamic CML Tobit Model

The model design is

$$
\begin{aligned}
d_{it} &= x_{it}\beta + \alpha_{1i} + \epsilon_{1it}, && (38)\\
y^*_{i0} &= d_{i0}\theta_1 + \alpha_{2i} + \epsilon_{2i0}, && (39)\\
y^*_{it} &= d_{it}\theta_1 + y_{i,t-1}\theta_2 + \alpha_{2i} + \epsilon_{2it}, && (40)\\
y_{it} &= (y^*_{it} > 0)y^*_{it}, \qquad (i = 1, ..., n; t = 1, ..., T-1) && (41)
\end{aligned}
$$

where $\beta = 1$; $\theta_1 = \theta_2 = 1$; x_{it} is a $\mathcal{N}(0,1)$ variable; $\alpha_{1i} = \alpha_{2i} = \sum_{t=0}^{T-1} x_{it}/\sqrt{T}$; ϵ_{1it} and ϵ_{2it} are jointly distributed as a bivariate normal with correlation $\rho = .6$ and common variances $\sigma^2_\epsilon = 1 + (T-1)\rho^2/T$.[11] All data are generated *i.i.d.* across individuals and over time. We generate panel data sets with $n = 100$ individuals and three different total time periods T: 6, 8, and 12. For the trimming parameter that determines the number of lags used in the estimation of biases and variances, we choose a bandwidth parameter $m = 1$ following Hahn and Kuersteiner (2003).

A Tobit ML estimator for the primary equation is inconsistent here due to the endogeneity of d_{it}, even in the absence of dynamics and individual effects. Consistent estimates can be obtained by generalizing the two-step procedure of Smith and Blundell (1986) and Rivers and Vuong (1989) from cross sectional data to panel data with fixed effects and lagged dependent variables. The control function corresponds to the reduced form equation error, $\lambda_{it} = \epsilon_{1it}$, and can be estimated as the residuals of a FE regression. Including this estimated control function and fixed effects in the primary equation allows us to account for both time varying and time invariant endogeneity. A bias correction is required, however, to reduce the incidental parameters problem due to the nonlinearity of the Tobit estimator.

Table 4 presents the finite sample results for the coefficient of the endogenous continuous explanatory variable d_{it}. $TOBIT$ denotes the FE Tobit estimator that does not account for the endogeneity of d_{it} and is inconsistent. $CMLE-1$ denotes the unfeasible estimator that uses the true (unobserved) control function in the second stage; whereas $CMLE$ is the feasible version of $CMLE-1$ that replaces the true ϵ_{1it}'s with the OLS FE residuals of the reduced form for d_{it}. Overall, all the FE estimators that control for

[11] We choose the value of σ^2_ϵ such that the variance of the error term in the estimating equation that includes the estimated control function is approximately equal to 1.

endogeneity have small finite sample biases, although the inference procedures are oversized due to underestimation of the dispersion.[12] The results here agree with the Honoré (1993) and Greene (2004) numerical findings of small biases in Tobit FE estimators of the slope parameters.

Table 5 reports the results for the coefficient of the lagged dependent variable. Here, as in the uncensored linear case, uncorrected FE estimators are biased downward even when we use the true control function. The bias corrections remove an important part of this bias, and have rejection probabilities closer to their nominal value, although the test is still oversized. Table 6 shows the results for the coefficient of the control function ρ. This coefficient captures the correlation between the error terms of the control equation and primary equation. Here we find small biases for both uncorrected and bias-corrected estimators, with the endogeneity tests having bigger size than their nominal level due to the underestimation of the dispersion of the parameters.

6 Empirical Illustration: Estimating the Impact of Union Status on Wages

To illustrate our approach we estimate a two equation model which describes the manner in which union status affects wages where the union status decision, which is endogenous to wages, is treated as a dynamic binary choice outcome. The model is similar to that considered in Vella and Verbeek (1998), hereafter VV, noting that there the individual components are treated as random effects. In particular, we estimate the following equations

$$Union_{it} = \mathbf{1}\left\{\beta_1 \times Union_{i,t-1} + x_{it}'\beta_2 + \alpha_{1i} + \epsilon_{1it} > 0\right\}, \tag{42}$$

$$wage_{it} = \theta_1 \times Union_{it} + x_{it}'\theta_2 + \alpha_{2i} + \epsilon_{2it}, \tag{43}$$

where $Union$ is a binary variable denoting that the individual is a member of a union and $wage$ is the log of the individual's hourly wage rate. The vector x_{it} includes completed years of schooling, log of potential experience (age - schooling - 6), and married, rural area, health disability, region, industry, time, and occupation dummies. The model is interesting in the context of the methods presented here as the binary union decision

[12] The expression for the standard errors accounts for the estimation of the regressors.

equation has a FE and a lagged dependent variable. Also, the wage equation has a binary endogenous regressor where the endogeneity is the result of potentially time varying heterogeneity. This specification is similar to VV.

The sample, selected from the National Longitudinal Survey (Youth Sample), consists of full-time young working males followed over the period 1980 to 1988. We exclude individuals who fail to provide sufficient information for each year, are in the active forces in any year, have negative potential experience in at least one year, their schooling decreases in any year or increases by more than two years between two interviews, or report too high (more than $500 per hour) or too low (less than $1 per hour) wages. The final sample includes 545 men. The first period is used as the initial condition for the lagged union variable.[13]

Table 7 reports descriptive statistics for the sample used. Union membership is based on a question reflecting whether or not the individual had his wage set in a collective bargaining agreement. Roughly 26 % of the sample are union members. Union and nonunion workers have similar observed characteristics, though union workers are slightly less educated, more likely to be married, more likely to live in the northern central region, and less likely to live in the South. Across industries, there are relatively more union workers in transportation, manufacturing and public administration, and fewer in trade and business. Union membership reduces wage dispersion and has high persistence. Note that all variables, except for the Black and Hispanic dummies, display time variation over the period considered. The unconditional union premium is around 23 %.

Table 8 presents the estimates for the dynamic probit model of union membership. The left panel of this Table excludes the occupational dummies while the right panel includes them. We make this distinction to remain comparable to VV. In each panel, the first column reports pooled probit estimates that do not account for individual time invariant heterogeneity, the second column shows the unadjusted FE probit estimates, while the third column presents the corresponding bias corrected estimates. The fourth and fifth columns give the average marginal effects for each of the FE models.[14] We include time dummies in the specification to remain comparable to VV, even though they

[13] Although we do not use the identical data to VV the time period and the summary statistics of the data sets are very similar.

[14] The bias corrected estimates reported correspond to the $BC3$ method. The other methods give similar results.

are not covered by the regularity conditions. In results not reported, however, we find that excluding the time dummies does not have any significant effect on the estimates.[15]

First, note that for the estimate of primary interest in this table, the coefficient of the lag dependent variable, the pooled probit estimator that does not account for heterogeneity leads to an important overstatement of the importance of the state dependence. This result can be seen by comparing ratios of coefficients, for example with respect to the coefficient of log experience, since the pooled and FE estimators use different normalizations. Comparing with the VV estimates of .611 and .632 for the left and right panels, respectively, our unadjusted FE probit estimates are smaller with values of .354 and .318. More interestingly, however, are the adjusted results. The bias adjusted estimates of the lagged union variable coefficients are approximately .73 for the specification excluding the occupational dummies, and .70 for that with the occupation dummies included. These estimates are more similar to those reported by VV. The effects of the bias corrections in the FE estimators are easier to interpret by looking at the estimates of the average marginal effects. While the unadjusted estimates already reveal a substantial degree of state dependence with average marginal effects of 4 to 5 percentage points, recalling that the mean of the union membership variable is only 26 percent, the adjusted estimates of state dependence are approximately 100 percent higher with estimates of 9 to 10 percentage points. An inspection of the other marginal effects indicates there is little difference between the adjusted and unadjusted estimates.

Table 9 presents the estimates of the wage equation. We consider a range of estimators depending on whether or not they account for possible endogenous time varying sample selection and/or individual time invariant heterogeneity. We again provide estimates that include and exclude occupational dummies in the left and right panel, respectively. We start from a pooled OLS estimator that does not account for any source of heterogeneity ($P - OLS$). Then, we control for possible sample selection using a pooled Heckman estimator ($P - Heckit$). Next, we introduce a FE estimator that controls for individual heterogeneity but not for time varying sample selection (FE). Finally, we consider a FE Heckman estimator that controls for both time varying and time invariant heterogeneity ($FE - Heck$), together with a bias corrected version of this estimator that reduces the incidental parameters problem of $FE - Heck$ ($BC - Heck$).[16] The corrected estimator

[15]These results are available from the authors upon request.

[16]Note that the control function in this case is the generalized residual for the probit model.

employs the bias corrected estimates of Table 8 to construct the control function, and performs an additional bias correction of the estimates of the wage equation to fix the bias problem due to the non linearity of the control function in the estimates of the first stage individual effects.

Pooled OLS produces estimates of the union effect of 16 and 18 percent. As in previous studies, these estimates increase when possible non random selection into unions is taken into account, and decrease when individual heterogeneity is controlled for using longitudinal estimators. More interestingly, the effect of endogenous selection is more important for estimators that account also for individual heterogeneity. Thus, the union effect raises from 15-18 percent to 24-28 percent for pooled estimators, whereas it jumps from 10-11 percent to 30-32 percent for FE estimators. The difference is even more acute when we correct the bias problem of the Heckman FE estimator. Thus, the corrected estimates give a union effect of about 40 to 42 percent. These results are also in line with VV estimates, which find a union effect of about 39 percent. For the other coefficients we only observe significant differences between corrected and uncorrected estimates for the coefficient of the control function.

Overall the evidence leads to a number of conclusions. First, of the parameters of interest in this empirical investigation it appears that the ones most subject to bias are that for the lagged dependent variable in the union membership equation, and those for the union variable and selection correction in the wage equation. Second, the results here confirm the finding in VV that the increase in the union effect which results from OLS estimation is due to time varying heterogeneity rather than time invariant heterogeneity. Finally, the empirical evidence indicates that there is significant interaction between the individual heterogeneity in the wage equation and the selection mechanism of workers into unions.

7 Summary and conclusion

This paper introduces bias-corrected estimators for nonlinear and dynamic panel models with both time invariant and time variant heterogeneity. These estimators have closed analytical form and are easy to implement. A major attraction of our approach is that it does not require any assumption on the parametric form of the distribution of the unobserved individual heterogeneity.

Our estimation strategy is very flexible and can accommodate other models of interest with minor adjustments. For example, the estimation method for the dynamic tobit model with endogenous regressors can be extended to the case where the lag of the latent dependent variable, instead of the lag of the censored dependent variable, is included as explanatory variable. This model is arguably more plausible for economic applications where the censoring comes from data limitations due, for example, to top-coding (see, e.g., Hu, 2002).

Our simulation evidence strongly suggests that our approach is a very effective bias reduction method for two important models. Moreover, it is very likely that this performance would also extend to other models of interest to economists. Finally, an empirical example which investigates the effect of endogenous union membership on wages illustrates the importance of accounting for both unobserved time invariant and time varying heterogeneity and highlights the need to bias adjust the fixed effects estimates.

Appendix

Throughout the appendices O_{up} and o_{up} denote uniform orders in probability. For example, for a sequence of random variables $\{\xi_i, i = 1, ..., n\}$, $\xi_i = O_{up}(1)$ means $\max_{1\leq i\leq n} \xi_i = O_p(1)$, and $\xi_i = o_{up}(1)$ means $\max_{1\leq i\leq n} \xi_i = o_p(1)$. For a matrix $A = (a_{ij}), i = 1, ..., m, j = 1, ..., n$, $|A|$ denotes Euclidean norm, that is $|A|^2 = trace[AA']$. HK refers to Hahn and Kuersteiner (2003).

