Centre de Recherche en Economie Publique et de la Population

CREPP WP No 2011/13

Nonlinear Panel Data Models with Expected a Posteriori Values of Correlated Random

Amaresh Tiwari
Franz Palm

December 2011

CREPP - Centre de recherche en Economie Publique et de la Population
Boulevard du Rectorat 7 (B31)
4000 Liège Belgium

CREPP Working Paper Series

Nonlinear Panel Data Models with Expected á Posteriori Values of Correlated Random Effects

Amaresh K Tiwari*and Franz C. Palm†‡

November, 2011

Abstract

We develop a two step estimation procedure to estimate nonlinear panel data models. Our approach combines the "correlated random effect" and the "control function" approach to handel endogeneity of regressors that are correlated with both the unobserved heterogeneity as well as the idiosyncratic component. The novelty here lies in integrating out the unobserved heterogeneity on which the structural equations are conditioned. The integration is performed with respect to the posterior distribution of the individual effects obtained from the first stage reduced form estimation. Our framework suggests separate tests for correlation between unobserved heterogeneity and the covariates, and correlation between idiosyncratic component and the covariates. Average partial effects (APEs) of covariates are also easily obtained.

JEL Classification: C13, C18, C33

Key Words: Correlated Random Effects (CRE), Endogeneity, Average Partial Effects (APE), Expected a Posteriori (EAP), Multidimensional Numerical Integration.

1 Introduction

Panel data, consisting of observations a cross time for different individual, allow the possibility of controlling for unobserved time invariant individual heterogeneity. Such heterogeneity can be an important phenomenon, and failure to control for it can result in misleading inferences. This problem is particularly severe when the unobserved heterogeneity is correlated with explanatory variables. Models and methods of controlling for unobserved heterogeneity in linear models are well established; see Chamberlain (1984) and Arellano and Honore (2001) for references and discussion. Controlling for unobserved heterogeneity is much more difficult

*University of Liege, A.Tiwari@ulg.ac.be

†University of Maastricht and CESifo fellow, F.Palm@maastrictuniversity.nl

‡The authors wish to thank seminar participants at Center for Operations Research and Econometrics (CORE), University of Liege, and Bernard Lejeune for their helpful suggestions. All remaining errors are our own.

in nonlinear models. Chamberlain (2010) and Arellano and Bonhomme (2011) point out that when panel data outcomes are discrete, serious identification issues arise.

The identification question posed by Chamberlain (2010) is the following: is there a unique value of structural parameters such that one can estimate the "average partial effect" (APE) by integrating out the individual effects out of the likelihood for some conditional distribution of individual effects? In the static binary choice model with two time periods and exogenous covariates with bounded support, Chamberlain (2010) finds that the structural parameters are not point-identified, unless the distribution of the idiosyncratic component is the logistic distribution. Notwithstanding this underidentification result, various methods have evolved to estimate the structural parameters of interest and APE. Weidner (2011) provides with a brief overview, and categorizes, of some of the methods developed to estimate the quantities of interest.

Now, what is desirable in panel data analysis is point identification of the quantities of interest, such as APE, where the number of time periods, T, remain fixed while the number of cross sectional units, N, become large. However, at fixed T a non-linear panel data model may not be point identified, or may not possess a $\sqrt{N}$ consistent estimator, as discussed by Chamberlain (2010) for the binary choice model. One of the leading methods in the literature is the fixed effect (FE) approach that treat the heterogeneity or individual effects as parameters to be estimated. But we know that an incidental parameter problem (see Neyman and Scott (1948) and Lancaster (2000) for a review) usually appears in fixed T estimation of non-linear panel data models since the number of incidental parameters (individual effects) grows with the sample size. More recently, it has been argued that the incidental parameter problem can be viewed as time-series finite-sample bias when T tends to infinity. Following this perspective, several approaches have been proposed to correct for the time-series bias. Some of the papers that follow the bias reduction technique for estimating the quantities of interest are Hahn and Newey (2004), Hahn and Kuersteiner (2011), Woutersen (2002), Arellano (2003), Carro (2007), Arellano and Hahn (2007) Arellano and Hahn (2006), Bester and Hansen (2009), and Fernandez-Val (2009).

