TESTING AND CORRECTING FOR ENDOGENEITY IN NONLINEAR UNOBSERVED EFFECTS MODELS

Wei Lin
Institute for Economics and Social Research
Jinan University

Jeffrey M. Wooldridge
Department of Economics
Michigan State University

This version: September 10, 2018

Abstract

We study testing and estimation in panel data models with two potential sources of endogeneity: that due to correlation of covariates with time-constant, unobserved heterogeneity and that due to correlation of covariates with time-varying idiosyncratic errors. In the linear case, we show that two control function approaches allow us to test exogeneity with respect to the idiosyncratic errors while being silent on exogeneity with respect to heterogeneity. The linear case suggests a general approach for nonlinear models. We consider two leading cases of nonlinear models: an exponential conditional mean function for nonnegative responses and a probit conditional mean function for binary or fractional responses. In the former case, we exploit the full robustness of the fixed effects Poisson quasi-MLE, and for the probit case we propose correlated random effects

1. Introduction

The availability of panel data can greatly facilitate the estimation of causal effects from nonexperimental data. For example, for studying policy interventions using linear models, the methods of fixed effects (FE) estimation and first differencing (FD) estimation are used routinely. The primary attractiveness of the FE and FD methods is due to their eliminating additive, unobserved heterogeneity that is thought to be correlated with the policy variable or variables of interest. Fixed effects-type approaches are available in special cases for nonlinear models, although in such cases they are best viewed as conditional maximum likelihood, or conditional quasi-maximum likelihood, estimators, where a conditioning argument essentially removes the dependence of an objective function on unobserved heterogeneity. The leading cases are the so-called FE logit and FE Poisson estimators. To handle heterogeneity more generally in a microeconometric setting, where the number of avalaible time periods, T, is typically small, the correlated random effects (CRE) approach can be effective. Wooldridge (2010) shows how the CRE approach can be used for a variety of nonlinear panel data models used in practice. See also Wooldridge (2016) for some recent developments using unbalanced panels.

One drawback to FE, FD, and CRE approaches is that they allow for only one kind of endogeneity: correlation between the time-varying explanatory variables, often through sometime like the time average of these variables, and time-constant heterogeneity. But in many contexts we may be worried about correlation between at least some of the covariates and unobserved shocks – often called idiosyncratic errors. In the case of a linear model, combining instrumental variables (IV) approaches with the FE and FD transformations can be quite powerful. For example, Levitt (1996, 1997) uses IV approaches after eliminating

heterogeneity at either the state or city level.

Fixed effects IV approaches explicitly recognize two potential sources of endogeneity. We will call these "heterogeneity endogeneity," which arises when one or more explanatory variables is correlated with time-constant heterogeneity, and "idiosyncratic endogeneity," which arises when one or more explanatory variables is correlated with time-varying unobservables. Both kinds of endogeneity can be present in nonlinear models, too. Papke and Wooldridge (2008) [hereafter, PW (2008)], in the context of a probit fractional response model, show how to combine the CRE and control function approaches to allow for heterogeneity and idiosyncratic endogeneity. [More recently, Murtazashvili and Wooldridge (2016) use a similar approach for panel data switching regression models with lots of heterogeneity.] The approach is largely parametric, although it is robust to distributional misspecificiation other than the conditional mean, and it allows unrestricted serial dependence over time – a feature not allowed, for example, by random effects probit or fixed effects logit approaches. The PW (2008) approach is attractive because it leads to simple estimation methods, robust inference, and easy calculation of average partial effects. It does, however, have a couple of potential drawbacks. The first is that the method does not allow one to tell whether a rejection of the null hypothesis of exogeneity of the covariates is due to heterogeneity or idiosyncratic endogeneity. Second, the explanatory variables that are potentially endogenous in the structural equation are not rendered *strictly* exogenous in the estimating equation. Rather, they are only contemporaneously exogenous, which means that only pooled methods, or method of moments versions of them, produce consistent estimators. This leaves out the possibility of applying quasi-generalized least squares approaches, such as the generalized estimating equations (GEE) approach that is popular in fields outside

economics.

In this paper, we show how to modify, in a straightforward way, the CRE/CF approach of PW (2008) so that we can easily separate the two kinds of endogeneity. One benefit is that we can test the null hypothesis of idiosyncratic exogeneity while allowing for heterogeneity exogeneity, which effectively allows us to determine whether an IV approach is needed. Section 2 covers the linear case, where we show that our new control function approach leads to a test statistic that is identical to the variable addition Hausman test discussed in Wooldridge (2010, Chapter 11). This sets the stage for two leading cases of nonlinear models, an exponential mean function and a probit mean function. The exponential mean case, treated in Section 3, is an interesting case because the robustness properties of the Poisson FE estimator can be combined with the control function approach to obtain a test for idiosyncratic exogeneity that is fully robust to distributional misspecification, as well as to serial dependence of arbitrary form. We also cover the issue of estimating average partial effects, and discuss the merits of a CRE/CF approach. In section 4 we turn to a probit response function – as in PW (2008) – and show how to modify PW's CRE approach to separately analyze the two kinds of endogeneity. Section 5 discusses how the approach applied to general nonlinear unobserved effects models, and provides a discussion of the pros and cons of using a joint MLE – such as random effects probit or random effects Tobit – in the second stage. Two empirical applications in section 6 show how the methods are easily applied, and section 7 contains concluding remarks.

2. Models Linear in Parameters

We start with a "structural" equation

$$y_{it1} = \mathbf{x}_{it1}\boldsymbol{\beta}_1 + c_{i1} + u_{it1} \tag{2.1}$$

where, for now, the explanatory variables are

$$\mathbf{x}_{it1} = (\mathbf{y}_{it2}, \mathbf{z}_{it1}).$$

The vector $\mathbf{z}_{it1}$ would typically include a full set of time effects to allow for secular changes over time. We suspect the vector $\mathbf{y}_{it2}$ is endogenous in that it may be correlated with the unobserved effect (or heterogeneity), c_{i1}, and possibly with the idiosyncratic error, u_{it1}. In what follows we allow all exogenous variables, which include the vector $\mathbf{z}_{it1}$ and variables excluded, $\mathbf{z}_{it2}$, to be correlated with the heterogeneity. Therefore, we proceed as if all explanatory variables can be correlated with the unobserved heterogeneity, c_{i1}. In other words, we are not taking a traditional random effects approach.

The difference between $\mathbf{y}_{it2}$ and $\mathbf{z}_{it}$ is that we take the latter to be strictly exogenous with respect to $\{u_{it1}\}$:

$$Cov(\mathbf{z}_{it}, u_{ir1}) = 0, \text{ all } t, r = 1, \ldots, T.$$

By contrast, $\{\mathbf{y}_{it2}\}$ may be correlated with $\{u_{it1}\}$, either contemporaneously or across time periods.

Given a suitable rank condition, which is discussed in Wooldridge (2010, Chapter 11), $\boldsymbol{\beta}_1$ can be estimated by fixed effects 2SLS (FE2SLS), sometimes called FEIV. To describe the estimator, define the deviations from time averages as

$$\ddot{y}_{it1} = y_{it1} - T^{-1}\sum_{r=1}^{T} y_{ir1}, \ddot{\mathbf{y}}_{it2} = \mathbf{y}_{it2} - T^{-1}\sum_{r=1}^{T} \mathbf{y}_{ir2}, \ddot{\mathbf{z}}_{it} = \mathbf{z}_{it} - T^{-1}\sum_{r=1}^{T} \mathbf{z}_{ir}.$$

Given a random sample (in the cross section) of size N, one characterization of FE2SLS estimator is that it is pooled 2SLS applied to the equation

$$\ddot{y}_{it1} = \ddot{\mathbf{x}}_{it1}\boldsymbol{\beta}_1 + \ddot{u}_{it1}, t = 1, \ldots, T$$

using IVs $\ddot{\mathbf{z}}_{it}$. With fixed T and $N \to \infty$, the estimator is generally consistent and $\sqrt{N}$-asymptotically normal. Fully robust inference that allows arbitrary serial correlation and heteroskedasticity in $\{u_{it1}\}$ is straightforward.

In terms of precision, the FE2SLS estimator can have large standard errors. We are first removing much of the variation in the data by removing the time averages, and then we are applying 2SLS. At a minimum, we require sufficient variation in the excluded exogenous variables that serve as instruments for $\mathbf{y}_{it2}$. Therefore, it is of some interest to test the null hypothesis that $\{\mathbf{y}_{it2}\}$ is exogenous with respect to $\{u_{it1}\}$.

A common approach is to apply the Hausman (1978) principle, where the two estimators being compared are the usual FE estimator and the FE2SLS estimator. The usual FE estimator is consistent if we add the assumption

$$Cov(\mathbf{y}_{it2}, u_{ir1}) = 0, \text{ all } t, r = 1, \ldots, T.$$

The FE2SLS estimator does not require this stronger form of exogeneity of $\mathbf{y}_{it2}$.