A Notation

A.1 First Stage

Let $\hat{\alpha}_{1i0} = \hat{\alpha}_{1i}(\theta_{10})$. Then, a standard higher-order asymptotic expansion gives (see, e.g., Hahn and Kuersteiner, 2003), as $T \to \infty$,

$$\begin{aligned}
\hat{\alpha}_{1i0} &= \alpha_{1i0} + \frac{1}{T}\sum_{t=1}^{T}\psi_{1it} + \frac{1}{T}\bar{\beta}_{1i} + \frac{1}{T^{3/2}}R_{1i0}, \quad \psi_{1it} = -E_T\left[v_{1it\alpha}\right]^{-1} v_{1it}, \; R_{1i0} = o_{up}(\sqrt{T}) \\
\bar{\beta}_{1i} &= -E_T\left[v_{1it\alpha}\right]^{-1}\left\{\bar{E}_T\left[v_{1it\alpha}\psi_{1is}\right] + \frac{1}{2}E_T\left[v_{1it\alpha\alpha}\right]\bar{\sigma}_{1i}^2\right\}, \quad \bar{\sigma}_{1i}^2 = \bar{E}_T\left[\psi_{1it}\psi_{1is}\right]. \qquad (44)
\end{aligned}$$

The asymptotic bias of the first stage FE estimator has the form $\bar{\mathcal{B}}_1 := -\mathcal{J}_1^{-1}\bar{b}_1$, where

$$\begin{aligned}
\mathcal{J}_1 &= E_n\left[E_T\left[u_{1it\theta}\right] - E_T\left[u_{1it\alpha}\right]E_T\left[v_{1it\theta}\right]/E_T\left[v_{1it\alpha}\right]\right], \qquad (45) \\
\bar{b}_1 &= E_n\left[\bar{E}_T\left[u_{1it\alpha}\psi_{1is}\right] + E_T\left[u_{1it\alpha}\right]\bar{\beta}_{1i} + \frac{1}{2}\bar{\sigma}_{1i}^2 E_T\left[u_{1it\alpha\alpha}\right]\right]. \qquad (46)
\end{aligned}$$

A.2 Control Function

Let

$$\lambda_{it}(\theta, \alpha) := \lambda(w_{it}; \theta, \alpha), \qquad (47)$$

denote the control function. Additional subscripts refer to partial derivatives, e.g., $\lambda_{it\theta}(\theta, \alpha) := \partial\lambda_{it}(\theta, \alpha)/\partial\theta'$. For notational convenience the arguments are omitted when the expressions are evaluated at the true parameter value, i.e., $\lambda_{it} := \lambda_{it}(\theta_{10}, \alpha_{1i0})$; and arguments are omitted and a tilde is added when the expressions are evaluated at bias corrected estimates of the parameters, i.e., $\tilde{\lambda}_{it} := \lambda_{it}(\tilde{\theta}_1, \tilde{\alpha}_{1i})$. Then, by Lemma 2, we have

$$\tilde{\lambda}_{it} = \lambda_{it} + \tilde{\psi}_{it}^{\lambda}/\sqrt{T} + \bar{\beta}_{it}^{\lambda}/T + R_{it}^{\lambda}/T^{3/2}, \qquad (48)$$

where $\tilde{\psi}_{it}^{\lambda} = \lambda_{it\alpha}\psi_{1i} + \mathcal{J}_{it}^{\lambda}\varphi_1/\sqrt{n}$, $\mathcal{J}_{it}^{\lambda} = \lambda_{it\theta} - E_T\left[v_{1it\alpha}\right]^{-1} E_T\left[v_{1it\theta}\right]\lambda_{it\alpha}$, and $\bar{\beta}_{it}^{\lambda} = \lambda_{it\alpha}\bar{\beta}_{1i} + \lambda_{it\alpha\alpha}\psi_{1i}^2/2$.

A.3 Second Stage

Let $\hat{\alpha}_{2i0} = \hat{\alpha}_{2i}(\theta_{20})$. Then, a higher-order asymptotic expansion gives (see Lemmas 8 and 9 in Appendix B), for $T \to \infty$

$$\begin{aligned}
\hat{\alpha}_{2i0} &= \alpha_{2i0} + \frac{1}{T}\sum_{t=1}^{T}\psi_{2it} + \frac{1}{T}\bar{\beta}_{2i} + o_p(T^{-1}), && (49)\\
\psi_{2it} &= -E_T\left[v_{2it\alpha}\right]^{-1}\left\{v_{2it} + E_T\left[v_{2it\lambda}\lambda_{it\alpha}\right]\psi_{1it} + E_T\left[v_{2it\lambda}\mathcal{J}_{it}^{\lambda}\right]\varphi_{1it}/\sqrt{n}\right\}, && (50)\\
\bar{\beta}_{2i} &= -E_T\left[v_{2it\alpha}\right]^{-1}\left\{\bar{E}_T[v_{2it\alpha}\psi_{2is}] + E_T\left[v_{2it\alpha\alpha}\right]\bar{\sigma}_{2i}^2/2 + E_T\left[v_{2it\lambda\alpha}\lambda_{it\alpha}\right]\bar{\sigma}_{12i}\right.\\
&\quad + \bar{E}_T\left[v_{2it\lambda}\lambda_{it\alpha}\psi_{1is}\right] + E_T\left[v_{2it\lambda}\left(\lambda_{it\alpha}\bar{\beta}_{1i} + \frac{1}{2}\lambda_{it\alpha\alpha}\bar{\sigma}_{1i}^2\right)\right]\\
&\quad \left. + E_T\left[v_{2it\lambda\lambda}\lambda_{it\alpha}^2\right]\bar{\sigma}_{1i}^2/2\right\}, && (51)\\
\bar{\sigma}_{2i}^2 &= \bar{E}_T\left[\psi_{2it}\psi_{2is}\right], \quad \bar{\sigma}_{12i} = \bar{E}_T\left[\psi_{1it}\psi_{2is}\right]. && (52)
\end{aligned}$$

B Proofs of Main Results

B.1 Proof of Lemma 1

The result for $\tilde{\theta}_1$ follows from Theorem 2 in HK. For $\tilde{\alpha}_{1i}$, note that

$$\tilde{\alpha}_{1i} = \hat{\alpha}_{1i}\left(\theta_{10}\right) + \frac{\partial\hat{\alpha}_{1i}(\bar{\theta}_1)}{\partial\theta_1'}(\tilde{\theta}_1 - \theta_{10}), \tag{53}$$

where $\bar{\theta}_1$ lies between $\tilde{\theta}_1$ and θ_{10}. Then, the asymptotic expansion for $\tilde{\alpha}_{1i}$ can be derived using the asymptotic expansion for $\hat{\alpha}_{1i0}$ in (44), the first result of the Lemma, and

$$\frac{\partial\hat{\alpha}_{1i}(\bar{\theta}_1)}{\partial\theta_1'} = -E_T\left[v_{1it\alpha}\right]^{-1} E_T\left[v_{1it\theta}\right] + o_{up}(1). \tag{54}$$

The last result follows by differentiating the first order condition of $\hat{\alpha}_{1i}(\theta_1)$, $0 = \hat{E}_T\left[v_{1it}\left(\theta_1, \hat{\alpha}_{1i}(\theta_1)\right)\right]$, with respect to $\hat{\alpha}_{1i}$ and θ_1. The remainder terms are uniformly bounded in probability by Lemmas 10, 12, 14, and 16 in HK.

B.2 Consistency of $\hat{\theta}_2$ and $\hat{\alpha}_{2i}(\hat{\theta}_2)$

Lemma 3 *Assume that Conditions 1, 2, and 3 hold. Let $h(w_{it}, \lambda_{it}; \theta, \alpha)$ be a function such that (i) $h(w_{it}, \lambda_{it}; \theta, \alpha)$ is continuous in λ_{it}; (ii) there exists a function $M(w_{it}, \lambda_{it})$ such that $|h(w_{it}, \lambda_{it}; \theta, \alpha)| \leq M(w_{it}, \lambda_{it})$ and $|\partial h(w_{it}, \lambda_{it}; \theta, \alpha)/\partial \lambda_{it}| \leq M(w_{it}, \lambda_{it})$, with $\sup_i E[M(w_{it}, \lambda_{it})] < \infty$. Then, for any $\eta > 0$, we have*

$$\Pr\left\{\max_{1\leq i\leq n} \sup_{(\theta,\alpha)\in\Gamma_2} \left|\hat{\bar{h}}_i(\theta,\alpha) - \hat{h}_i(\theta,\alpha)\right| \geq \eta\right\} = o(T^{-1}). \tag{55}$$

where

$$\hat{\bar{h}}_i(\theta,\alpha) = \hat{E}_T\left[h_i(w_{it}, \bar{\lambda}_{it}; \theta, \alpha)\right], \tag{56}$$

$$\hat{h}_i(\theta,\alpha) = \hat{E}_T\left[h_i(w_{it}, \lambda_{it}; \theta, \alpha)\right], \tag{57}$$

with $\bar{\lambda}_{it} := \lambda(w_{it}; \bar{\theta}_1, \bar{\alpha}_{1i})$, and $(\bar{\theta}_1', \bar{\alpha}_{1i})$ lies between $(\theta_{10}', \alpha_{1i0})$ and $(\tilde{\theta}_1', \tilde{\alpha}_{1i})$.

Proof. By Condition 3 and assumption (ii), we can write

$$\max_{1\leq i\leq n} \sup_{(\theta,\alpha)\in\Gamma_2} \left|\hat{\bar{h}}_i(\theta,\alpha) - \hat{h}_i(\theta,\alpha)\right| \leq \sup_i E\left[M(w_{it}, \lambda_{it})\right] \sup_i E\left[M_\lambda(w_{it})\right] \max_{1\leq i\leq n} \left|(\bar{\theta}_1, \bar{\alpha}_{1i}) - (\theta_{10}, \alpha_{1i0})\right|. \tag{58}$$

Then, the result follows by Lemma 1, Condition 3, and assumption (ii) of the Lemma. ■

Lemma 4 *Assume that Conditions 1, 2, 3, and 4 hold. Then, for any $\eta > 0$, we have*

$$\Pr\left\{\max_{1\leq i\leq n} \sup_{(\theta,\alpha)\in\Gamma_2} \left|\hat{\bar{g}}_{2i}(\theta,\alpha) - g_{2i}(\theta,\alpha)\right| \geq \eta\right\} = o(T^{-1}), \tag{59}$$

where

$$\hat{\bar{g}}_{2i}(\theta,\alpha) = \hat{E}_T\left[g_{2i}(w_{it}, \bar{\lambda}_{it}; \theta, \alpha)\right], \tag{60}$$

$$g_{2i}(\theta,\alpha) = E_T\left[g_{2i}(w_{it}, \lambda_{it}; \theta, \alpha)\right], \tag{61}$$

where $\bar{\lambda}_{it} := \lambda(w_{it}; \bar{\theta}_1, \bar{\alpha}_{1i})$, and $(\bar{\theta}_1', \bar{\alpha}_{1i})$ lies between $(\theta_{10}', \alpha_{1i0})$ and $(\tilde{\theta}_1', \tilde{\alpha}_{1i})$.

Proof. By triangle inequality, note that

$$\left|\hat{\bar{g}}_{2i}(\theta,\alpha) - g_{2i}(\theta,\alpha)\right| \leq \left|\hat{\bar{g}}_{2i}(\theta,\alpha) - \hat{g}_{2i}(\theta,\alpha)\right| + \left|\hat{g}_{2i}(\theta,\alpha) - g_{2i}(\theta,\alpha))\right|, \tag{62}$$

where $\hat{g}_{2i}(\theta,\alpha) = \hat{E}_T\left[g_{2i}(w_{it}, \lambda_{it}; \theta, \alpha)\right]$. Then, the conclusion follows by Lemma 3 applied to $h(w_{it}, \lambda_{it}; \theta, \alpha) = g_2(w_{it}, \lambda_{it}; \theta, \alpha)$, and Lemma 1 in HK. ■

Proposition 1 *Assume that Conditions 1, 2, 3, and 4 hold, and $m \to \infty$ such that $m/T^{1/2} \to 0$. Then, for any $\eta > 0$, we have*

$$\Pr\left\{\left|\hat{\theta}_2 - \theta_{20}\right| \geq \eta\right\} = o(T^{-1}). \tag{63}$$

Proof. This Lemma can be proven using the same argument as in the proof of Theorem 3 in HK, replacing Lemma 4 for of Lemma 1 in HK. ■

Proposition 2 *Assume that Conditions 1, 2, 3, and 4 hold, and $m \to \infty$ such that $m/T^{1/2} \to 0$. Then, for any $\eta > 0$, we have*

$$\Pr\left\{\max_{1\leq i\leq n}\left|\hat{\alpha}_{2i}(\hat{\theta}_2) - \alpha_{2i0}\right| \geq \eta\right\} = o(T^{-1}). \tag{64}$$

Proof. The result follows using the same argument as in the proof of Theorem 4 in HK, replacing Lemma 4 and Proposition 1 for Lemma 1 and Theorem 3 in HK. ■

Corollary 1 *Assume that Conditions 1, 2, 3, and 4 hold, and $m \to \infty$ such that $m/T^{1/2} \to 0$. Then,*

$$\Pr\left\{\max_{1\leq i\leq n}\left|\hat{\alpha}_{2i0} - \alpha_{2i0}\right| \geq \eta\right\} = o(T^{-1}), \tag{65}$$

for any $\eta > 0$, where $\hat{\alpha}_{2i0} := \hat{\alpha}_{2i}(\theta_{20})$.

Proof. Same arguments as in the proof of Proposition 2 replacing $\hat{\theta}_2$ for θ_{20}. ■

B.3 Asymptotic Expansion for $\tilde{\lambda}_{it}$

Lemma 5 *Assume that Conditions 1, 2, 3, and 4 hold, and $m \to \infty$ such that $m/T^{1/2} \to 0$. Let $\bar{\lambda}_{it}$ denote the estimator of the control functions, $\lambda(w_{it}; \bar{\theta}_1, \bar{\alpha}_{1i})$, where $\bar{\theta}_1$ and $\bar{\alpha}_{1i}$ lie between the bias corrected estimators of the control equation parameters, $\tilde{\theta}_1$ and $\tilde{\alpha}_{1i} := \hat{\alpha}_{1i}(\tilde{\theta}_1)$; and the true parameter values θ_{10} and α_{1i0}, respectively. Let $\bar{\lambda}_{it\alpha^{d_1}\theta^{d_2}}$ denote its derivatives*

$$\frac{\partial^{d_1+d_2}\lambda(w_{it};\theta,\alpha)}{\partial^{d_1}\alpha\partial^{d_2}\theta}, \tag{66}$$

evaluated at $(\bar{\theta}_1', \bar{\alpha}_{1i})$, and $\lambda_{it\alpha^{d_1}\theta^{d_2}}$ denote the derivatives evaluated at $(\theta_{10}', \alpha_{1i0})$, for $0 \leq d_1 + d_2 \leq 3$. Then, for almost every w_{it}, we have

$$\sqrt{T}\left(\bar{\lambda}_{it\alpha^{d_1}\theta^{d_2}} - \lambda_{it\alpha^{d_1}\theta^{d_2}}\right) \;=\; o_{up}\left(T^{1/10+\upsilon}\right), \tag{67}$$

for some $0 < \upsilon < (100q_1 + 120)^{-1}$.

Proof. The result follows by Condition 3, Lemma 1, and Lemma 11 in HK. ■

Proof of Lemma 2

Proof. By a standard Taylor expansion around $(\theta_{10}, \alpha_{1i0})$, we have

$$\tilde{\lambda}_{it} = \lambda_{it} + \lambda_{it\alpha}(\tilde{\alpha}_{1i} - \alpha_{1i0}) + \lambda_{\theta}(w_{it}; \overline{\theta}_1, \overline{\alpha}_{1i})(\tilde{\theta}_1 - \theta_{10}) + \frac{1}{2}\lambda_{\alpha\alpha}(w_{it}; \breve{\theta}_1, \breve{\alpha}_{1i})(\tilde{\alpha}_{1i} - \alpha_{1i0})^2, \quad (68)$$

where $(\overline{\theta}_1, \overline{\alpha}_{1i})$ and $(\breve{\theta}_1, \breve{\alpha}_{1i})$ lie between $(\tilde{\theta}_1, \tilde{\alpha}_{1i})$ and $(\theta_{10}, \alpha_{10i})$. The expressions for $\tilde{\psi}^{\lambda}_{it}$, and $\bar{\beta}^{\lambda}_{1it}$ can be obtained using the expansions for $\tilde{\theta}_1$ and $\tilde{\alpha}_{1i}$ in Lemma 1, after some algebra. For the remainder term, we have

$$\begin{aligned} R^{\lambda}_{it} &= \lambda_{it\alpha} R_{1i} + \sqrt{T}\left(\lambda_{\theta}(w_{it}; \overline{\theta}_1, \overline{\alpha}_{1i}) - \lambda_{it\theta}\right) T(\tilde{\theta}_1 - \theta_{10}) + \lambda_{it\theta}\sqrt{T} R_1/n \\ &+ \frac{1}{2}\lambda_{it\alpha\alpha}\left(-\sqrt{T} E_T\left[v_{it\alpha}\right]^{-1} E_T\left[v_{it\theta}\right]\varphi_1/\sqrt{n} + \bar{\beta}_{1i} + R_{1i}/\sqrt{T}\right)\left[\sqrt{T}\left(\tilde{\alpha}_{1i} - \alpha_{1i0}\right) + \psi_{1i}\right] \\ &+ \sqrt{T}\left(\lambda_{\alpha\alpha}(w_{it}; \breve{\theta}_1, \breve{\alpha}_{1i}) - \lambda_{it\alpha\alpha}\right) T(\tilde{\alpha}_{1i} - \alpha_{1i})^2/2 = o_{up}\left(\sqrt{T}\right). \end{aligned} \quad (69)$$

The uniform rate of convergence then follows by Lemmas 1 and 5. ■

B.4 Proof of Theorem 1

Proof. From a Taylor Expansion of the FOC for $\hat{\theta}_2$ around θ_{20}, we have

$$0 = \hat{E}_n\left[\hat{E}_T\left[\hat{\tilde{u}}_{2it}(\hat{\theta}_2)\right]\right] = \hat{E}_n\left[\hat{E}_T\left[\hat{\tilde{u}}_{2it}(\theta_{20})\right]\right] + \hat{E}_n\left[\hat{E}_T\left[\frac{d\hat{\tilde{u}}_{2it}(\overline{\theta}_2)}{d\theta'}\right]\right](\hat{\theta}_2 - \theta_{20}), \quad (70)$$

where $\overline{\theta}_2$ lies between $\hat{\theta}_2$ and θ_0.

Part I: Asymptotic limit for $\hat{\mathcal{J}}_{2i}(\bar{\theta}_2) := \hat{E}_T\left[\frac{d\hat{\tilde{u}}_{2it}(\overline{\theta}_2)}{d\theta'}\right]$. Note that

$$\hat{\mathcal{J}}_{2i}(\bar{\theta}_2) = \hat{E}_T\left[\tilde{u}_{2it\theta}(\overline{\theta}_2, \hat{\alpha}_{2i}(\overline{\theta}_2))\right] + \hat{E}_T\left[\tilde{u}_{2it\alpha}(\overline{\theta}_2, \hat{\alpha}_{2i}(\overline{\theta}_2))\right]\frac{\partial\hat{\alpha}_{2i}(\overline{\theta}_2)}{\partial\theta} \quad (71)$$

Then, differentiation of the FOC for $\hat{\alpha}_{2i}$, $\hat{E}_T\left[\tilde{v}_{2it}(\hat{\alpha}_{2i}(\overline{\theta}_2), \overline{\theta}_2)\right] = 0$, with respect to θ_2 and $\hat{\alpha}_{2i}$ gives

$$\hat{E}_T\left[\tilde{v}_{2it\theta}(\hat{\alpha}_{2i}(\overline{\theta}_2), \overline{\theta}_2)\right] + \hat{E}_T\left[\tilde{v}_{2it\alpha}(\hat{\alpha}_{2i}(\overline{\theta}_2), \overline{\theta}_2)\right]\frac{\partial\hat{\alpha}_{2i}(\overline{\theta}_2)}{\partial\theta} = 0. \quad (72)$$

By repeated application of Lemma 11 in HK and Lemma 3, we can write

$$\frac{\partial\hat{\alpha}_{2i}(\overline{\theta}_2)}{\partial\theta} = -E_T\left[v_{2it\alpha}\right]^{-1} E_T\left[v_{2it\theta}\right] + o_{up}(1). \quad (73)$$

From Lemma 3 and Lemma 11 in HK, we have

$$\hat{E}_T\left[\tilde{u}_{2it\theta}(\bar{\theta}_2,\hat{\alpha}_{2i}(\bar{\theta}))\right] = \hat{E}_T\left[u_{2it\theta}(\bar{\theta}_2,\hat{\alpha}_{2i}(\bar{\theta}))\right]+o_{up}(1)=E_T\left[u_{2it\theta}\right]+o_{up}(1), \quad (74)$$

$$\hat{E}_T\left[\tilde{u}_{2it\alpha}(\bar{\theta}_2,\hat{\alpha}_{2i}(\bar{\theta}_2))\right] = \hat{E}_T\left[u_{2it\alpha}(\bar{\theta}_2,\hat{\alpha}_{2i}(\bar{\theta}_2))\right]+o_{up}(1)=E_T\left[u_{2it\alpha}\right]+o_{up}(1). \quad (75)$$

Finally, replacing the expressions for the components in (71) we have

$$\hat{\mathcal{J}}_{2i}(\bar{\theta}_2) = E_T\left[u_{2it\theta}\right]-E_T\left[v_{2it\alpha}\right]^{-1}E_T\left[u_{2it\alpha}\right]E_T\left[v_{2it\theta}\right]+o_{up}(1):=\mathcal{J}_{2i}+o_{up}(1), \quad (76)$$

where $E_n\left[\mathcal{J}_{2i}\right]\xrightarrow{d}\mathcal{J}_2$ by LLN.

Part II: Asymptotic Expansion for $\hat{\theta}_2-\theta_{20}$. From part I and Lemma 12, we have

$$0 = \underbrace{\sqrt{nT}\hat{E}_n\left[\hat{E}_T\left[\hat{\tilde{u}}_{2it}(\theta_{20})\right]\right]}_{O_p(1)}+\underbrace{\hat{E}_n\left[\hat{\mathcal{J}}_{2i}(\bar{\theta}_2)\right]}_{O_p(1)}\sqrt{nT}(\hat{\theta}_2-\theta_{20}). \quad (77)$$

Therefore, $\sqrt{nT}(\hat{\theta}-\theta_0)=O_p(1)$. Then, by (76), Condition 4, and Lemma 12, we have

$$\sqrt{nT}(\hat{\theta}_2-\theta_{20})=-\mathcal{J}_2^{-1}\sqrt{nT}\hat{E}_n\left[\hat{E}_T\left[\hat{\tilde{u}}_{2it}(\theta_{20})\right]\right]+o_p(1)\xrightarrow{d}\mathcal{N}\left(\sqrt{\rho}\bar{\mathcal{B}}_2,\mathcal{J}_2^{-1}\bar{\Omega}_2\mathcal{J}_2^{-1}\right). \quad (78)$$

The expression for $\bar{\mathcal{B}}_2$ follows from part I and Lemma 11. ■

Corollary 2 *Assume that Conditions 1, 2, 3, and 4 hold, and $m\to\infty$ such that $m/T^{1/2}\to 0$. Then, we have*

$$\sqrt{nT}\left(\hat{\theta}_2-\theta_{20}-\frac{1}{T}\bar{\mathcal{B}}_2\right)=\varphi_2+o_p(1)\xrightarrow{d}\mathcal{N}\left(0,\mathcal{J}_2^{-1}\bar{\Omega}_2\mathcal{J}_2^{-1}\right). \quad (79)$$

Proof. The result follows from Theorem 1 and Slutsky Theorem. ■

B.5 Proof of Theorem 2

Proof. First, note that

$$\sqrt{nT}\left(\tilde{\theta}_2-\theta_{20}\right)=\sqrt{nT}\left(\hat{\theta}_2-\theta_{20}-\frac{1}{T}\bar{\mathcal{B}}_2\right)-\sqrt{\frac{n}{T}}\left(\hat{\bar{\mathcal{B}}}_2(\hat{\theta}_2)-\bar{\mathcal{B}}_2\right). \quad (80)$$

By Corollary 2 and Condition 1 we only need to show that $\hat{\bar{\mathcal{B}}}_2(\hat{\theta}_2)-\bar{\mathcal{B}}_2=o_p(1)$.