There are a couple of drawbacks to the traditional Hausman test. Most importantly, because it assumes that one estimator is relatively efficient – in this case, the FE estimator plays the role of the efficient estimator – it is not robust to serial correlation or heteroskedasticity in $\{u_{it1}\}$. If we make our inference concerning $\boldsymbol{\beta}_1$ robust to departures from the standard, usually unrealistic, assumptions, then it is logically inconsistent to use nonrobust specification tests. Wooldridge (1990) makes this point in the context of a variety of specification tests. The second problem with the traditional Hausman test is the asymptotic variance required is singular, and this can lead to computational problems as well as incorrect calculation of

degrees of freedom.

A simpler approach is to obtain a variation addition test (VAT), which is based on the control function approach. Wooldridge (2010, Chapter 11) describes the procedure:

Procedure 2.1 (FE Variable Addition Test):

1. Estimate the reduced form of $\mathbf{y}_{it2}$,

$$\mathbf{y}_{it2} = \mathbf{z}_{it}\boldsymbol{\Pi}_2 + \mathbf{c}_{i2} + \mathbf{u}_{it2},$$

by fixed effects, and obtain the FE residuals,

$$\ddot{\hat{\mathbf{u}}}_{it2} = \ddot{\mathbf{y}}_{it2} - \ddot{\mathbf{z}}_{it}\hat{\boldsymbol{\Pi}}_2$$
$$\ddot{\mathbf{y}}_{it2} = \mathbf{y}_{it2} - T^{-1}\sum_{r=1}^{T}\mathbf{y}_{ir2}$$

2. Estimate the equation

$$y_{it1} = \mathbf{x}_{it1}\boldsymbol{\beta}_1 + \ddot{\hat{\mathbf{u}}}_{it2}\boldsymbol{\rho}_1 + c_{i1} + error_{it1}$$

by *usual* FE and compute a robust Wald test of $H_0 : \boldsymbol{\rho}_1 = \mathbf{0}$. □

The VAT version of the Hausman test has a simple interepretation, because the $\hat{\boldsymbol{\beta}}_1$ obtained in the second step is actually the FEIV estimate. If we set $\boldsymbol{\rho}_1$ to zero we are using the usual FE estimator. If we estimate $\boldsymbol{\rho}_1$, we obtain the FEIV estimator. Importantly, it is very easy to make the test robust to arbitrary serial correlation and heteroskedasticity. As a practical matter, it is important to understand that the nature of $\mathbf{y}_{it2}$ is unrestricted. It can be continuous, discrete (including binary), or some mixture. Below we will discuss what happens if we allow more general functional forms.

In motivating our general approach for nonlinear models, it is useful to obtain a test based on Mundlak's (1978) CRE approach. We must use some care to obtain a test that rejects only

in the presence of idiosyncratic endogeneity. We start with a linear reduced form for $\mathbf{y}_{it2}$, but we emphasize that, for linear models, this equation is not restrictive. A linear unobserved effects reduced form is

$$\mathbf{y}_{it2} = \mathbf{z}_{it}\boldsymbol{\Pi}_2 + \mathbf{c}_{i2} + \mathbf{u}_{it2}$$

where $\boldsymbol{\Pi}_2$ is dimension $L \times G_1$ where G_1 is the dimension of $\mathbf{y}_{it2}$. Now we apply the Mundlak (1978) to the vector of unobserved heterogeneity, $\mathbf{c}_{i2}$:

$$\mathbf{c}_{i2} = \boldsymbol{\psi}_2 + \bar{\mathbf{z}}_i\boldsymbol{\Xi}_2 + \mathbf{a}_{i2},$$

where $\bar{\mathbf{z}}_i = T^{-1}\sum_{t=1}^{T}\mathbf{z}_{it}$ is the row vector of time averages of all exogenous variables and $\boldsymbol{\Xi}_2$ is $L \times G_1$. Plugging into the previous equation gives

$$\mathbf{y}_{it2} = \boldsymbol{\psi}_2 + \mathbf{z}_{it}\boldsymbol{\Pi}_2 + \bar{\mathbf{z}}_i\boldsymbol{\Xi}_2 + \mathbf{a}_{i2} + \mathbf{u}_{it2}, \; t = 1, \ldots, T.$$

In what follows, we operate *as if*

$$Cov(\mathbf{z}_{it}, \mathbf{u}_{is2}) = \mathbf{0}, \text{ all } t, s$$
$$Cov(\mathbf{z}_{it}, \mathbf{a}_{i2}) = \mathbf{0}, \text{ all } t,$$

but, as we will see, even these mild assumptions need not actually hold.

The key now in obtaining a test of idiosyncratic endogeneity is how we apply the Mundlak device to c_{i1} in the structural equation

$$y_{it1} = \mathbf{x}_{it1}\boldsymbol{\beta}_1 + c_{i1} + u_{it1}$$

One possibility is to project c_{i1} only onto $\bar{\mathbf{z}}_i$. It turns out that this approach is fine for estimating $\boldsymbol{\beta}_1$ but, for testing endogeneity of $\mathbf{y}_{it2}$, it does not distinguish between

$$Cov(\mathbf{y}_{it2}, c_{i1}) \neq \mathbf{0}$$

and

$$Cov(\mathbf{y}_{it2}, u_{is1}) \neq \mathbf{0}.$$

Instead, it is better to project c_{i1} onto $(\bar{\mathbf{z}}_i, \bar{\mathbf{v}}_{i2})$ where

$$\mathbf{v}_{it2} = \mathbf{a}_{i2} + \mathbf{u}_{it2}.$$

Then we have

$$\begin{aligned} c_{i1} &= \eta_1 + \bar{\mathbf{z}}_i\boldsymbol{\lambda}_1 + \bar{\mathbf{v}}_{i2}\boldsymbol{\pi}_1 + a_{i1} \\ Cov(\mathbf{z}_i, a_{i1}) &= \mathbf{0} \\ Cov(\mathbf{y}_{i2}, a_{i1}) &= \mathbf{0} \end{aligned}$$

Importantly, the remaining heterogeneity, a_{i1}, us uncorrelated not only with $\mathbf{z}_i = \{\mathbf{z}_{it} : t = 1, \ldots, T\}$ but also with $\mathbf{y}_{i2} = \{\mathbf{y}_{it2} : t = 1, \ldots, \}$. Plugging into the structure equation produces the following estimating equation:

$$\begin{aligned} y_{it1} &= \mathbf{x}_{it1}\boldsymbol{\beta}_1 + \eta_1 + \bar{\mathbf{z}}_i\boldsymbol{\lambda}_1 + \bar{\mathbf{v}}_{i2}\boldsymbol{\pi}_1 + a_{i1} + u_{it1} \\ &= \mathbf{x}_{it1}\boldsymbol{\beta}_1 + \eta_1 + \bar{\mathbf{z}}_i\boldsymbol{\lambda}_1 + (\bar{\mathbf{y}}_{i2} - \boldsymbol{\psi}_2 - \bar{\mathbf{z}}_i\boldsymbol{\Lambda}_2)\boldsymbol{\pi}_1 + a_{i1} + u_{it1} \\ &\equiv \mathbf{x}_{it1}\boldsymbol{\beta}_1 + \psi_1 + \bar{\mathbf{y}}_{i2}\boldsymbol{\pi}_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + a_{i1} + u_{it1}. \end{aligned}$$

Now, by the Mundlak device, a_{i1} is uncorrelated with *all* RHS observables, that is, $(\mathbf{y}_{it2}, \mathbf{z}_{it1}, \bar{\mathbf{y}}_{i2}, \bar{\mathbf{z}}_i)$. By the strict exogeneity assumption on $\{\mathbf{z}_{it} : t = 1, \ldots, T\}$, u_{it1} is uncorrelated with $(\mathbf{z}_{it1}, \bar{\mathbf{z}}_i)$. Therefore, we can now test whether $\mathbf{y}_{it2}$, is uncorrelated with u_{it1} by testing whether $\mathbf{v}_{it2}$ is uncorrelated with u_{it1}.

Procedure 2.2 (CRE/CF Variable Addition Test):

1. Run a pooled OLS regression

$$\mathbf{y}_{it2} = \boldsymbol{\psi}_2 + \mathbf{z}_{it}\boldsymbol{\Pi}_2 + \bar{\mathbf{z}}_i\boldsymbol{\Xi}_2 + \mathbf{v}_{it2},$$

and obtain the residuals, $\hat{\mathbf{v}}_{it2}$.