Recall that $\bar{\mathcal{B}}_2$ is a (continuous) function of expectations of derivatives of the objective function evaluated at the true parameter values, i.e., expressions of the form $h_i(\theta_{20},\alpha_{2i0}) = E_T\left[h(w_{it},\lambda_{it};\theta_{20},\alpha_{2i0})\right]$. $\hat{\bar{\mathcal{B}}}_2(\hat{\theta}_2)$, the fixed effects estimator of $\bar{\mathcal{B}}_2$, replaces expected values by sample analogs, and the true values of the parameters and

control functions by fixed effects estimates, i.e., $\hat{\bar{\mathcal{B}}}_2(\hat{\theta}_2)$ has components of the form $\hat{\bar{h}}_i(\hat{\theta}_2, \hat{\alpha}_{2i}) = \hat{E}_T\left[h(w_{it}, \tilde{\lambda}_{it}; \hat{\theta}_2, \hat{\alpha}_{2i})\right]$. Propositions 1 and 2, and Lemma 6 establish the uniform consistency of the components of the estimator of $\bar{\mathcal{B}}_2$. The result for the entire expression then follows by the continuous mapping theorem and a LLN, where the consistency of the truncated estimators of the spectral variances and covariances follow by Lemma 6 in HK. ■

Lemma 6 *Assume that Conditions 1, 2, 3, and 4 hold. Let $\hat{\theta}_2$ and $\hat{\alpha}_{2i}$, $i = 1, ..., n$, be (uniformly) consistent estimators of θ_{20} and α_{2i0}, i.e., $\hat{\theta}_2 - \theta_{20} = o_p(1)$ and $\hat{\alpha}_{2i} - \alpha_{2i0} = o_{up}(1)$. Let $h(w_{it}, \lambda_{it}; \theta_2, \alpha_{2i})$ be a function such that (i) $h(w_{it}, \lambda_{it}; \theta, \alpha)$ is continuous in $\gamma_{it} := (\lambda_{it}, \theta_2', \alpha_{2i})$; (ii) there exists a function $M(w_{it}, \lambda_{it})$ such that $|h(w_{it}, \lambda_{it}; \theta, \alpha)| \leq M(w_{it}, \lambda_{it})$ and $|\partial h(w_{it}, \lambda_{it}; \theta, \alpha)/\partial \gamma_{it}| \leq M(w_{it}, \lambda_{it})$, with $\sup_i E[M(w_{it}, \lambda_{it})^2] < \infty$. Then, for any $\eta > 0$, we have*

$$\Pr\left\{\max_{1\leq i\leq n}\left|\hat{\bar{h}}_i(\bar{\theta}_2, \bar{\alpha}_2) - h_i(\theta_{20}, \alpha_{2i0})\right| \geq \eta\right\} = o(T^{-1}). \tag{81}$$

where $(\bar{\theta}_2', \bar{\alpha}_{2i})'$ lies between $(\hat{\theta}_2', \hat{\alpha}_{2i})'$ and $(\theta_{20}', \alpha_{2i0})'$, and

$$\hat{\bar{h}}_i(\theta, \alpha) = \hat{E}_T\left[h_i(w_{it}, \bar{\lambda}_{it}; \theta, \alpha)\right], \tag{82}$$

$$h_i(\theta, \alpha) = E_T\left[h_i(w_{it}, \lambda_{it}; \theta, \alpha)\right], \tag{83}$$

with $\bar{\lambda}_{it} := \lambda(w_{it}; \bar{\theta}_1, \bar{\alpha}_{1i})$, and $(\bar{\theta}_1', \bar{\alpha}_{1i})$ lies between $(\theta_{10}', \alpha_{1i0})$ and $(\tilde{\theta}_1', \tilde{\alpha}_{1i})$.

Proof. By triangle inequality,

$$\begin{aligned}\max_{1\leq i\leq n}\left|\hat{\bar{h}}_i(\hat{\theta}_2, \hat{\alpha}_{2i}) - h_i(\theta_{20}, \alpha_{2i0})\right| &\leq \max_{1\leq i\leq n}\left|\hat{\bar{h}}_i(\hat{\theta}_2, \hat{\alpha}_{2i}) - h_i(\hat{\theta}_2, \hat{\alpha}_{2i})\right| \\ &+ \max_{1\leq i\leq n}\left|h_i(\hat{\theta}_2, \hat{\alpha}_{2i}) - h_i(\theta_{20}, \alpha_{2i0})\right| \end{aligned} \tag{84}$$

Then, the first term is uniformly bounded by a similar argument as in the proof of Lemma 4. The second term is bounded by $\sup_i E\left[M(w_{it}, \lambda_{it})\right]\left(|\bar{\theta}_2 - \theta_{20}| + \max_{1\leq i\leq n}|\bar{\alpha}_{2i} - \alpha_{2i0}|\right) = o_p(1)$. ■

B.6 Stochastic Expansion for $\hat{\alpha}_{2i0} = \hat{\alpha}_{2i}(\theta_{20})$

Lemma 7 *Assume that Conditions 1, 2, 3, and 4 hold, and $m \to \infty$ such that $m/T^{1/2} \to 0$. We then have*

$$\sqrt{T}(\hat{\alpha}_{2i0} - \alpha_{2i0}) = \psi_{2i} + \frac{1}{\sqrt{T}}R_{21i}, \tag{85}$$

where

$$\begin{aligned}
\psi_{2i} &= -\sqrt{T}E_T\left[v_{2it\alpha}\right]^{-1}\left\{\hat{E}_T\left[v_{2it}+E_T\left[v_{2it\lambda}\lambda_{it\alpha}\right]\psi_{1it}\right]+E_T\left[u_{2it\lambda}\mathcal{J}_{it}^{\lambda}\right]\varphi_1/\sqrt{nT}\right\} \\
&= o_{up}(T^{1/10+\upsilon}), \\
R_{21i} &= o_{up}(T^{2/10+2\upsilon}) = o_{up}(\sqrt{T}). \qquad (86)
\end{aligned}$$

for some $0<\upsilon<(100q+120)^{-1}$, *where* $q=\max\{q_1,q_2\}$.

Proof. By two first order Taylor expansions of the FOC for $\hat{\alpha}_{2i0}$ with respect to $\tilde{\lambda}_{it}$ and $\hat{\alpha}_{2i0}$, respectively, and Lemma 2, we have

$$\begin{aligned}
0 &= \hat{E}_T\left[\tilde{v}_{2it}(\theta_{20},\hat{\alpha}_{2i0})\right]=\hat{E}_T\left[v_{2it}(\theta_{20},\hat{\alpha}_{2i0})\right]+\hat{E}_T\left[\bar{v}_{2it\lambda}(\hat{\alpha}_{2i},\theta_{20})(\tilde{\lambda}_{it}-\lambda_{it})\right] \\
&= \hat{E}_T\left[v_{2it}\right]+\hat{E}_T\left[v_{2it\alpha}(\theta_{20},\bar{\alpha}_{2i0})\right](\hat{\alpha}_{2i0}-\alpha_{2i0})+\hat{E}_T\left[\bar{v}_{2it\lambda}(\hat{\alpha}_{2i},\theta_{20})(\tilde{\lambda}_{it}-\lambda_{it})\right], \qquad (87)
\end{aligned}$$

where $\bar{\alpha}_{2i}$ lies between α_{2i0} and $\hat{\alpha}_{2i0}$, and $\bar{v}_{2it\lambda}(\cdot)=v_{2\lambda}(w_{it},\bar{\lambda}_{it};\cdot)$ where $\bar{\lambda}_{it}$ lies between λ_{it} and $\tilde{\lambda}_{it}$. Next,

$$\begin{aligned}
\sqrt{T}(\hat{\alpha}_{2i0}-\alpha_{2i0}) &= -\underbrace{E_T\left[v_{2it\alpha}\right]^{-1}}_{=O_{up}(1)}\Bigg\{\underbrace{\sqrt{T}\hat{E}_T\left[v_{2it}(\theta_{20},\bar{\alpha}_{2i0})\right]}_{=o_{up}(T^{1/10+\upsilon})}+\underbrace{E_T\left[v_{2it\lambda}\sqrt{T}(\tilde{\lambda}_{it}-\lambda_{it})\right]}_{=o_{up}(T^{1/10+\upsilon})} \\
&+ \underbrace{\left(\hat{E}_T\left[v_{2it\alpha}\right]-E_T\left[v_{2it\alpha}\right]\right)}_{=o_{up}(1)}\sqrt{T}(\hat{\alpha}_{2i0}-\alpha_{2i0}) \\
&+ \underbrace{\sqrt{T}\left(\hat{E}_T\left[\bar{v}_{2it\lambda}(\hat{\alpha}_{2i},\theta_{20})(\tilde{\lambda}_{it}-\lambda_{it})\right]-E_T\left[v_{2it\lambda}(\tilde{\lambda}_{it}-\lambda_{it})\right]\right)}_{=o_{up}(1)}\Bigg\} \\
&= o_{up}(T^{1/10+\upsilon})+o_{up}\left(\sqrt{T}(\hat{\alpha}_{2i0}-\alpha_{2i0})\right) \qquad (88)
\end{aligned}$$

by Lemma 2 and Lemma 11 in HK. Next, the expression for ψ_{2i} follows from Lemma 2.

Finally, for the remainder term, we have

$$
\begin{aligned}
R_{21i} &= -\underbrace{E_T\left[v_{2it\alpha}\right]^{-1}}_{=O_{up}(1)}\Bigg\{\underbrace{\sqrt{T}\left(\hat{E}_T\left[v_{2it\alpha}(\theta_{20},\bar{\alpha}_{2i0})\right]-E_T\left[v_{2it\alpha}\right]\right)}_{=o_{up}(T^{1/10+\upsilon})}\underbrace{\sqrt{T}(\hat{\alpha}_{2i0}-\alpha_{2i0})}_{=o_{up}(T^{1/10+\upsilon})} \\
&+ \underbrace{\sqrt{T}\left(\hat{E}_T\left[\bar{v}_{2it\lambda}(\hat{\alpha}_{2i},\theta_{20})\sqrt{T}(\tilde{\lambda}_{it}-\lambda_{it})\right]-E_T\left[v_{2it\lambda}\sqrt{T}(\tilde{\lambda}_{it}-\lambda_{it})\right]\right)}_{=o_{up}(T^{2/10+2\upsilon})} \\
&+ \underbrace{\hat{E}_T\left[v_{2it\lambda}(\bar{\beta}^{\lambda}_{it}+R^{\lambda}_{it}/\sqrt{T})\right]}_{=o_{up}(T^{2/10+2\upsilon})}\Bigg\} = o_{up}(T^{2/10+2\upsilon}), \qquad (89)
\end{aligned}
$$

by Lemma 11 in HK. ■

Lemma 8 *Assume that Conditions 1, 2, 3, and 4 hold, and $m \to \infty$ such that $m/T^{1/2} \to 0$. We then have*

$$
\sqrt{T}(\hat{\alpha}_{2i0}-\alpha_{2i0}) = \psi_{2i} + \frac{1}{\sqrt{T}}Q_{21i} + \frac{1}{T}R_{22i}, \qquad (90)
$$

where

$$
\begin{aligned}
Q_{21i} &= -E_T\left[v_{2it\alpha}\right]^{-1}\Bigg\{\sqrt{T}\left(\hat{E}_T\left[v_{2it\alpha}\right]-E_T\left[v_{2it\alpha}\right]\right)\psi_{2i}+\frac{1}{2}E_T\left[v_{2it\alpha\alpha}\right]\psi_{2i}^2 \\
&+ E_T\left[v_{2\lambda\alpha it}\lambda_{it\alpha}\right]\psi_{1i}\psi_{2i}+\sqrt{T}\left(\hat{E}_T\left[v_{2i\lambda}\lambda_{i\alpha}\right]-E_T\left[v_{2i\lambda}\lambda_{i\alpha}\right]\right)\psi_{1i} \\
&+ E_T\left[v_{2it\lambda}\bar{\beta}^{\lambda}_{it}\right]+\frac{1}{2}E_T\left[v_{2it\lambda}\left(\tilde{\psi}^{\lambda}_{it}\right)^2\right]\Bigg\} = o_{up}(T^{2/10+2\upsilon}), \\
R_{22i} &= o_{up}(\sqrt{T}), \qquad (91)
\end{aligned}
$$

for some $0 < \upsilon < (100q+120)^{-1}$, where $q = \max\{q_1, q_2\}$.