2. Estimate

$$y_{it1} = \mathbf{x}_{it1}\boldsymbol{\beta}_1 + \psi_1 + \bar{\mathbf{y}}_{i2}\boldsymbol{\pi}_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + \hat{\mathbf{v}}_{it2}\boldsymbol{\rho}_1 + error_{it1} \tag{2.2}$$

by POLS or RE and use a robust Wald test of $H_0 : \boldsymbol{\rho}_1 = \mathbf{0}$. □

Because the derivation of the estimating equation in Procedure 2.2 uses the Mundlak device, it nominally appears that it is less robust than that based on fixed effects in Procedure 2.1. This turns out not to be the case; in fact, the two approaches yield identical estimates of $\boldsymbol{\beta}_1$ and $\boldsymbol{\rho}_1$. The estimate of $\boldsymbol{\beta}_1$ is still the FEIV estimate. Therefore, we can use either the FE approach or the Mundlak CRE approach, and it does not matter whether the residuals we add to the equation are the FE residuals, $\ddot{\hat{\mathbf{u}}}_{it2}$, or the Mundlak residuals, $\hat{\mathbf{v}}_{it2}$. These residuals are not the same, but in the appendix it is shown that

$$\hat{\mathbf{v}}_{it2} = \ddot{\hat{\mathbf{u}}}_{it2} + \hat{\mathbf{r}}_{i2}$$

where

$$\hat{\mathbf{r}}_{i2} = \bar{\mathbf{y}}_{i2} - \hat{\boldsymbol{\kappa}}_2 - \bar{\mathbf{z}}_i\hat{\boldsymbol{\Lambda}}_2$$

are the between residuals from regressing $\bar{\mathbf{y}}_{i2}$ on 1, $\bar{\mathbf{z}}_i$. In particular, $\hat{\mathbf{r}}_{i2}$ is a linear combination of $(\bar{\mathbf{y}}_{i2}, 1, \bar{\mathbf{z}}_i)$. It follows immediately that replacing $\hat{\mathbf{v}}_{it2}$ in (2.2) does not change $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\rho}}_1$. Only $\hat{\psi}_1$, $\hat{\boldsymbol{\pi}}_1$ and $\hat{\boldsymbol{\xi}}_1$ would change.

Interestingly, if we drop $\bar{\mathbf{y}}_{i2}$ from step (2) in Procedure 2.2, the resulting estimate of $\boldsymbol{\beta}_1$ is still the FEIV estimate. But we obtain a different estimate of $\boldsymbol{\rho}_1$, and basing a test of endogeneity on the equation without including $\bar{\mathbf{y}}_{i2}$ conflates heterogeneity endogeneity and idiosyncratic endogeneity. Evidently, this point has gone unnoticed, probably because Procedure 2.1 is the usual VAT in testing for idiosyncratic endogeneity. Neverthless, this observation is very important when we must use the Mundlak CRE approach in nonlinear models (because an FE approach is not available).

The conclusion from this section is that, for using the CRE/CF approach for testing

$$H_0 : Cov(\mathbf{y}_{it2}, u_{is1}) = \mathbf{0},$$

we should use the equations

$$\begin{aligned}\mathbf{y}_{it2} &= \hat{\boldsymbol{\psi}}_2 + \mathbf{z}_{it}\hat{\boldsymbol{\Pi}}_2 + \bar{\mathbf{z}}_i\hat{\boldsymbol{\Xi}}_2 + \hat{\mathbf{v}}_{it2} \\ y_{it1} &= \mathbf{x}_{it1}\boldsymbol{\beta}_1 + \psi_1 + \bar{\mathbf{y}}_{i2}\boldsymbol{\pi}_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + \hat{\mathbf{v}}_{it2}\boldsymbol{\rho}_1 + error_{it1},\end{aligned}$$

being sure to include $\bar{\mathbf{y}}_{i2}$.

As an aside, one might want to know what happens if the seemingly less restrictive Chamberlain (1982) version of the CRE approach is used in place of Mundlak. The answer is: nothing. At least not if we use the basic estimation methods that do not attempt to exploit serial correlation or heteroskedasticity in the $\{u_{it1}\}$. To be clear, letting

$$\mathbf{z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{iT}), \mathbf{y}_{i2} = (\mathbf{y}_{i12}, \ldots, \mathbf{y}_{iT2}),$$

the equations

$$\begin{aligned}\mathbf{y}_{it2} &= \hat{\boldsymbol{\psi}}_2 + \mathbf{z}_{it}\hat{\boldsymbol{\Pi}}_2 + \mathbf{z}_i\hat{\boldsymbol{\Xi}}_2 + \hat{\mathbf{v}}_{it2} \\ y_{it1} &= \mathbf{x}_{it1}\boldsymbol{\beta}_1 + \psi_1 + \mathbf{z}_i\boldsymbol{\xi}_1 + \mathbf{y}_{i2}\boldsymbol{\pi}_1 + \hat{\mathbf{v}}_{it2}\boldsymbol{\rho}_1 + error_{it1}\end{aligned}$$

result in the same estimates of $\boldsymbol{\beta}_1$ and $\boldsymbol{\rho}_1$ as the Mundlak approach, provided we use either pooled OLS or RE in the second equation.

How can one use the test of idiosyncratic endogeneity? Guggenberger (2010) shows that the pretesting problem that exists from using the Hausman test to determine an appropriate estimation strategy can be severe. Nevertheless, such practice is common in empirical work. If the VAT rejects at, say, the 5% significance level, one typically uses the FEIV estimator. If one fails to reject, it provides some justification for dropping the IV approach and instead using the usual FE estimator.

3. Exponential Model

If y_{it1} is nonnegative, and especially if it can take the value zero, an exponential conditional mean function is attractive. (The common alternative when $y_{it1} > 0$ is to use $\log(y_{it1})$ in a linear model, but some researchers prefer to model y_{it1} directly.) An unobserved effects model that allows for heterogeneity endogeneity and idiosyncratic endogeneity is

$$E(y_{it1}|\mathbf{y}_{i2}, \mathbf{z}_i, c_{i1}, r_{it1}) = E(y_{it1}|\mathbf{y}_{it2}, \mathbf{z}_{it1}, c_{i1}, r_{it1}) = c_{i1} \exp(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + r_{it1}), \tag{3.1}$$

where, again, $\mathbf{x}_{it1} = (\mathbf{y}_{it2}, \mathbf{z}_{it1})$. Now the heterogeneity, c_{i1}, is nonnegative and multiplicative. We use r_{it1} to denote time-varying omitted factors that we suspect are correlated with $\mathbf{y}_{it2}$. We could make r_{it1} multiplicative but it is slightly more convenient to have it appear inside the exponential function.

An FE Poisson/CF Approach

As discussed in Wooldridge (1999) and Wooldridge (2010, Chapter 18), without r_{it1} an appealing estimator is what is typically called the fixed effects Poisson estimator. In Hausman, Hall, and Griliches (1984), the FE Poisson estimator was obtained as a conditional MLE, where the Poisson assumption was assumed to hold along with conditional independence. Wooldridge (1999) showed that the neither assumption is needed to ensure consistency and asymptotic normality of the FE Poisson estimator. Viewed as a quasi-MLE, the estimator is fully robust in the sense that it only requires, in the current notation (with idiosyncratic endogeneity),

$$E(y_{it1}|\mathbf{x}_{i1}, c_{i1}) = E(y_{it1}|\mathbf{x}_{it1}, c_{i1}) = c_{i1} \exp(\mathbf{x}_{it1}\boldsymbol{\beta}_1).$$

The first equality imposes a strict exogeneity requirement with respect to idiosyncratic shocks. It will be violated if r_{it1} is present and correlated with $\mathbf{y}_{is2}$ for any time period s, including, of

course, $s = t$.

To obtain a test of the null hypothesis that there is no idiosyncratic endogeneity, we again need time-varying, strictly exogenous instruments that are excluded from $\mathbf{z}_{it1}$. Formally, the null hypothesis is

$$E(y_{it1}|\mathbf{y}_{i2}, \mathbf{z}_i, c_{i1}) = E(y_{it1}|\mathbf{y}_{it2}, \mathbf{z}_{it1}, c_{i1}) = c_{i1} \exp(\mathbf{x}_{it1}\boldsymbol{\beta}_1),$$

where the key is that $\mathbf{z}_{it2}$ is exclused from the mean function. Also, all variables are strictly exogenous conditional on c_{i1}. In order to obtain a test, we need to specify an alternative, and this is where explicitly introducing a time-varying unobservables into the structural model, and a reduced form for $\mathbf{y}_{it2}$, come into play. But we emphasize that these do not play a role under the null hypothesis. They are used only to obtain a test. In addition to (3.1), we write

$$\mathbf{y}_{it2} = \mathbf{z}_{it}\boldsymbol{\Pi}_2 + \mathbf{c}_{i2} + \mathbf{u}_{it2}, \, t = 1, \ldots, T,$$

and, because the $\{\mathbf{z}_{it}\}$ is strictly exogenous, we test for correlation between $\{r_{it1}\}$ and functions of $\{\mathbf{u}_{it2}\}$. We use the analog of the test from Procedure 2.1.