Proof. By two second order Taylor expansions of the FOC for $\hat{\alpha}_{2i0}$, $0 = \hat{E}_T\left[\tilde{v}_{2it}(\theta_{20},\hat{\alpha}_{2i0})\right]$, with respect to $\tilde{\lambda}_{it}$ and $\hat{\alpha}_{2i0}$, respectively, we have

$$
\begin{aligned}
0 &= \hat{E}_T\left[v_{2it}(\theta_{20},\hat{\alpha}_{2i0})\right]+\hat{E}_T\left[v_{2it\lambda}(\theta_{20},\hat{\alpha}_{2i0})(\tilde{\lambda}_{it}-\lambda_{it})\right]+\frac{1}{2}\hat{E}_T\left[\bar{v}_{2it\lambda\lambda}(\hat{\alpha}_{2i},\theta_{20})(\tilde{\lambda}_{it}-\lambda_{it})^2\right] \\
&= \hat{E}_T\left[v_{2it}\right]+\hat{E}_T\left[v_{2it\alpha}\right](\hat{\alpha}_{2i0}-\alpha_{2i0})+\frac{1}{2}\hat{E}_T\left[v_{2it\alpha\alpha}(\theta_{20},\bar{\alpha}_{2i0})\right](\hat{\alpha}_{2i0}-\alpha_{2i0})^2 \\
&+\hat{E}_T\left[v_{2it\lambda}(\tilde{\lambda}_{it}-\lambda_{it})\right]+\hat{E}_T\left[v_{2it\lambda\alpha}(\bar{\alpha}_{2i0},\theta_{20})(\tilde{\lambda}_{it}-\lambda_{it})\right](\hat{\alpha}_{2i0}-\alpha_{2i0}) \\
&+\frac{1}{2}\hat{E}_T\left[\bar{v}_{2it\lambda\lambda}(\hat{\alpha}_{2i},\theta_{20})(\tilde{\lambda}_{it}-\lambda_{it})^2\right], \qquad (92)
\end{aligned}
$$

where $\bar{\alpha}_{2i}$ lies between α_{2i0} and $\hat{\alpha}_{2i0}$, and $\bar{v}_{2it\lambda\lambda}(\cdot) = v_{2\lambda\lambda}(w_{it}, \bar{\lambda}_{it}; \cdot)$ where $\bar{\lambda}_{it}$ lies between λ_{it} and $\tilde{\lambda}_{it}$. The expression for Q_{21i} can be obtained from the expansion of $\tilde{\lambda}_{it}$ in Lemma 2 following a similar argument as in Lemma A4 in Newey and Smith (2004). The uniform rate of convergence for Q_{1i} follows by Lemma 7, and Lemma 11 in HK. For the remainder term, we have

$$\begin{aligned}
R_{22i} &= -E_T\left[v_{2it\alpha}\right]^{-1}\Big\{\sqrt{T}\left(\hat{E}_T\left[v_{2it\alpha}\right] - E_T\left[v_{2it\alpha}\right]\right)R_{21i} \\
&+ \frac{\sqrt{T}}{2}\left(\hat{E}_T\left[v_{2i\alpha\alpha}(\theta_{20}, \bar{\alpha}_{2i0})\right] - E_T\left[v_{2i\alpha\alpha}\right]\right)T(\hat{\alpha}_{2i0} - \alpha_{2i0})^2 R_{21i} \\
&+ \frac{1}{2}E_T\left[v_{2it\alpha\alpha}\right]R_{21i}\left(\sqrt{T}(\hat{\alpha}_{2i0} - \alpha_{2i0}) + \psi_{2i}\right) \\
&+ \sqrt{T}\left(\hat{E}_T\left[v_{2it\lambda}\left(\bar{\beta}_{it}^{\lambda} + R_{it}^{\lambda}/\sqrt{T}\right)\right] - E_T\left[v_{2it\lambda}\left(\bar{\beta}_{it}^{\lambda} + R_{it}^{\lambda}/\sqrt{T}\right)\right]\right) \\
&+ E_T\left[v_{2it\lambda}R_{it}^{\lambda}\right] + E_T\left[v_{2it\alpha\lambda}\sqrt{T}\left(\tilde{\lambda}_{it} - \lambda\right)\right]R_{21i} \\
&+ E_T\left[v_{2it\alpha\lambda}\left(\bar{\beta}_{it}^{\lambda} + R_{it}^{\lambda}\right)\right]\sqrt{T}(\hat{\alpha}_{2i0} - \alpha_{2i0}) \\
&+ \sqrt{T}\left(\hat{E}_T\left[v_{2it\alpha\lambda}(\theta_{20}, \bar{\alpha}_{2i0})\sqrt{T}\left(\tilde{\lambda}_{it} - \lambda\right)\right] - E_T\left[v_{2it\alpha\lambda}\sqrt{T}\left(\tilde{\lambda}_{it} - \lambda\right)\right]\right)\sqrt{T}(\hat{\alpha}_{2i0} - \alpha_{2i0}) \\
&+ \frac{\sqrt{T}}{2}\left(\hat{E}_T\left[\bar{v}_{2it\lambda\lambda}(\theta_{20}, \hat{\alpha}_{2i0})T\left(\tilde{\lambda}_{it} - \lambda_{it}\right)^2\right] - E_T\left[v_{2it\lambda\lambda}T\left(\tilde{\lambda}_{it} - \lambda_{it}\right)^2\right]\right) \\
&+ \frac{1}{2}E_T\left[v_{2it\lambda\lambda}\left(\bar{\beta}_{it}^{\lambda} + R_{it}^{\lambda}\right)\left(\sqrt{T}\left(\tilde{\lambda}_{it} - \lambda_{it}\right) + \tilde{\psi}_{it}^{\lambda}\right)\right] = o_{up}\left(\max\left\{T^{3/10+3v}, \sqrt{T}\right\}\right). \quad (93)
\end{aligned}$$

The uniform rates of convergence follow by Lemma 2 and Lemma 11 in HK. ■

Lemma 9 *Assume that Conditions 1, 2, 3, and 4 hold, and $m \to \infty$ such that $m/T^{1/2} \to 0$. We then have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{2i} \xrightarrow{d} N(0, \bar{\sigma}_2^2 := E_n\left[\bar{\sigma}_{2i}^2\right]), \quad (94)$$

$$\frac{1}{n}\sum_{i=1}^{n}Q_{21i} \xrightarrow{p} \bar{\beta}_2 := E_n\left[\bar{\beta}_{2i}\right], \quad (95)$$

where

$$\bar{\sigma}_{2i}^2 = \bar{E}_T\left[\psi_{2it}\psi_{2is}\right], \tag{96}$$

$$\psi_{2it} = -E_T\left[v_{2it\alpha}\right]^{-1}\left\{v_{2it} + E_T\left[v_{2it\lambda}\lambda_{it\alpha}\right]\psi_{1it} + E_T\left[v_{2it\lambda}\mathcal{J}_{it}^{\lambda}\right]\varphi_{1it}\right\}, \tag{97}$$

$$\begin{aligned}\bar{\beta}_{2i} = & -E_T\left[v_{2it\alpha}\right]^{-1}\Bigg\{\bar{E}_T\left[v_{2it\alpha}\psi_{2is}\right] + \frac{1}{2}E_T\left[v_{2it\alpha\alpha}\right]\bar{\sigma}_{2i}^2 + \bar{E}_T\left[v_{2it\lambda}\lambda_{it\alpha}\psi_{1is}\right] \\ & + E_T\left[v_{2\lambda\alpha it}\lambda_{it\alpha}\right]\bar{\sigma}_{12i}^2 + E_T\left[v_{2it\lambda}\lambda_{it\alpha}\right]\bar{\beta}_{1i} + \frac{1}{2}E_T\left[v_{2it\lambda}\lambda_{it\alpha\alpha}\right]\bar{\sigma}_{1i}^2 \\ & + \frac{1}{2}E_T\left[v_{2it\lambda\lambda}\lambda_{it\alpha}^2\right]\bar{\sigma}_{1i}^2\Bigg\},\end{aligned} \tag{98}$$

$$\bar{\sigma}_{12i}^2 = \bar{E}_T\left[\psi_{1it}\psi_{2is}\right]. \tag{99}$$

Proof. The result for the influence functions ψ_{2i}'s follows by Lemma 7, and Lemma 3 in HK. The result for the Q_{21i}'s can be shown using a similar argument as in the derivation of the limiting behavior of $\theta^{\epsilon\epsilon}(0)$ in the proof of Theorem 1 in HK. In particular, uniform convergence of Q_{21i}, $i = 1, ..., n$, can be established using Corollary A.2 of Hall and Heyde (1980), and Lemma 3 in HK. ■

B.7 Stochastic Expansion for $\hat{E}_T\left[\hat{\tilde{u}}_{2it}(\theta_{20})\right]$

Lemma 10 *Assume that Conditions 1, 2, 3, and 4 hold, and* $m \to \infty$ *such that* $m/T^{1/2} \to 0$. *We then have*

$$\hat{E}_T\left[\hat{\tilde{u}}_{2it}(\theta_{20})\right] = \frac{1}{\sqrt{T}}\psi_{2i}^u + \frac{1}{T}Q_{21i}^u + \frac{1}{T^{3/2}}R_{22i}^u, \tag{100}$$

where

$$\psi_{2i}^u = \sqrt{T}\hat{E}_T\left[u_{2it}\right] + E_T\left[u_{2it\alpha}\right]\psi_{2i} + E_T\left[u_{2it\lambda}\tilde{\psi}_{it}^{\lambda}\right] = o_{up}(T^{1/10+\upsilon}), \tag{101}$$

$$\begin{aligned}Q_{21i}^u = & E_T\left[u_{2it\alpha}\right]Q_{21i} + \sqrt{T}\left(\hat{E}_T\left[u_{2i\alpha}\right] - E_T\left[u_{2i\alpha}\right]\right)\psi_{2i} + \frac{1}{2}E_T\left[u_{2it\alpha\alpha}\right]\psi_{2i}^2 + E_T\left[u_{2it\lambda}\bar{\beta}_{it}^{\lambda}\right] \\ + & E_T\left[u_{2it\lambda\alpha}\tilde{\psi}_{it}^{\lambda}\right]\psi_{2i} + \sqrt{T}\left(\hat{E}_T\left[u_{2it\lambda}\tilde{\psi}_{it}^{\lambda}\right] - E_T\left[u_{2it\lambda}\tilde{\psi}_{it}^{\lambda}\right]\right) + \frac{1}{2}E_T\left[u_{2it\lambda\lambda}\left(\tilde{\psi}_{it}^{\lambda}\right)^2\right] \\ = & o_{up}(T^{2/10+2\upsilon}),\end{aligned} \tag{102}$$

$$R_{22i}^u = o_{up}(\sqrt{T}), \tag{103}$$

for some $0 < \upsilon < (100 + 120q)^{-1}$, *where* $q = \max\{q_1, q_2\}$.

Proof. By two second order expansions of $\hat{E}_T\left[\hat{\tilde{u}}_{2it}(\theta_{20})\right]$ with respect to $\tilde{\lambda}_{it}$ and $\hat{\alpha}_{2i0}$,

respectively, we have

$$\begin{aligned}
\hat{E}_T\left[\hat{\tilde{u}}_{2it}(\theta_{20})\right] &= \hat{E}_T\left[\hat{u}_{2it\lambda}\right] + \hat{E}_T\left[\hat{u}_{2it\lambda}\left(\tilde{\lambda}_{it}-\lambda_{it}\right)\right] + \frac{1}{2}\hat{E}_T\left[\hat{\tilde{u}}_{2it\lambda\lambda}\left(\tilde{\lambda}_{it}-\lambda_{it}\right)^2\right] \\
&= \hat{E}_T\left[u_{2it\alpha}\right] + \hat{E}_T\left[u_{2it\alpha}\right](\hat{\alpha}_{2i0}-\alpha_{2i0}) + \frac{1}{2}\hat{E}_T\left[u_{2it\alpha}(\theta_{20},\bar{\alpha}_{2i0})\right](\hat{\alpha}_{2i0}-\alpha_{2i0})^2 \\
&+ \hat{E}_T\left[u_{2it\lambda}\left(\tilde{\lambda}_{it}-\lambda_{it}\right)\right] + \hat{E}_T\left[u_{2it\lambda\alpha}(\theta_{20},\bar{\alpha}_{2i0})\left(\tilde{\lambda}_{it}-\lambda_{it}\right)\right](\hat{\alpha}_{2i0}-\alpha_{2i0}) \\
&+ \frac{1}{2}\hat{E}_T\left[\hat{\tilde{u}}_{2it\lambda\lambda}\left(\tilde{\lambda}_{it}-\lambda_{it}\right)^2\right], \qquad (104)
\end{aligned}$$

where $\hat{\tilde{u}}_{2it\lambda\lambda} = u_2(w_{it},\bar{\lambda}_{it};\theta_{20},\hat{\alpha}_{2i0})$, $\bar{\lambda}_{it}$ lies between $\tilde{\lambda}_{it}$ and λ_{it}, and $\bar{\alpha}_{2i0}$ is between $\hat{\alpha}_{2i0}$ and α_{2i0}. The expressions for ψ^u_{2i} and Q^u_{21i} can be obtained using Lemmas 2 and 8, after some algebra. The properties for these terms follow by Lemma 11 from HK. For the remainder term, we have

$$\begin{aligned}
R^u_{22i} &= E_T\left[u_{2it\alpha}\right]R_{22i} + \sqrt{T}\left(\hat{E}_T\left[u_{2i\alpha}\right] - E_T\left[u_{2i\alpha}\right]\right)R_{21i} \\
&+ \frac{1}{2}E_T\left[u_{2it\alpha}\right]\left(R_{21i}\sqrt{T}(\hat{\alpha}_{2i0}-\alpha_{2i0}) + \psi_{2i}R_{21i}\right) \\
&+ \frac{1}{2}\sqrt{T}\left(\hat{E}_T\left[u_{2it\alpha\alpha}(\theta_{20},\bar{\alpha}_{2i0})\right] - E_T\left[u_{2it\alpha\alpha}\right]\right)T\left(\hat{\alpha}_{2i0}-\alpha_{2i0}\right)^2 + E_T\left[u_{it\lambda}R^\lambda_{it}\right] \\
&+ \sqrt{T}\left(\hat{E}_T\left[u_{2it\lambda}\left(\bar{\beta}^\lambda_{it}+R^\lambda_{it}\right)\right] - E_T\left[u_{2it\lambda}\left(\bar{\beta}^\lambda_{it}+R^\lambda_{it}\right)\right]\right) \\
&+ E_T\left[u_{2it\lambda\alpha}\left(\bar{\beta}^\lambda_{it}+R^\lambda_{it}\right)\right]\sqrt{T}\left(\hat{\alpha}_{2i0}-\alpha_{2i0}\right) \\
&+ \sqrt{T}\left(\hat{E}_T\left[u_{2it\lambda\alpha}(\theta_{20},\bar{\alpha}_{2i0})\sqrt{T}\left(\tilde{\lambda}_{it}-\lambda_{it}\right)\right] - E_T\left[u_{2it\lambda\alpha}\sqrt{T}\left(\tilde{\lambda}_{it}-\lambda_{it}\right)\right]\right)\sqrt{T}\left(\hat{\alpha}_{2i0}-\alpha_{2i0}\right) \\
&+ E_T\left[u_{2it\lambda\alpha}\tilde{\psi}^\lambda_{it}\right]R_{21i} + \frac{1}{2}E_T\left[u_{2it\lambda\lambda}\left(\bar{\beta}^\lambda_{it}+R^\lambda_{it}\right)\left(\sqrt{T}\left(\tilde{\lambda}_{it}-\lambda_{it}\right)+\tilde{\psi}^\lambda_{it}\right)\right] \\
&+ \frac{1}{2}\sqrt{T}\left(\hat{E}_T\left[u_{2it\lambda\lambda}(\theta_{20},\bar{\alpha}_{2i0})T\left(\tilde{\lambda}_{it}-\lambda_{it}\right)^2\right] - E_T\left[u_{2it\lambda\lambda}\left(\tilde{\lambda}_{it}-\lambda_{it}\right)^2\right]\right) \\
&= o_{up}\left(T^{3/10+3\upsilon},\sqrt{T}\right). \qquad (105)
\end{aligned}$$

Then, the uniform order in probability for R^u_{22i} follows by the properties of the components in the expansion of $\hat{\alpha}_{2i0}$ and $\tilde{\lambda}_{it}$, and Lemma 11 in HK. ■

Lemma 11 *Assume that Conditions 1, 2, 3, and 4 hold, and* $m \to \infty$ *such that* $m/T^{1/2} \to 0$. *We then have*

$$\begin{aligned}
\sqrt{n}E_n\left[\psi^u_{2i}\right] &\xrightarrow{d} N(0,\bar{\Omega}_2 := E_n\left[\bar{\Omega}_{2i}\right]), \qquad (106) \\
\hat{E}_n\left[Q^u_{21i}\right] &\xrightarrow{p} \bar{b}_2 := E_n\left[\bar{b}_{2i}\right], \qquad (107)
\end{aligned}$$

where

$$\bar{\Omega}_{2i} = \bar{E}_T\left[U_{2it}U'_{2is}\right], \tag{108}$$

$$U_{2it} = u_{2it} + E_T\left[u_{2it\alpha}\right]\psi_{2it} + E_T\left[u_{2it\lambda}\lambda_{it\alpha}\right]\psi_{1it} + E_T\left[u_{2it\lambda}\mathcal{J}^{\lambda}_{it}\right]\varphi_{1it}, \tag{109}$$

$$\begin{aligned}\bar{b}_{2i} &= \bar{E}_T\left[Q^u_{21i}\right] = \bar{E}_T\left[u_{2it\alpha}\psi_{2is}\right] + E_T\left[u_{2it\alpha}\right]\bar{\beta}_{2i} + \frac{1}{2}\bar{\sigma}^2_{2i}E_T\left[u_{2it\alpha\alpha}\right] + E_T\left[u_{2it\lambda\alpha}\lambda_{it\alpha}\right]\bar{\sigma}_{12i} \\ &\quad + \bar{E}_T\left[u_{2it\lambda}\lambda_{it\alpha}\psi_{1is}\right] + E_T\left[u_{2it\lambda}\left(\lambda_{it\alpha}\bar{\beta}_{1i} + \frac{1}{2}\lambda_{it\alpha\alpha}\bar{\sigma}^2_{1i}\right)\right] + \frac{1}{2}E_T\left[u_{2it\lambda\lambda}\lambda^2_{it\alpha}\right]\bar{\sigma}^2_{1i},\end{aligned} \tag{110}$$

Proof. The result for the influence functions ψ^u_{2i}'s follows by Lemma 3 in HK. The result for the Q^u_{21i}'s can be shown using a similar argument as in the derivation of the limiting behavior of $\theta^{\epsilon\epsilon}(0)$ in the proof of Theorem 1 in HK. In particular, uniform convergence of Q^u_{21i}, $i = 1, ..., n$, can be established using Corollary A.2 of Hall and Heyde (1980), and Lemma 3 in HK. ■

Lemma 12 *Assume that Conditions 1, 2, 3, and 4 hold, and* $m \to \infty$ *such that* $m/T^{1/2} \to 0$. *We then have, for* $\rho = \lim_{n,T\to\infty} n/T$,

$$\sqrt{nT}E_n\left[\hat{E}_T\left[\hat{\bar{u}}_{2it}(\theta_{20})\right]\right] \xrightarrow{d} \mathcal{N}\left(\sqrt{\rho}\,\bar{b}_2, \bar{\Omega}_2\right), \tag{111}$$

where $\bar{b}_2$ *and* $\bar{\Omega}_2$ *are defined in Lemma 11.*

Proof. From Lemma 10, we have

$$\sqrt{nT}E_n\left[\hat{E}_T\left[\hat{\bar{u}}_{2it}(\theta_{20})\right]\right] = \underbrace{\sqrt{n}\hat{E}_n\left[\psi^u_{2i}\right]}_{=O_p(1)} + \underbrace{\sqrt{\frac{n}{T}}\hat{E}_n\left[Q^u_{21i}\right]}_{=O_p(1)} + \underbrace{\frac{\sqrt{n}}{T}\hat{E}_n\left[R^u_{2i}\right]}_{=o_p(1)}. \tag{112}$$

Then, the result follows by Lemma 11 and Slutsky Theorem. ■

References

[1] Arellano, M., and R. Carrasco (2003), "Discrete Choice Panel Data Models with Predetermined Variables," *Journal of Econometrics* 115 (1), 125-157.

[2] Arellano, M., and J. Hahn (2005), "Understanding Bias in Nonlinear Panel Models: Some Recent Developments," Mimeo, CEMFI.

[3] Carro, J. M. (2006), "Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects," *Journal of Econometrics*, doi: 1016/j-jeconom.2006.07.023.

[4] Fernández-Val, I. (2005), "Estimation of Structural Parameters and Marginal Effects in Binary Choice Panel Data Models with Fixed Effects," working paper, MIT Department of Economics.

[5] Gayle, G. L., and C. Viauroux (2005), "Root-N Consistent Semi-parametric Estimators of A Dynamic Panel Sample Selection Model," *Journal of Econometrics* forthcoming.

[6] Gourieroux, C., Monfort, A., Renault E., and A. Trognon (1987), "Generalised residuals," *Journal of Econometrics* 34, 5-32.

[7] Greene, W.H. (2002), "The Behavior of the Fixed Effects Estimator in Nonlinear Models,"unpublished manuscript, New York University.

[8] Greene, W.H. (2004), "Fixed Effects and Bias Due to the Incidental Parameters Problem in the Tobit Model,"unpublished manuscript, New York University.

[9] 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.

[10] Hahn, J., and G. Kuersteiner (2003), "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects," unpublished manuscript. UCLA.

[11] Hahn, J., and W. Newey (2004), "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica* 72, 1295-1319.

[12] Hall, P., and C. Heyde (1980), *Martingale Limit Theory and its Applications*, Academic Press.

[13] Heckman, J. J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931-959.

[14] Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica,* 47, 153-161.

[15] 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 Panel Data with Econometric Applications*, 179-195.

[16] Honoré, B. E. (1993), "Orthogonality conditions for Tobit models with fixed effects and lagged dependent variables," *Journal of Econometrics* 59, pp 35-61.

[17] Honoré, B. E., and A. Lewbel (2002), "Semiparametric Binary Choice Panel Data Models Without Strictly Exogenous Regressors," *Econometrica* 70, 2053-2063.

[18] Hu, L. (2002), "Estimation of a Censored Dynamic Panel Data Model," *Econometrica* 70, 2499-2517.

[19] Kyriazidou, E. (1997), "Estimation of a Panel Data Sample Selection Model," *Econometrica*, 65, 1335-1364.

[20] Kyriazidou, E. (2001), "Estimation of Dynamic Panel Data Sample Selection Models," *Review of Economic Studies*, 68, 543–572.

[21] Lancaster, T. (2002), "Orthogonal Parameters and Panel Data," *Review of Economic Studies* 69, 647-666.

[22] Lee, L-F., Maddala, G S, and R.P. Trost (1980), "Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity," *Econometrica*, 48(2), 491-503.

[23] MacKinnon, J. G., and H. White (1985), "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties," *Journal of Econometrics* 29(3), pp 305-325.