Procedure 3.1 (Poisson FE/VAT):

1. Estimate the reduced form for $\mathbf{y}_{it2}$ by fixed effects and obtain the FE residuals,

$$\ddot{\hat{\mathbf{u}}}_{it2} = \ddot{\mathbf{y}}_{it2} - \ddot{\mathbf{z}}_{it}\hat{\boldsymbol{\Pi}}_2$$

2. Use FE Poisson on the mean function

$$\text{“}E(y_{it1}|\mathbf{z}_{it1}, \mathbf{y}_{it2}, \hat{\ddot{\mathbf{u}}}_{it2}, c_{i1}) = c_{i1} \exp(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + \hat{\ddot{\mathbf{u}}}_{it2}\boldsymbol{\rho}_1)\text{”}$$

and use a robust Wald test of $H_0 : \boldsymbol{\rho}_1 = \mathbf{0}$. □

It turns out that, as in the linear case, the fixed effects residuals can be replaced with the Mundlak residuals. Again let $\hat{\mathbf{v}}_{it2}$ be the OLS residuals from estimating

$$\mathbf{y}_{it2} = \boldsymbol{\psi}_2 + \mathbf{z}_{it}\boldsymbol{\Pi}_2 + \bar{\mathbf{z}}_i\boldsymbol{\Xi}_2 + \mathbf{v}_{it2}.$$

Then, as shown in the appendix, step (2) in Procedure 3.1 produces the same estimates of $(\boldsymbol{\beta}_1, \boldsymbol{\rho}_1)$. This follows from the form of the FE Poisson quasi-log-likelihood function and the fact that $\hat{\mathbf{v}}_{it2} = \ddot{\hat{\mathbf{u}}}_{it2} + \hat{\mathbf{r}}_{i2}$, and so removing the time averages of $\hat{\mathbf{v}}_{it2}$ produces the FE residuals $\ddot{\hat{\mathbf{u}}}_{it2}$.

As in the linear case, it is useful to remember that, under the null hypothesis, no restrictions are placed on $\mathbf{y}_{it2}$. In fact, the EEVs could include binary variables, in which case the reduced forms are linear probability models estimated by FE or the CRE approach. Under the null that $\{\mathbf{y}_{it2}\}$ is exogenous we can use any way of generating residuals that we want. More power might be obtained by using different models for the elements of $\mathbf{y}_{it2}$, but that is a power issue.

The equivalance between the between using the FE residuals $\ddot{\hat{\mathbf{u}}}_{it2}$ and the Mundlak residuals $\hat{\mathbf{v}}_{it2}$ means that we can obtain sufficient conditions for Procedure 3.1 to correct for idiosyncratic endogeneity when it is present. But now we need to make assumptions on the reduced form of $\mathbf{y}_{it2}$. We can get by with somewhat less, but a convenient assumption is

$$(\mathbf{r}_{i1}, \mathbf{u}_{i2}) \text{ is independent of } (c_{i1}, \mathbf{c}_{i2}, \mathbf{z}_i),$$

where $\mathbf{r}_{i1}$ is the vector of omitted variables in (3.1) and $\mathbf{u}_{i2}$ is the reduced form error. This assumption that v_{it2} is independent of means that the Mundlak equation is in fact a conditional expectation. Moreover, there cannot be heteroskedasticity.

Now, if we make a functional form assumption,

$$E[\exp(r_{it1})|\mathbf{u}_{i2}] = \exp(\theta_1 + \mathbf{u}_{it2}\boldsymbol{\rho}_1) = \exp[\theta_1 + (\mathbf{v}_{it2} - \mathbf{a}_{i1})\boldsymbol{\rho}_1],$$

which follows under joint normality of $(\mathbf{r}_{i1}, \mathbf{u}_{i2})$ but can hold more generally. The structural expectation is in (3.1), where now we also assume this is the expectation when we add $\mathbf{c}_{i2}$ to

the conditioning set. Then

$$E(y_{it1}|\mathbf{y}_{i2},\mathbf{z}_i,c_{i1},\mathbf{c}_{i2},\mathbf{v}_{i2}) = c_{i1}\exp[\mathbf{x}_{it1}\boldsymbol{\beta}_1 + \theta_1 + (\mathbf{v}_{it2} - \mathbf{a}_{i1})\boldsymbol{\rho}_1]$$
$$= g_{i1}\exp(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + \mathbf{v}_{it2}\boldsymbol{\rho}_1)$$

where $g_{i1} = c_{i1}\exp(-\mathbf{a}_{i1}\boldsymbol{\rho}_1)$. Now we can use Procedure 3.1, with either the FE residuals or the Mundlak residuals, to consistently estimate $\boldsymbol{\beta}_1$, along with $\boldsymbol{\rho}_1$, using the Poisson FE estimator. We require nothing more about the Poisson distribution to be correctly specified, and serial independence is entirely unrestricted. However, because we now allow $\boldsymbol{\rho}_1 \neq \mathbf{0}$, the standard errors need to be adjusted for the two-step estimation. One can use the delta method, or use a panel bootstrap, where both estimating steps are done with each bootstrap sample.

Estimating Average Partial Effects

In addition to consistently estimating $\boldsymbol{\beta}_1$, we may want to obtain partial effects on the conditional expectation itself. One possibility is to estimate the average structural function (Blundell and Powell, 2004), which averages out the unobservables for fixed $\mathbf{x}_{t1}$:

$$ASF_t(\mathbf{x}_{t1}) = E_{(c_{i1},r_{it1})}[c_{i1}\exp(\mathbf{x}_{t1}\boldsymbol{\beta}_1 + r_{it1})]$$
$$= E_{(c_{i1},r_{it1})}[c_{i1}\exp(r_{it1})]\exp(\mathbf{x}_{t1}\boldsymbol{\beta}_1).$$

Let

$$v_{it1} = c_{i1}\exp(r_{it1})$$
$$\theta_{t1} \equiv E(v_{it1}).$$

Because we have a consistent estimate of $\boldsymbol{\beta}_1$ – which would typically include time effects – we just need to estimate θ_{t1} for each t (or, we might assume these are constant across t). Write

$$y_{it1} = v_{it1}\exp(\mathbf{x}_{it1}\boldsymbol{\beta}_1)e_{it1}$$
$$E(e_{it1}|\mathbf{x}_{i1},c_{i1},\mathbf{r}_{i1}) = 1.$$

In particular,

$$E(v_{it1}e_{it1}) = E[v_{it1}E(e_{it1}|v_{it1})] = E(v_{it1}) = \theta_{t1}.$$

Therefore,

$$\theta_{t1} = E\left[\frac{y_{it}}{\exp(\mathbf{x}_{it1}\boldsymbol{\beta}_1)} \right]$$

and so a consistent estimator of θ_{t1} is

$$\hat{\theta}_{t1} = N^{-1}\sum_{i=1}^{N}\left[\frac{y_{it}}{\exp(\mathbf{x}_{it1}\hat{\boldsymbol{\beta}}_1)} \right].$$

Therefore, a consistent and $\sqrt{N}$-asymptotically normal estimator of $ASF_t(\mathbf{x}_{t1})$ is

$$\widehat{ASF}_t(\mathbf{x}_{t1}) = \hat{\theta}_{t1}\exp(\mathbf{x}_{t1}\hat{\boldsymbol{\beta}}_1).$$

One can compute derivatives or changes with respect to the elements of $\mathbf{x}_{t1}$, and insert interesting values. Obtaining a valid standard error for the resulting partial effects can be done via the delta method or bootstrapping.

Sometimes one wishes to have a single measure of partial effects, averaged across both the unobservables and observables. If x_{t1j} is continuous – for example, an element of $\mathbf{y}_{t2}$ – we usually obtain the derivative and then average. The average partial effect (APE) is

$$APE_{tj} = \beta_{1j}E_{(\mathbf{x}_{it1},c_{i1},r_{it1})}[c_{i1}\exp(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + r_{it1})]$$

and this is particularly easy to estimate because, by iterated expectations,

$$E_{(\mathbf{x}_{it1},c_{i1},r_{it1})}[c_{i1}\exp(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + r_{it1})] = E(y_{it}).$$

(This simplification comes because of the exponential mean function.) Therefore, for each t,

$$APE_{tj} = \beta_{1j}E(y_{it}),$$

and a simple, consistent estimator is $\hat{\beta}_{1j}\left(N^{-1}\sum_{i=1}^{N} y_{it}\right)$. In many cases one would average

across t as well to obtain a single partial effect.

A CRE/Control Function Approach

A CRE/CF approach can be used, although it requires more assumptions. Let

$$\begin{aligned} E(y_{it1}|\mathbf{y}_{it2}, \mathbf{z}_{it1}, c_{i1}, r_{it1}) &= c_{i1}\exp(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + r_{it1}) \\ v_{it1} &= c_{i1}\exp(r_{it1}) \\ \mathbf{y}_{it2} &= \boldsymbol{\psi}_2 + \mathbf{z}_{it}\boldsymbol{\Pi}_2 + \bar{\mathbf{z}}_i\boldsymbol{\Xi}_2 + \mathbf{v}_{it2} \end{aligned}$$

Then there are two possibilities. Papke and Wooldridge (2008) suggest modeling the conditional disribution

$$D(v_{it1}|\mathbf{z}_i, \mathbf{v}_{it2}),$$

where and assuming that this depends only on $(\bar{\mathbf{z}}_i, \mathbf{v}_{it2})$. While this approach leads to consistent estimation under maintained parametric assumptions, it does not lead to a straightforward test of idiosyncratic endogeneity: v_{it1} may be related to $\mathbf{v}_{it2}$ because of heterogeneity or idiosyncratic endogeneity. In addition. because we obtain an equation for $E(y_{it1}|\mathbf{x}_{it1}, \mathbf{z}_i, \mathbf{v}_{it2})$, only contemporaneous exogeneity holds because we are only conditioning on $\mathbf{v}_{it2}$ at time t. Therefore, only pooled methods can be used for consistent estimation.

Drawing on the linear case, a second possibility is attractive: Model the distribution

$$D(v_{it1}|\mathbf{z}_i, \mathbf{v}_{i2}).$$

Here, we use a Mundlak assumption:

$$\begin{aligned} D(v_{it1}|\mathbf{z}_i, \mathbf{v}_{i2}) &= D(v_{it1}|\bar{\mathbf{z}}_i, \bar{\mathbf{v}}_{i2}, \mathbf{v}_{it2}) \\ &= D(v_{it1}|\bar{\mathbf{z}}_i, \bar{\mathbf{y}}_{i2}, \mathbf{v}_{it2}). \end{aligned}$$

By construction, strict exogeneity holds for the conditioning variables, and so GLS-type procedures can be used. Moreover, even before we use a parametric model, this approach endogeneity of $\{\mathbf{y}_{it2}\}$ with respect to c_{i1} and $\{u_{it1}\}$.

If we use a linear index structure, the estimating equation is

$$E(y_{it1}|\mathbf{z}_i, \mathbf{y}_{i2}) = \exp(\psi_1 + \mathbf{x}_{it1}\boldsymbol{\beta}_1 + \bar{\mathbf{y}}_{i2}\boldsymbol{\pi}_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + \mathbf{v}_{it2}\boldsymbol{\rho}_1).$$

Identification of the parameters follows because the time-varying exogenous variables $\mathbf{z}_{it2}$ are excluded from $\mathbf{x}_{it1}$, and therefore generates variation in $\mathbf{v}_{it2}$. The presence of $\bar{\mathbf{y}}_{i2}$ and $\bar{\mathbf{z}}_i$ allows the unobserved heterogeneity to be correlated with all explanatory variables and the excluded exogenous variables. The test of $H_0 : \boldsymbol{\rho}_1 = \mathbf{0}$ is a clean test of idiosyncratic endogeneity, provided we assume our instruments are strictly exogenous and that the Mundlak device holds.

There are several approaches to estimating. The simplest is to use the pooled Poisson QMLE; naturally, we need to use fully robust inference to allow serial correlation and violations of the Poisson assumption. But we can also used a generalized least squares approach, where a "working" variance-covariance matrix is used to potentially increase efficiency over pooled estimation. Typically, one would use the Poisson variance, up to a scaling factor, as the "working" variances, and then choose a simple working correlation matrix – such as an exchangeable one, or at least one with constant pairwise correlations. Wooldridge (2010, Chapter 12) shows how the GEE approach is essentially multivariate weighted nonlinear least squares with a particular weighting matrix.

Because of the properties of the exponential function, it is possible to estimate the parameters $\boldsymbol{\beta}_1$ using a generalized method of moments approach on a particular set of nonlinear moment conditions. The GMM approach does not restrict that nature of $\mathbf{y}_{it2}$. See Wooldridge (1997) and Windmeijer (2000). At a minimum, one can use the test for idiosyncratic endogeneity based on the Poisson FE estimator before proceeding to a more complicated GMM proccedure.

4. Probit Response Function

With a probit conditional mean function, there are no versions of a fixed effects estimator that have attractive statistical properties, at least when T is not fairly large. Therefore, we consider only CRE/CF approaches to testing and correcting for endogeneity.

A probit conditional mean for $y_{it1} \in [0,1]$, which we consider the "structural" equation, is

$$E(y_{it1}|\mathbf{z}_i, \mathbf{y}_{i2}, c_{i1}, u_{it1}) = E(y_{it1}|\mathbf{z}_{it1}, \mathbf{y}_{it2}, c_{i1}, u_{it1}) = \Phi(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + c_{i1} + u_{it1}), \tag{4.1}$$

and this can hold when y_{it1} is binary or when it is a fractional response. We assume that $\mathbf{y}_{it2}$ continuous and write a Mundlak reduced form, as before:

$$\mathbf{y}_{it2} = \boldsymbol{\psi}_2 + \mathbf{z}_{it}\boldsymbol{\Pi}_2 + \bar{\mathbf{z}}_i\boldsymbol{\Xi}_2 + \mathbf{v}_{it2}$$

The important restriction (which can be relaxed to some degree) is

$$\mathbf{v}_{it2} \text{ is independent of } \mathbf{z}_i.$$

Define

$$r_{it1} = c_{i1} + u_{it1}.$$

Now we assume

$$D(r_{it1}|\mathbf{z}_i, \mathbf{v}_{i2}) = D(r_{it1}|\bar{\mathbf{z}}_i, \bar{\mathbf{v}}_{i2}, \mathbf{v}_{it2}) = D(r_{it1}|\bar{\mathbf{z}}_i, \bar{\mathbf{y}}_{i2}, \mathbf{v}_{it2}),$$

where the second equality holds simply because of the relationships among $\bar{\mathbf{z}}_i, \bar{\mathbf{y}}_{i2}$, and $\bar{\mathbf{v}}_{i2}$. In the leading case, we use a homoskedastic normal with linear mean:

$$r_{it1}|\bar{\mathbf{z}}_i, \bar{\mathbf{y}}_{i2}, \mathbf{v}_{it2} \sim Normal(\psi_1 + \bar{\mathbf{y}}_{i2}\boldsymbol{\pi}_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + \mathbf{v}_{it2}\boldsymbol{\rho}_1, 1).$$

We set the variance to unity because we cannot identify a separate variance, and it has no effect on estimating the average partial effects – see Papke and Wooldridge (2008) for further discussion. Then, an argument similar to that in Papke and Wooldridge (2008) gives the

estimating equation

$$E(y_{it1}|\mathbf{z}_i, \mathbf{y}_{i2}) = \Phi(\psi_1 + \mathbf{x}_{it1}\boldsymbol{\beta}_1 + \bar{\mathbf{y}}_{i2}\boldsymbol{\pi}_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + \mathbf{v}_{it2}\boldsymbol{\rho}_1),$$

which is clearly similar to the estimating equation in the exponential case.

Procedure 4.1 (CRE/CF Probit):

1. Obtain the Mundlak residuals, $\hat{\mathbf{v}}_{it2}$, by pooled OLS.

2. Insert $\hat{\mathbf{v}}_{it2}$ in place of $\mathbf{v}_{it2}$, use pooled (fractional) probit of

$$y_{it1} \text{ on } 1, \mathbf{x}_{it1}, \bar{\mathbf{y}}_{i2}, \bar{\mathbf{z}}_i, \hat{\mathbf{v}}_{it2}, t = 1, \ldots, T; \; i = 1, \ldots, N. \; \square$$

As in the linear case, Procedure 2.2, because $\hat{\mathbf{v}}_{it2} = \hat{\ddot{\mathbf{u}}}_{it2} + \hat{\mathbf{r}}_{i2}$ we can replace $\hat{\mathbf{v}}_{it2}$ with $\hat{\ddot{\mathbf{u}}}_{it2}$ and not change $\hat{\boldsymbol{\beta}}_1$ or $\hat{\boldsymbol{\rho}}_1$; only $\hat{\psi}_1$, $\hat{\boldsymbol{\pi}}_1$ and $\hat{\boldsymbol{\xi}}_1$ would change.

As before, we can use a cluster-robust Wald test of $H_0 : \boldsymbol{\rho}_1 = \mathbf{0}$ as a test of idiosyncratic exogeneity. Compared with Papke and Wooldridge (2008), $\bar{\mathbf{y}}_{i2}$ has been added to the equation, and doing so allows one to separate the two sources of endogeneity. Further, because the conditional mean satisfies a strict exogeneity assumption, we can use a GEE (quasi-GLS) procedure, although bootstrapping should be used to obtain valid standard errors. Technically, the assumptions under which Procedure 4.1 is consistent are different than those for the PW procedure, but in practice the difference is unlikely to be important. Procedure 4.1 leads to a cleaner test and also has the potential to produce more efficient estimators. Namely, GEE approaches can be used in place of the pooled probit estimation.

Consistent estimation of the APEs is also stratiforward. Using the same arguments in Papke and Wooldridge (2008),

$$APE_{tj} = \beta_{1j} E_{(\mathbf{x}_{it1},\bar{\mathbf{z}}_i,\bar{\mathbf{y}}_{i2},\mathbf{v}_{it2})}[\phi(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + \psi_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + \bar{\mathbf{y}}_{i2}\boldsymbol{\pi}_1 + \mathbf{v}_{it2}\boldsymbol{\rho}_1)]$$

$$\widehat{APE_{tj}} = \hat{\beta}_{1j}\left[N^{-1}\sum_{i=1}^{N} \phi(\mathbf{x}_{it1}\hat{\boldsymbol{\beta}}_1 + \hat{\psi}_1 + \bar{\mathbf{z}}_i\hat{\boldsymbol{\xi}}_1 + \bar{\mathbf{y}}_{i2}\hat{\boldsymbol{\pi}}_1 + \hat{\mathbf{v}}_{it2}\hat{\boldsymbol{\rho}}_1) \right]$$

To obtain a single value, $\widehat{APE_{tj}}$ can be averaged across t, too, and this is what would be produced by applying the Stata "margins" command after pooled estimation or GEE estimation. The standard error of the APE is complicated because of the two-step estimation and the averaging. Bootstrapping the entire procedure is practically sensible and not difficult computationally.