[24] Newey, W. K. (1984), "A method of moments interpretation of sequential estimators," *Economics Letters* 14(2-3), pp 201-206.

[25] Newey, W.K., and R. Smith (2004), "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica* 72, pp 219-255.

[26] Neyman, J., and E.L. Scott (1948), "Consistent Estimates Based on Partially Consistent Observations," *Econometrica* 16, 1-32.

[27] Ridder, G. (1990), "Attrition in Multi-Wave Panel Data," in J. Hartog, G. Ridder and J. Theeuwes, eds., Panel Data and Labor Market Studies, Elsevier, North Holland.

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

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

[30] Vella, F. (1993), "A Simple Estimator for Simultaneous Models with Censored Endogenous Regressors, *International Economic Review,* 34, 441-457.

[31] 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, 163-183.

[32] Vella, f., and M. Verbeek (1999), "Two-Step Estimation of Panel Data Models with Censored Endogenous Regressors and Selection Bias", *Journal of Econometrics*, 90, 239-263.

[33] Verbeek, M., and T. Nijman (1992), "Testing for Selectivity Bias in Panel Data Models", *International Economic Review*, 33(3), 681-703.

[34] Wooldridge, J. M. (1995), "Selection Corrections for Panel Data Models under Conditional Mean Independence Assumptions", *Journal of Econometrics*, 68, 115-132.

[35] Wooldridge, J. (2001), "A Framework for Estimating Dynamic, Unobserved Effects Panel Data Models with Possible Feedback to Future Explanatory Variables," *Economics Letters* 68, 245-250.

[36] Woutersen, T. M. (2002), "Robustness Against Incidental Parameters," unpublished manuscript, University of Western Ontario.

Table 1: Heckman Selection Model - Probit First Stage, n =100

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE	nobs
				A. T = 6				
A.1. Coefficient β_1 (true = 1)								
H-MLE	1.32	1.31	0.190	0.51	0.62	0.827	0.308	463
H-BC1	1.05	1.05	0.131	0.04	0.08	1.061	0.094	463
H-BC2	1.14	1.13	0.165	0.15	0.25	0.881	0.142	463
H-BC3	0.98	0.97	0.117	0.04	0.07	1.113	0.082	463
A.2. Coefficient β_2 (true = 1)								
H-MLE	1.32	1.31	0.189	0.52	0.64	0.833	0.306	463
H-BC1	1.05	1.05	0.128	0.03	0.08	1.083	0.089	463
H-BC2	1.14	1.13	0.164	0.14	0.23	0.884	0.139	463
H-BC3	0.98	0.98	0.114	0.03	0.07	1.144	0.077	463
				B. T = 8				
B.1. Coefficient β_1 (true = 1)								
H-MLE	1.22	1.22	0.138	0.43	0.55	0.892	0.218	671
H-BC1	1.07	1.06	0.115	0.08	0.14	1.000	0.089	671
H-BC2	1.08	1.07	0.119	0.10	0.17	0.970	0.091	671
H-BC3	1.00	1.00	0.103	0.04	0.07	1.059	0.071	671
B.2. Coefficient β_2 (true = 1)								
H-MLE	1.22	1.21	0.137	0.41	0.54	0.895	0.212	671
H-BC1	1.06	1.06	0.114	0.08	0.14	1.003	0.083	671
H-BC2	1.07	1.07	0.119	0.09	0.17	0.968	0.089	671
H-BC3	1.00	1.00	0.102	0.03	0.07	1.063	0.072	671
				C. T = 12				
C.1. Coefficient β_1 (true = 1)								
H-MLE	1.14	1.13	0.102	0.31	0.44	0.890	0.133	1089
H-BC1	1.04	1.04	0.092	0.09	0.13	0.953	0.065	1089
H-BC2	1.03	1.03	0.091	0.09	0.13	0.959	0.061	1089
H-BC3	1.00	1.00	0.086	0.06	0.11	0.985	0.056	1089
C.2. Coefficient β_2 (true = 1)								
H-MLE	1.13	1.13	0.098	0.30	0.41	0.934	0.131	1089
H-BC1	1.04	1.03	0.088	0.06	0.12	0.997	0.062	1089
H-BC2	1.03	1.03	0.087	0.05	0.11	1.006	0.059	1089
H-BC3	1.00	1.00	0.082	0.05	0.08	1.033	0.055	1089

Notes : 1,000 replications. MLE denotes uncorrected probit FE estimator; BC1 denotes Hahn and Newey (2004) bias-corrected estimator based on Bartlett equalities; BC2 denotes Hahn and Newey (2004) bias-corrected estimator based on general estimating equations; BC3 denotes Fernandez-Val (2005) bias corrected estimator.

Table 2: Heckman Selection Model - OLS Second Stage, coefficient θ, n =100

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE	nobs
				A. T = 6*				
OLS	1.18	1.18	0.102	0.52	0.63	0.848	0.176	285
H-1	1.01	1.01	0.119	0.10	0.15	0.865	0.077	285
H-MLE	1.07	1.07	0.115	0.14	0.21	0.897	0.094	285
H-JACK	1.00	1.00	0.133	0.15	0.23	0.737	0.090	285
H-BC1	1.03	1.03	0.130	0.11	0.17	0.836	0.090	285
H-BC2	1.04	1.04	0.122	0.12	0.18	0.858	0.090	285
H-BC3	1.03	1.03	0.124	0.11	0.17	0.864	0.089	285
				B. T = 8*				
OLS	1.17	1.17	0.082	0.65	0.73	0.892	0.174	390
H-1	1.00	0.99	0.098	0.08	0.15	0.894	0.066	390
H-MLE	1.04	1.03	0.096	0.10	0.16	0.924	0.069	390
H-JACK	0.99	0.99	0.107	0.12	0.19	0.773	0.075	390
H-BC1	1.01	1.01	0.103	0.09	0.15	0.881	0.071	390
H-BC2	1.01	1.01	0.101	0.09	0.15	0.888	0.070	390
H-BC3	1.01	1.01	0.102	0.08	0.15	0.889	0.070	390
				C. T = 12				
OLS	1.18	1.18	0.063	0.83	0.88	0.941	0.176	592
H-1	1.00	1.00	0.078	0.08	0.14	0.908	0.054	592
H-MLE	1.03	1.03	0.078	0.09	0.15	0.922	0.055	592
H-JACK	1.00	1.00	0.084	0.12	0.19	0.786	0.056	592
H-BC1	1.01	1.01	0.083	0.09	0.14	0.889	0.053	592
H-BC2	1.01	1.01	0.082	0.09	0.14	0.892	0.053	592
H-BC3	1.00	1.01	0.082	0.09	0.14	0.892	0.052	592

Notes : 1,000 replications. Numerical algorithm fails to converge 98 times for T = 6 and 4 times for T = 8 for the Jackknife. These replications are not used for any of the estimators. H-1 denotes unfeasible estimator that uses (unobserved) true control function; MLE is the feasible version of H-1 that uses estimated control function; JACK denotes the two-step Jackknife bias-corrected estimator; BC1 denotes Hahn and Newey (2004) bias-corrected estimator based on Bartlett equalities; BC2 denotes Hahn and Newey (2004) bias-corrected estimator based on general estimating equations; BC3 denotes Fernandez-Val (2005) bias-corrected estimator. Standard errors account for heteroskedasticity and generated regressors, when relevant.

Table 3: Heckman Selection Model - OLS Second Stage, coefficient ρ (control function), n =100

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE	nobs
				A. T = 6*				
H-1	0.99	0.98	0.382	0.09	0.16	0.849	0.249	285
H-MLE	0.78	0.77	0.362	0.16	0.24	0.864	0.289	285
H-JACK	0.97	0.96	0.476	0.16	0.24	0.729	0.312	285
H-BC1	1.08	1.07	0.551	0.17	0.23	0.682	0.346	285
H-BC2	0.90	0.89	0.420	0.11	0.18	0.831	0.285	285
H-BC3	1.02	1.01	0.468	0.09	0.16	0.835	0.307	285
				B. T = 8*				
H-1	1.02	1.01	0.321	0.10	0.16	0.847	0.203	390
H-MLE	0.85	0.84	0.305	0.14	0.22	0.870	0.228	390
H-JACK	1.01	1.00	0.381	0.15	0.22	0.727	0.248	390
H-BC1	1.03	1.03	0.380	0.12	0.21	0.784	0.251	390
H-BC2	0.96	0.96	0.347	0.11	0.17	0.846	0.223	390
H-BC3	1.02	1.02	0.368	0.09	0.17	0.846	0.243	390
				C. T = 12				
H-1	1.01	1.01	0.244	0.08	0.15	0.891	0.157	592
H-MLE	0.90	0.90	0.240	0.11	0.18	0.897	0.173	592
H-JACK	1.01	1.00	0.275	0.12	0.20	0.772	0.191	592
H-BC1	1.01	1.01	0.270	0.09	0.16	0.862	0.177	592
H-BC2	0.99	0.99	0.265	0.09	0.15	0.881	0.179	592
H-BC3	1.02	1.01	0.271	0.08	0.14	0.882	0.178	592

Notes : 1,000 replications. Numerical algorithm fails to converge 98 times for T = 6 and 4 times for T = 8 for the Jackknife. These replications are not used for any of the estimators. H-1 denotes unfeasible estimator that uses (unobserved) true control function; MLE is the feasible version of H-1 that uses estimated control function; JACK denotes the two-step Jackknife bias-corrected estimator; BC1 denotes Hahn and Newey (2004) bias-corrected estimator based on Bartlett equalities; BC2 denotes Hahn and Newey (2004) bias-corrected estimator based on general estimating equations; BC3 denotes Fernandez-Val (2005) bias-corrected estimator. Standard errors account for heteroskedasticity and generated regressors, when relevant.

Table 4: Tobit CMLE, θ_1 (endogenous regressor), n =100

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE	nobs
				T = 6				
TOBIT	1.31	1.31	0.058	1.00	1.00	0.863	0.314	415
CMLE-1	0.98	0.98	0.074	0.11	0.17	0.868	0.050	415
CMLE	0.98	0.99	0.081	0.11	0.19	0.817	0.056	415
BC1	1.02	1.02	0.094	0.18	0.24	0.727	0.063	415
BC2	1.00	1.00	0.084	0.13	0.19	0.798	0.056	415
BC3	0.98	0.98	0.085	0.14	0.20	0.783	0.058	415
				T = 8				
TOBIT	1.33	1.33	0.048	1.00	1.00	0.882	0.328	614
CMLE-1	0.99	0.99	0.060	0.08	0.14	0.909	0.041	614
CMLE	0.99	0.99	0.066	0.10	0.17	0.841	0.043	614
BC1	1.01	1.01	0.072	0.13	0.19	0.796	0.047	614
BC2	1.00	1.00	0.068	0.11	0.16	0.832	0.044	614
BC3	0.99	0.99	0.067	0.11	0.17	0.833	0.044	614
				T = 12				
TOBIT	1.34	1.34	0.035	1.00	1.00	0.953	0.336	1016
CMLE-1	1.00	0.99	0.046	0.06	0.12	0.945	0.032	1016
CMLE	1.00	0.99	0.050	0.09	0.14	0.894	0.034	1016
BC1	1.00	1.00	0.052	0.10	0.16	0.860	0.035	1016
BC2	1.00	1.00	0.051	0.09	0.14	0.888	0.034	1016
BC3	1.00	1.00	0.050	0.09	0.15	0.886	0.033	1016

Notes: 1,000 replications. TOBIT denotes Tobit FE maximum likelihood estimator that does not account for endogeneity; CMLE-1 denotes unfeasible estimator that uses (unobserved) true control function; CMLE is the feasible version of CME-1 that uses estimated control function; BC1 denotes Hahn and Kuersteiner (2003) bias-corrected estimator based on Bartlett equalities; BC2 denotes Hahn and Kuersteiner (2003) bias-corrected estimator based on general estimating equations; BC3 denotes Fernandez-Val (2005) bias-corrected estimator; Standard errors account for generated regressors, when relevant.