It can be shown that, just like the parameters, estimation of the APEs does not depend on whether $\hat{\mathbf{v}}_{it2}$ or $\hat{\hat{\mathbf{u}}}_{it2}$ is used as the control function.

It is easy to make Procedure 4.1 more flexible. For example, rather than just entering eace variable linearly, any nonlinear functions of

$$(\mathbf{x}_{it1}, \bar{\mathbf{z}}_i, \bar{\mathbf{y}}_{i2}, \hat{\mathbf{v}}_{it2})$$

can be included. These would typically include squares and cross products, but maybe higher order terms, too. One can still obtain the APEs by differentiating or differencing with respect to the elements of $\mathbf{x}_{t1}$ and then averaging across everything. For example, if we extend the estimating equation to

$$E(y_{it1}|\mathbf{z}_i, \mathbf{y}_{i2}) = \Phi(\psi_1 + \mathbf{x}_{it1}\boldsymbol{\beta}_1 + \bar{\mathbf{y}}_{i2}\boldsymbol{\pi}_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + \mathbf{v}_{it2}\boldsymbol{\rho}_1 + (\mathbf{x}_{it1} \otimes \bar{\mathbf{x}}_{i1})\boldsymbol{\psi}_1 + (\mathbf{x}_{it1} \otimes \mathbf{v}_{it2})\boldsymbol{\delta}_1),$$

then we simply add the terms $\mathbf{x}_{it1} \otimes \bar{\mathbf{x}}_{i1}$ and $\mathbf{x}_{it1} \otimes \hat{\mathbf{v}}_{it2}$ to the probit or fractional probit estimation. We then have to account for the interactions when taking derivatives, and then average the resulting function.

Another possibility is to allow the variance in the probit equation, whether fractional or

not, to depend on

$$(\bar{\mathbf{z}}_i, \bar{\mathbf{y}}_{i2}, \mathbf{v}_{it2}).$$

Then, one uses heteroskedastic probit or "fractional heteroskedastict probit" to allow c_{i1} to have nonconstant variance.

5. Other Nonlinear Models

5.1. Pooled Methods

The approach taken in the previous section applies to other nonlinear models, including the unobserved effects Tobit model. The approach is unchanged from the model with a probit response function. First, model the heterogeneity as a function of the history of the exogenous and endogenous variables, $(\mathbf{z}_i, \mathbf{y}_{i2})$, typically (but not necessarily) through simple functions, such as the time averages, $(\bar{\mathbf{z}}_i, \bar{\mathbf{y}}_{i2})$. Then add reduced-form Mundlak residuals, $\hat{\mathbf{v}}_{it2}$, in a pooled Tobit estimation. The key assumption is that for each t, y_{it1} conditional on $(\mathbf{z}_i, \mathbf{y}_{i2})$ follows a Tobit model with linear index $\psi_1 + \mathbf{x}_{it1}\boldsymbol{\beta}_1 + \bar{\mathbf{y}}_{i2}\boldsymbol{\pi}_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + \mathbf{v}_{it2}\boldsymbol{\rho}_1$ and constant variance. If we used a pooled estimation method then abitrary serial dependence is allowed. As usual, we must account for two-step estimation in calculating standard errors, and we must cluster to account for the serial dependence.

If y_{it1} is a count variable, and we prefer to use, say, a negative binomial model, then we can simple assume that, conditional on $(\mathbf{z}_{it1}, \mathbf{y}_{it2}, \bar{\mathbf{z}}_i, \bar{\mathbf{y}}_{i2}, \mathbf{v}_{it2})$, y_{it1} follows the appropriate model. Notice that we would not be able to derive such a model if we start with the assumption that the structural model for y_{it1} – conditional unobservables (c_{i1}, u_{it1}) as in the previous section – follow a negative binomial model. Therefore, purists may be reluctant to adopt such a strategy even though it would perhaps provide a good approximation that accounts for the count nature

of y_{it1}.

One can even apply the approach to less obvious situations, such as two-part models. For example, suppose the Tobit model is replaced by the Cragg (1971) truncated normal hurdle model – see also Wooldridge (2010, Section 17.6). Then one can model the two parts both as functions of $(\mathbf{z}_{it1}, \mathbf{y}_{it2}, \bar{\mathbf{z}}_i, \bar{\mathbf{y}}_{i2}, \mathbf{v}_{it2})$, and then separately test for endogeneity of $\mathbf{y}_{it2}$ in each part by testing coefficients on $\hat{\mathbf{v}}_{it2}$. Average partial effects are easily obtained by averaging out $(\bar{\mathbf{z}}_i, \bar{\mathbf{y}}_{i2}, \hat{\mathbf{v}}_{it2})$, across i or across (i,t), in the partial derivatives with respect to $\mathbf{x}_{t1}$. The form of the partial effects is given in, for example, Wooldridge (2010, equation (17.48)).

5.2. Joint Estimation Methods

So far our discussion has centered on pooled estimation methods. There are two reasons for this. First, pooled two-step methods are computationally simple, and panel bootstrap methods run quickly in most cases for obtaining valid standard errors. Second, and just as importantly, pooled methods are robust to any kind of serial dependence.

It is possible to apply the CRE/CF approach to joint MLE estimation in the second stage. For example, rather than using pooled probit, as in Section 5, one might want to estimate a so-called random effects probit in the second stage. The explanatory variables would be

$$(\mathbf{x}_{it1}, \bar{\mathbf{z}}_i, \bar{\mathbf{y}}_{i2}, \hat{\mathbf{v}}_{it2}),$$

where recall $\mathbf{x}_{it1}$ is a function of $(\mathbf{z}_{it1}, \mathbf{y}_{it2})$. Or, we could use more flexible functions of the histories $(\mathbf{z}_i, \mathbf{y}_{i2})$. While joint MLEs can be used in the second stage, one should be aware of the costs of doing so. First, computationally joint MLEs are usually significantly more difficult to obtain than pooled MLEs. While the difference in computational times is often irrelevant for one pass through the data, adding $\hat{\mathbf{v}}_{it2}$ to account for idiosyncratic endogeneity of $\mathbf{y}_{it2}$ requires

some sort of adjustment for inference, althoug testing the null that $\hat{\mathbf{v}}_{it2}$ has zero coefficients does not require an adjustment. If one uses the bootstrap then the increased computational burden can be nontrivial.

The second cost to use joing MLE in the second step is lack of robustness to distributional misspecification and serial dependence. Standard joint MLEs used for nonlinear random effects models maintain that innovations – what we would call $\{u_{it1}\}$ in equations (2.1) and (4.1) – are independent over time, as well as being independent of c_{i1} and $\mathbf{z}_i$. None of random effects probit, RE logit, RE Tobit, RE Poisson, and so on have robustness properties in the presence of serial correlation of the innovations. Moreover, even if the innovations in (4.1) are serially independent, the RE probit joint MLE is not known to be consistent

When we apply a joint MLE in the second step, there is another subtle point. Suppose we express the relationship between innovations in, say, (4.1) and those in the reduced form of $\mathbf{y}_{it2}$, $\mathbf{v}_{it2}$, as

$$u_{it1} = \mathbf{v}_{it2}\boldsymbol{\rho}_1 + e_{it1}.$$

The relevant innovations underlying the joint MLE in the second step are $\{e_{it1}\}$, not $\{u_{it1}\}$ – unless $\boldsymbol{\rho}_1 = \mathbf{0}$. Consequently, serial correlation in the reduced form of $\mathbf{y}_{it2}$ can cause serial correlation in the second stage MLE, even though there was none in the original innovations.

For robustness and computational reasons, the pooled methods are generally preferred. Future research could focus on how to improve in terms of efficiency over the pooled methods without adding assumptions.

6. Empirical Example

Papke and Wooldridge (2008) estimate the effect of spending on fourth-grade math test

past rates using data from Michigan. The years straddle the Michigan School Reform, which was passed in 1995. The response variable, *math*4, is a pass rate, and so we use a fractional probit model response in addition to a linear model estimated by fixed effects IV. The variable of interest is the natural log of real per pupil spending, averaged over the current and previous three years. The instrumental variable is the "foundation allowance." which is the amount given by the state to each school district – after the spending reform. A kinked relationship between the allowance and pre-reform per-pupil revenue means that, once a district effect is controlled for, the foundation allowance is exogenous. Not surprisingly, its log is a very strong instrument for the log of average real spending. Other controls include the proportion of students eligble for free and reduced lunch and the log of district enrollment. A full set of year effects is also included. There are $N = 501$ school districts over the seven years 1995 to 2001.

The results of the test are given in Table 1 for the spending variable. The linear fixed effects estimate, $.377$, implies that a 10% increase in average spending increases the pass rate by about 3.8 percentage points, and the effect is very statistically significant. The FEIV estimate actually increases to $.420$, and remains strong significant. The fully robust test of idiosyncratic endogeneity, where the null is exogeneity, gives $t = -.41$, which is not close to being statistically significant. Therefore, the evidence is that, once spending is allowed to be correlated with the district heterogeneity, spending is not endogenous with respect to idiosyncratic shocks.