Table 5: Tobit CMLE, θ_2 (lagged dep. variable), n =100

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE	nobs
				T = 6				
TOBIT	0.81	0.81	0.077	0.77	0.84	0.874	0.189	415
CMLE-1	0.84	0.84	0.073	0.71	0.78	0.859	0.162	415
CMLE	0.84	0.84	0.073	0.71	0.78	0.859	0.162	415
BC1	0.98	0.99	0.084	0.14	0.22	0.773	0.057	415
BC2	0.92	0.92	0.074	0.31	0.41	0.854	0.086	415
BC3	0.95	0.94	0.077	0.21	0.30	0.817	0.065	415
				T = 8				
TOBIT	0.86	0.86	0.063	0.71	0.79	0.875	0.143	614
CMLE-1	0.88	0.88	0.058	0.64	0.74	0.889	0.121	614
CMLE	0.88	0.88	0.058	0.64	0.75	0.890	0.122	614
BC1	0.99	0.99	0.062	0.11	0.17	0.846	0.044	614
BC2	0.94	0.94	0.057	0.24	0.34	0.904	0.062	614
BC3	0.96	0.96	0.058	0.15	0.23	0.886	0.049	614
				T = 12				
TOBIT	0.91	0.91	0.047	0.58	0.68	0.931	0.093	1016
CMLE-1	0.92	0.92	0.044	0.50	0.63	0.928	0.079	1016
CMLE	0.92	0.92	0.044	0.50	0.63	0.928	0.079	1016
BC1	0.99	0.99	0.044	0.07	0.14	0.924	0.032	1016
BC2	0.96	0.96	0.043	0.17	0.25	0.938	0.040	1016
BC3	0.97	0.97	0.043	0.12	0.19	0.933	0.035	1016

Notes: 1,000 replications. TOBIT denotes Tobit FE maximum likelihood estimator that does not account for endogeneity; CMLE-1 denotes unfeasible estimator that uses (unobserved) true control function; CMLE is the feasible version of CME-1 that uses estimated control function; BC1 denotes Hahn and Kuersteiner (2003) bias-corrected estimator based on Bartlett equalities; BC2 denotes Hahn and Kuersteiner (2003) bias-corrected estimator based on general estimating equations; BC3 denotes Fernandez-Val (2005) bias-corrected estimator; Standard errors account for generated regressors, when relevant.

Table 6: Tobit CMLE, ρ (control function), n =100

Estimator	Mean	Median	SD	p; .05	p; .10	SE/SD	MAE	nobs
				T = 6				
CMLE-1	0.98	0.98	0.158	0.09	0.16	0.882	0.106	415
CMLE	0.98	0.99	0.168	0.10	0.16	0.846	0.113	415
BC1	1.02	1.02	0.193	0.14	0.21	0.760	0.127	415
BC2	1.00	1.00	0.174	0.10	0.17	0.829	0.115	415
BC3	1.00	1.00	0.173	0.10	0.18	0.833	0.117	415
				T = 8				
CMLE-1	1.00	1.00	0.131	0.09	0.14	0.892	0.086	614
CMLE	1.00	1.00	0.140	0.11	0.15	0.849	0.092	614
BC1	1.04	1.03	0.152	0.12	0.20	0.800	0.099	614
BC2	1.01	1.01	0.144	0.11	0.17	0.833	0.094	614
BC3	1.01	1.01	0.142	0.09	0.15	0.845	0.093	614
				T = 12				
CMLE-1	1.00	1.00	0.102	0.07	0.12	0.912	0.068	1016
CMLE	1.00	1.00	0.107	0.08	0.14	0.884	0.072	1016
BC1	1.03	1.03	0.112	0.10	0.18	0.854	0.080	1016
BC2	1.00	1.01	0.108	0.08	0.14	0.881	0.074	1016
BC3	1.00	1.00	0.108	0.08	0.14	0.884	0.074	1016

Notes: 1,000 replications. CMLE-1 denotes unfeasible estimator that uses (unobserved) true control function; CMLE is the feasible version of CME-1 that uses estimated control function; BC1 denotes Hahn and Kuersteiner (2003) bias-corrected estimator based on Bartlett equalities; BC2 denotes Hahn and Kuersteiner (2003) bias-corrected estimator based on general estimating equations; BC3 denotes Fernandez-Val (2005) bias-corrected estimator; Standard errors account for generated regressors, when relevant.

Table 7: Descriptive Statistics, 1981-1988

		Full sample			Union		Nonunion	
Variable	Definition	Mean	St. Dev.	Within (%)	Mean	St. Dev.	Mean	St. Dev.
SCHOOL	Years of Schooling	12.33	1.69	5	12.21	1.15	12.37	1.84
LEXPER	Log(1 + EXPER)	1.78	0.60	47	1.90	0.49	1.74	0.63
UNION	Wage set by collective bargaining	0.26	0.44	41	1.00	0.00	0.00	0.00
UNION1	Lag of UNION	0.27	0.44	43	0.73	0.44	0.10	0.30
MARRIED	Married	0.45	0.50	41	0.52	0.50	0.42	0.49
BLACK	Black	0.11	0.32	0	0.15	0.36	0.10	0.30
HISP	Hispanic	0.15	0.36	0	0.14	0.35	0.15	0.36
HEALTH	Has health disability	0.02	0.15	76	0.02	0.15	0.02	0.15
RURAL	Lives in rural area	0.20	0.40	22	0.19	0.39	0.20	0.40
NE	Lives in North East	0.21	0.40	3	0.20	0.40	0.21	0.41
NC	Lives in Northern Central	0.29	0.46	3	0.36	0.48	0.27	0.45
S	Lives in South	0.31	0.46	4	0.26	0.44	0.32	0.47
W	Lives in West	0.19	0.39	3	0.19	0.39	0.19	0.40
WAGE	Log of real hourly wage	1.74	0.46	41	1.91	0.41	1.68	0.46
Industry dummies								
AG	Agricultural	0.03	0.17	49	0.01	0.11	0.03	0.18
MIN	Mining	0.02	0.13	44	0.03	0.16	0.01	0.12
CON	Construction	0.10	0.30	41	0.09	0.29	0.11	0.31
TRAD	Trade	0.25	0.43	47	0.17	0.38	0.28	0.45
TRA	Transportation	0.07	0.26	45	0.13	0.34	0.05	0.22
FIN	Finance	0.03	0.16	38	0.01	0.08	0.03	0.18
BUS	Business and Repair Service	0.08	0.27	59	0.04	0.20	0.09	0.28
PER	Personal Service	0.01	0.12	77	0.01	0.09	0.02	0.13
ENT	Entertainment	0.01	0.12	61	0.00	0.05	0.02	0.14
MAN	Manufacturing	0.32	0.47	38	0.39	0.49	0.29	0.45
PRO	Professional and Related Service	0.05	0.21	56	0.04	0.20	0.05	0.21
PUB	Public Administration	0.04	0.18	49	0.08	0.26	0.02	0.14
Occupational dummies								
OCC1	Professional, Technical and kindred	0.07	0.25	53	0.03	0.17	0.08	0.27
OCC2	Managers, Officials and Proprietors	0.09	0.28	61	0.03	0.17	0.11	0.31
OCC3	Sales Workers	0.01	0.07	63	0.00	0.03	0.01	0.08
OCC4	Clerical and kindred	0.10	0.30	63	0.10	0.30	0.10	0.30
OCC5	Craftsmen, Foremen and kindred	0.23	0.42	50	0.23	0.42	0.24	0.42
OCC6	Operatives and kindred	0.23	0.42	54	0.32	0.47	0.21	0.40
OCC7	Laborers and farmers	0.11	0.31	69	0.16	0.36	0.09	0.29
OCC8	Farm Laborers and Foreman	0.01	0.11	55	0.00	0.04	0.02	0.13
OCC9	Service Workers	0.11	0.31	50	0.13	0.33	0.10	0.30
Number of Observations		4360		545 × 8	1141		3219	

Source : NLSY Men

Table 8: Fixed effects probit estimates of union membership (1981-1988)

	Index Coefficient			Average Marginal Effect		Index Coefficient			Average Marginal Effect	
	POOLED	FE	BC	FE	BC	POOLED	FE	BC	FE	BC
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
	Without Occupation					With Occupation				
UNION1	1.80	0.35	0.73	0.05	0.10	1.76	0.32	0.70	0.04	0.09
	(0.07)	(0.07)	(0.07)	(0.01)	(0.01)	(0.07)	(0.07)	(0.07)	(0.01)	(0.01)
SCHOOL	-0.01	0.09	0.07	0.01	0.01	0.04	0.19	0.16	0.02	0.02
	(0.02)	(0.11)	(0.11)	(0.01)	(0.01)	(0.02)	(0.11)	(0.11)	(0.01)	(0.01)
LEXPER	0.16	0.95	0.64	0.12	0.09	0.19	0.98	0.65	0.12	0.09
	(0.07)	(0.25)	(0.23)	(0.03)	(0.03)	(0.07)	(0.26)	(0.24)	(0.03)	(0.03)
RURAL	-0.09	0.15	0.13	0.02	0.02	-0.12	0.16	0.14	0.02	0.02
	(0.07)	(0.20)	(0.18)	(0.03)	(0.02)	(0.07)	(0.20)	(0.18)	(0.02)	(0.02)
MARRIED	0.14	0.26	0.21	0.03	0.03	0.15	0.26	0.21	0.03	0.03
	(0.06)	(0.11)	(0.11)	(0.01)	(0.01)	(0.06)	(0.11)	(0.11)	(0.01)	(0.01)
HEALTH	-0.18	-0.31	-0.28	-0.04	-0.04	-0.19	-0.32	-0.29	-0.04	-0.04
	(0.16)	(0.28)	(0.30)	(0.04)	(0.04)	(0.17)	(0.29)	(0.31)	(0.04)	(0.04)
BLACK	0.24					0.20				
	(0.08)					(0.08)				
HISP	0.02					0.01				
	(0.09)					(0.09)				
Observations	4360	2064	2064	4360	4360	4360	2064	2064	4360	4360

Notes : Standard errors in parantheses. All regressions include industry, region, and time dummies. Standard errors in columns [1] and [6] are clustered at the individual level.

Table 9: Wage Regressions with union effects (1981-1988)

	P - OLS	P - Heckit	FE	FE - Heck	BC - Heck	P - OLS	P - Heckit	FE	FE - Heck	BC - Heck
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
	Without Occupation					With Occupation				
UNION	0.16	0.24	0.10	0.30	0.40	0.18	0.28	0.11	0.32	0.42
	(0.02)	(0.04)	(0.02)	(0.06)	(0.05)	(0.02)	(0.04)	(0.02)	(0.05)	(0.05)
SCHOOL	0.08	0.08	0.09	0.09	0.09	0.07	0.07	0.08	0.07	0.07
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
LEXPER	0.21	0.20	0.29	0.27	0.26	0.20	0.19	0.28	0.25	0.24
	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)
RURAL	-0.16	-0.16	-0.01	-0.01	-0.02	-0.15	-0.15	-0.01	-0.01	-0.02
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
MARRIED	0.11	0.11	0.03	0.02	0.02	0.10	0.10	0.03	0.02	0.02
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
HEALTH	0.02	0.03	0.03	0.04	0.04	0.03	0.04	0.04	0.04	0.05
	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)
BLACK	-0.19	-0.20				-0.17	-0.18			
	(0.04)	(0.04)				(0.04)	(0.04)			
HISP	-0.08	-0.08				-0.07	-0.07			
	(0.04)	(0.04)				(0.03)	(0.03)			
λ		-0.07		-0.12	-0.18		-0.08		-0.13	-0.19
		(0.02)		(0.03)	(0.03)		(0.02)		(0.03)	(0.03)
R-squared	0.36	0.36	0.71	0.71	0.71	0.38	0.38	0.71	0.71	0.71
Obs.	4360	4360	4360	4360	4360	4360	4360	4360	4360	4360

Notes : Standard errors in parantheses. All regressions include industry, region, and time dummies. Standard errors in columns [1], [2], [6] and [7] are clustered at the individual level. Standard errors in columns [3], [4], [5], [8], [9], and [10] are robust to heteroskedasticity. Standard errors in columns [4], [5], [9], and [10] account for generated regressors (the standard errors in columns [2] and [7] do not account for the estimation of the mills ratio).