Columns (3) and (4) in Table 1 apply the fractional probit CRE/CF approaches. In column (3) we apply Procedure 4.1, which includes the time average of *lavgrexp* along with the time average of all exogenous variables, including *lfound*, the log of the foundation allowance. The coefficient is $.821$ and it is strongly statistically significant. The APE, which is comparable to

the FEIV estimate, is quite a bit lower: $.277$, but with $t = 2.47$ is still pretty significant. The test for idioysyncratic endogeneity fails to reject the null of exogeneity, with $t = .52$. This is entirely consistent with the linear model estimates and test. By contrast, when we apply the Papke-Wooldridge approach in column (4), the *t* statistic for the coefficient on the reduced form residual $\hat{v}_2$ is $t = -1.68$, which is significant at the 10% level. This is not a strong rejection of exogeneity, but it is much stronger than when the time average of *lavgrexp*. The outcomes in columsn (3) and (4) are consistent with the conclusion that spending is correlated with district-level heterogeneity but not district-level shocks, which is why the test in column (3) marginall rejected exogeneity and that in column (4) does not come close to rejecting. In the end, the new approach in column (3) and the PW approach in column (4) given very similar estimates of the APE of spending: $.277$ versus $.269$, and the standard errors are similar.

Table 1

Model:	**Linear**	**Linear**	**FProbit**		**FProbit**	
Estimation:	**FE**	**FEIV**	**PQMLE**		**PQMLE**	
	Coef	Coef	Coef	APE	Coef	APE
lavgrexp	.377 (.071)	.420 (.115)	.821 (.334)	.277 (.112)	.797 (.338)	.269 (.114)
$\hat{\bar{u}}_2$	—	−.060 (.146)	—	—	—	—
$\hat{v}_2$	—	—	.076 (.145)	—	−.666 (.396)	—
$\overline{lavgrexp}$?	—	—	Yes		No	

7. Extensions and Future Directions

The main message in this paper is that, when combining the correlated random effects and control function approaches in nonlinear panel data models, there is a good case to separately model – even if only implicitly – the distribution of the heterogeneity conditional on all explanatory variables and outside exogenous variables. In this way, adding the control functions to account for idiosyncratic endogeneity leads to a pure test of the null hypothesis of exogeneity. In linear models, a common variable addition test after fixed effects estimation achieves this goal. We have shown how the same goal can be acheived for two popular nonlinear models.

We have used parametric assumptions in our discussion and applications. Nevertheless, when the EEVs $\mathbf{y}_{it2}$ are continuous, there is a more general message when semiparametric, or even purely nonparametric, approaches are taken. For example, when applying the insights of Blundell and Powell (2004), it makes sense to separately include functions of the entirely history, $(\mathbf{y}_{i2}, \mathbf{z}_i)$, and the control functions, $\hat{\mathbf{v}}_{it2}$. We touched on this at the end of Section 5, where we showed a model with interactions between the variables of interests, the time averages, and the control functions can be added for flexibility. The general point is that by adding, say, $\bar{\mathbf{y}}_{i2}$ along with $\bar{\mathbf{z}}_i$ we then obtain an estimating equation where the addition of $\hat{\mathbf{v}}_{it2}$ is purely to account for possible idiosyncratic endogeneity.

In nonlinear models, the assumptions imposed on the reduced form of $\mathbf{y}_{it2}$ will not be met when $\mathbf{y}_{it2}$ has discreteness. Even allowing for a single binary EEV y_{it2} poses challenges for nonlinear unobserved effects panel data models. In particular, the parametric assumptions that can be viewed as convenient approximations when $\mathbf{y}_{it2}$ now have real bite when it comes to identifying the average partial effects. If one is willing to make distributional assumptions –

such as normality in the probit case – the methods in Wooldridge (2014) and Lin and Wooldridge (2016) can be extended to allow correlated random effects. As just one simple example, if y_{it2} is assumed to follow a reduced form probit, one can use as a control function the generalized residuals,

$$\widehat{gr}_{it2} = y_{it2}\lambda(\mathbf{w}_{it}\hat{\boldsymbol{\theta}}_2) - (1 - y_{it2})\lambda(-\mathbf{w}_{it}\hat{\boldsymbol{\theta}}_2),$$

where $\mathbf{w}_{it} = (1, \mathbf{z}_{it}, \bar{\mathbf{z}}_i)$. But then the issue of how to best model the relationship between heterogeneity and $(\mathbf{y}_{i2}, \mathbf{z}_i)$ arises. The Munklak device, or Chamberlain's version of it, may work reasonably well, but they may not be flexible enough. We leave investigations into the quality of CF approximations in discrete cases to future research.

As discussed in Wooldridge (2018), unbalanced panels pose challenges for the correlated random effects approach, although the challenges are not insurmountable. In the context of heterogeneity endogeneity only, Wooldridge suggests a modeling strategy where unobserved heterogeneity is a function of $\{(s_{it}, s_{it}\mathbf{x}_{it}) : t = 1, \ldots, T\}$, where s_{it} is a binary selection indicator which is unity when a complete set of data is observed for unit *i* in time *t*. This approach can be extended to the current setting, but the details remain to be worked out.

References

Altonji, J.G. and R.L. Matzkin (2005), "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica* 73, 1053-1102.

Blundell, R. and J.L. Powell (2004), "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies* 71, 655-679.

Chamberlain, G. (1982), "Multivariate Regression Models for Panel Data," *Journal of Econometrics* 1, 5-46.

Hausman, J.A., B.H. Hall, and Z. Griliches (1984), "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." *Econometrica* 52, 909-938.

Levitt, S.D. (1996), "The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation," *Quarterly Journal of Economics* 111, 319-351.

Levitt, S.D. (1997), "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime," *American Economics Review* 87, 270-290.

Lin, W. and J.M. Wooldridge (2016), "Binary and Fractional Response Models with Continuous and Binary Endogenous Explanatory Variables," working paper, Michigan State University Department of Economics.

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

Murtazashvili, I. and J.M. Wooldridge (2016), "A Control Function Approach to Estimating Switching Regression Models with Endogenous Explanatory Variables and Endogenous Switching," *Journal of Econometrics* 190, 252-266.

Papke, L.E. and J.M Wooldridge (2008), "Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates," *Journal of Econometrics* 145, 121-133.

Windmeijer, F. (2000), "Moment Conditions for Fixed Effects Count Data Models with Endogenous Regressors," *Economics Letters* 68, 21-24.

Wooldridge, J.M. (1990), "A Unified Approach to Robust, Regression-Based Specification Tests," *Econometric Theory* 6, 17-43.

Wooldridge, J.M. (1997), "Multiplicative Panel Data Models without the Strict Exogeneity Assumption," *Econometric Theory* 13, 667-678.

Wooldridge, J.M. (1999), "Distribution-Free Estimation of Some Nonlinear Panel Data Models," *Journal of Econometrics* 90, 77-97.

Wooldridge, J.M. (2010), *Econometric Analysis of Cross Section and Panel Data*, second edition. Cambridge, MA: MIT Press.

Wooldridge, J.M. (2014), "Quasi-Maximum Likelihood Estimation and Testing for Nonlinear Models with Endogenous Explanatory Variables," *Journal of Econometrics* 182, 226-234.

Wooldridge, J.M. (2015), "Control Function Methods in Applied Econometrics," *Journal of Human Resources* 50, 420-445.

Wooldridge, J.M. (2018), "Correlated Random Effects Models with Unbalanced Panels," forthcoming, *Journal of Econometrics*.

Appendix

This appendix verifies some of the algebraic claims made in Sections 2 and 3.

A.1. Relationship Between the FE and Mundlak Residuals

We firstfind a relationship between the FE residuals and the Mundlak residuals. Let $\mathbf{w}_i$ be any collection of time-constant variables. The FE and Mundlak residuals are, respectively,

$$\begin{aligned}\ddot{\hat{u}}_{it} &= \ddot{y}_{it} - \ddot{\mathbf{x}}_{it}\hat{\boldsymbol{\beta}}_{FE} \\ \hat{v}_{it} &= y_{it} - \mathbf{x}_{it}\hat{\boldsymbol{\beta}}_{FE} - \hat{\psi} - \bar{\mathbf{x}}_i\hat{\boldsymbol{\xi}} - \mathbf{w}_i\hat{\boldsymbol{\lambda}},\end{aligned}$$

where we use the fact that the estimates $\mathbf{x}_{it}$ are identical using FE and the Mundlak approaches. Further, because $\ddot{\mathbf{x}}_{it}$ is a nonsingular linear combination of $\mathbf{x}_{it}$ and $\bar{\mathbf{x}}_i$, we obtain the same Mundlak residuals if instead we run the pooled regression

$$y_{it} \text{ on } \ddot{\mathbf{x}}_{it},\ 1,\ \bar{\mathbf{x}}_i,\ \mathbf{w}_i$$

In fact, we can add on $\bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}_{FE}$ and subtract it off:

$$\begin{aligned}\hat{v}_{it} &= y_{it} - (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\hat{\boldsymbol{\beta}}_{FE} - \hat{\psi} - \bar{\mathbf{x}}_i\left(\hat{\boldsymbol{\xi}} + \hat{\boldsymbol{\beta}}_{FE}\right) - \mathbf{w}_i\hat{\boldsymbol{\lambda}} \\ &= y_{it} - \ddot{\mathbf{x}}_{it}\hat{\boldsymbol{\beta}}_{FE} - \hat{\psi} - \bar{\mathbf{x}}_i\left(\hat{\boldsymbol{\xi}} + \hat{\boldsymbol{\beta}}_{FE}\right) - \mathbf{w}_i\hat{\boldsymbol{\lambda}} \\ &\equiv y_{it} - \ddot{\mathbf{x}}_{it}\hat{\boldsymbol{\beta}}_{FE} - \hat{\psi} - \bar{\mathbf{x}}_i\hat{\boldsymbol{\delta}} - \mathbf{w}_i\hat{\boldsymbol{\lambda}}\end{aligned}$$

From Mundlak (1978), it is known that $(\hat{\psi}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\lambda}})$ are the between estimates, that is, from the cross section OLS regression

$$\bar{y}_i \text{ on } 1,\ \bar{\mathbf{x}}_i,\ \mathbf{w}_i.$$

This is easy to see directly in our setup. Define $\mathbf{z}_i = (1, \bar{\mathbf{x}}_i, \mathbf{w}_i)$ and let $\hat{\boldsymbol{\theta}}$ be the set of coefficients: $(\hat{\psi}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\lambda}})$. Then

$$\sum_{t=1}^{T} \mathbf{z}_i'\ddot{\mathbf{x}}_{it} = \mathbf{z}_i'\sum_{t=1}^{T}\ddot{\mathbf{x}}_{it} = \mathbf{0}$$

so that the regressors are orthogonal in sample. By Frisch-Waugh, $\hat{\boldsymbol{\theta}}$ is also obtained by dropping $\ddot{\mathbf{x}}_{it}$, that is, from

$$y_{it} \text{ on } \mathbf{z}_i,\ t = 1,\ldots,T; i = 1,\ldots,N.$$

But

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{z}_i'\mathbf{z}_i\right)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{z}_i'y_{it} \\ &= \left(T\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{z}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{z}_i'\sum_{t=1}^{T}y_{it} = \left(\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{z}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{z}_i'\left(T^{-1}\sum_{t=1}^{T}y_{it}\right) \\ &= \left(\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{z}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{z}_i'\bar{y}_i = \hat{\boldsymbol{\theta}}_B\end{aligned}$$

Now we can write

$$\begin{aligned}\hat{v}_{it} &\equiv y_{it} - \bar{y}_i - \ddot{\mathbf{x}}_{it}\hat{\boldsymbol{\beta}}_{FE} + \bar{y}_i - \hat{\psi}_B - \bar{\mathbf{x}}_i\hat{\boldsymbol{\delta}}_B - \mathbf{w}_i\hat{\boldsymbol{\lambda}}_B \\ &= \ddot{y}_{it} - \ddot{\mathbf{x}}_{it}\hat{\boldsymbol{\beta}}_{FE} + \left(\bar{y}_i - \hat{\psi}_B - \bar{\mathbf{x}}_i\hat{\boldsymbol{\delta}}_B - \mathbf{w}_i\hat{\boldsymbol{\lambda}}_B\right) \\ &= \hat{\ddot{u}}_{it} + \hat{r}_i,\end{aligned}$$

where $\hat{r}_i$ is the between residual. One important feature of this relationship is that $\hat{r}_i$ does not change over time. Therefore,

$$\sum_{t=1}^{T}\hat{r}_i\hat{\ddot{u}}_{it} = 0.$$

More importantly, for demeaned variables $\ddot{\mathbf{x}}_{it}$,

$$\sum_{t=1}^{T}\ddot{\mathbf{x}}_{it}'\hat{v}_{it} = \sum_{t=1}^{T}\ddot{\mathbf{x}}_{it}'\hat{\ddot{u}}_{it}$$

because $\sum_{t=1}^{T}\ddot{\mathbf{x}}_{it}'\hat{r}_i = 0$.

A.2. Equivalence in Using the FE and Mundlak Residuals in FE

Poisson Estimation

Now we obtain a general result that shows that adding time-constant variables to the explanatory variables does not affect $\hat{\boldsymbol{\beta}}$ in the Poisson FE case. For a cross-section observation i, the quasi-log likelihood is

$$\ell_i(\boldsymbol{\beta}) = \sum_{t=1}^{T} y_{it}\left\{\mathbf{x}_{it}\boldsymbol{\beta} - \log\left[\sum_{r=1}^{T}\exp(\mathbf{x}_{ir}\boldsymbol{\beta})\right]\right\},$$

and the score is

$$\mathbf{s}_i(\boldsymbol{\beta}) = \sum_{t=1}^{T} y_{it}\left\{\mathbf{x}_{it}' - \frac{\sum_{r=1}^{T}\mathbf{x}_{ir}'\exp(\mathbf{x}_{ir}\boldsymbol{\beta})}{\sum_{r=1}^{T}\exp(\mathbf{x}_{ir}\boldsymbol{\beta})}\right\}$$

Therefore, the FOC is

$$\sum_{i=1}^{N}\mathbf{s}_i\left(\hat{\boldsymbol{\beta}}\right) = \mathbf{0}.$$

Now suppose

$$\mathbf{x}_{it} = \mathbf{g}_{it} + \mathbf{h}_i,$$

which allows for the case that some $\mathbf{h}_i$ are identically zero for all i. Then for any i,

$$\begin{aligned}
\mathbf{s}_i\left(\hat{\boldsymbol{\beta}}\right) &= \sum_{t=1}^{T} y_{it}\left\{\mathbf{x}_{it}' - \frac{\sum_{r=1}^{T}\mathbf{x}_{ir}'\exp(\mathbf{g}_{ir}\hat{\boldsymbol{\beta}} + \mathbf{h}_i\hat{\boldsymbol{\beta}})}{\sum_{r=1}^{T}\exp(\mathbf{g}_{ir}\hat{\boldsymbol{\beta}} + \mathbf{h}_i\hat{\boldsymbol{\beta}})}\right\} = \sum_{t=1}^{T} y_{it}\left\{\mathbf{x}_{it}' - \frac{\exp(\mathbf{h}_i\hat{\boldsymbol{\beta}})\sum_{r=1}^{T}\mathbf{x}_{ir}'\exp(\mathbf{g}_{ir}\hat{\boldsymbol{\beta}})}{\exp(\mathbf{h}_i\hat{\boldsymbol{\beta}})\sum_{r=1}^{T}\exp(\mathbf{g}_{ir}\hat{\boldsymbol{\beta}})}\right\} \\
&= \sum_{t=1}^{T} y_{it}\left\{\mathbf{x}_{it}' - \frac{\sum_{r=1}^{T}\mathbf{x}_{ir}'\exp(\mathbf{g}_{ir}\hat{\boldsymbol{\beta}})}{\sum_{r=1}^{T}\exp(\mathbf{g}_{ir}\hat{\boldsymbol{\beta}})}\right\} = \sum_{t=1}^{T} y_{it}\left\{(\mathbf{g}_{it}' + \mathbf{h}_i') - \frac{\sum_{r=1}^{T}(\mathbf{g}_{ir}' + \mathbf{h}_i')\exp(\mathbf{g}_{ir}\hat{\boldsymbol{\beta}})}{\sum_{r=1}^{T}\exp(\mathbf{g}_{ir}\hat{\boldsymbol{\beta}})}\right\} \\
&= \sum_{t=1}^{T} y_{it}\left\{\mathbf{g}_{it}' - \frac{\sum_{r=1}^{T}\mathbf{g}_{ir}'\exp(\mathbf{g}_{ir}\hat{\boldsymbol{\beta}})}{\sum_{r=1}^{T}\exp(\mathbf{g}_{ir}\hat{\boldsymbol{\beta}})} + \mathbf{h}_i' - \mathbf{h}_i'\frac{\sum_{r=1}^{T}\exp(\mathbf{g}_{ir}\hat{\boldsymbol{\beta}})}{\sum_{r=1}^{T}\exp(\mathbf{g}_{ir}\hat{\boldsymbol{\beta}})}\right\} \\
&= \sum_{t=1}^{T} y_{it}\left\{\mathbf{g}_{it}' - \frac{\sum_{r=1}^{T}\mathbf{g}_{ir}'\exp(\mathbf{g}_{ir}\hat{\boldsymbol{\beta}})}{\sum_{r=1}^{T}\exp(\mathbf{g}_{ir}\hat{\boldsymbol{\beta}})}\right\}
\end{aligned}$$

Note that the final expression is the score with explanatory variables $\mathbf{g}_{it}$, and so we have shown

$\hat{\boldsymbol{\beta}}$ is the same whether we use $\mathbf{x}_{it}$ or $\mathbf{g}_{it}$.

The above result applies to the control function estimation in Section 3 because, as shown in Appendix A.1,

$$\hat{\mathbf{v}}_{it2} = \hat{\ddot{\mathbf{u}}}_{it2} + \hat{\mathbf{r}}_{i2},$$

where $\hat{\mathbf{r}}_{i2}$ are the between residuals and do not vary over time. The other explanatory variables are unchanged. Therefore, we obtain the same estimates whether we obtain the FE residuals in the first stage or the Mundlak residuals.