Wooldridge (2009), however, points out that the fixed effect approach, though promising, suffer from a number of shortcomings. First, the number of time periods needed for the bias adjustments to work well is often greater than is available in many applications. Second, an important point is that recent bias adjustments include the assumptions of stationarity and weak dependence; in some cases, the very strong assumption of serial independence (conditional on the heterogeneity) is maintained. But it has been found that in empirical work dealing with linear models that there are sources of serial correlation that arise due serial correlation in the idiosyncratic errors in addition to that caused by unobserved heterogeneity. The requirement of stationarity is also very strong and has substantive restrictions: it rules out staples in empirical work such as including separate year effects, which can be estimated very precisely given a large cross section. Also, there is the technical problem of allowing separate period effects when large-sample approximations involve a growing number of time periods, as it effectively introduces an incidental parameters problem in the time series dimension.

There is another class of models that acknowledge the fact that many non-linear panel data models are not point identified at fixed T and consequently discuss set identification (bound analysis) for the parameters of interest or for certain policy parameters like marginal effects. These papers show that show that the bounds become tighter as the number of time periods, T, increases. The papers that deal with bound analysis include for Chernozhukov, Hahn, and Newey (2005), Honore and Tamer (2006), Chernozhukov, Fernandez-val, Hahn, and Newey (2009) and Chernozhukov, Fernandez-val, and Newey (2009). However, with the exception of Honore and Tamer (2006), as Wooldridge (2009) points out, these methods are very promising but a still limited to discrete covariates. Moreover, these papers and papers utilizing FE approach only deal only with regressors, which conditional on unobserved heterogeneity, are exogenous or predetermined, and do not take endogeneity with respect to the idiosyncratic errors into account.

In this paper we adopt the "control function" approach to model a "correlated random effect" (CRE) estimator for non-linear panel data. The approach does entail restriction on the conditional distribution of the individual effects and the idiosyncratic component, but as Wooldridge (2009) argues, estimation using CRE and FE involve tradeoffs among assumptions and the type of quantities that can be estimated, and that no method provides consistent estimators of either parameters or APEs under a set of assumptions strictly weaker than the assumptions needed for the other procedures. Some of the recent papers that adopt the CRE approach to control for heterogeneity are Chernozhukov, Hahn, and Newey (2005), Bester and Hansen (2007), Papke and Wooldridge (2008) (henceforth PW), and Weidner (2011). While

Bester and Hansen (2007) and Weidner (2011) study semiparametric models, and do not specify the conditional distribution of the individual effects, PW assumes a parametric form. Weidner (2011) employs "generalized random effects" as form of constraint on the structure of this correlation between unobserved heterogeneity and covariates to draw inference on the conditional distribution of the individual effects. This form of constraint is employed to reduce the large dimensional support of the conditioning variables so that the curse of dimensionality is allayed when treating the individual effects non-parametrically. Bester and Hansen (2007) establish restrictions on the space of functions to which the conditional distribution of unobserved heterogeneity belongs so that both the parameters of interest and conditional distribution of unobserved heterogeneity are identified.

This paper explores ways to consistently estimate "average partial effects" (APE) in presence of endogenous regressors that are correlated with both the unobserved heterogeneity and idiosyncratic component. The paper that is closest to ours is the paper by PW, who also take endogeneity of regressors with respect to idiosyncratic component into account. While PW and our paper both consider a two-step procedure to construct control variables to account for endogeneity, our method and the form of the control functions differs from their's.

Typically, in a control function approach the structural parameters are estimated conditional on unobserved heterogeneity and unobserved idiosyncratic errors that appear in reduced form equations of a simultaneous triangular system of equations. In such an approach residuals obtained from the first stage reduced form estimates, that proxy for the idiosyncratic errors are used as control variables in the structural equations to account for the endogeneity of the regressors in the structural equations. However, in panel data models, where we want to account for unobserved individual effects, the residuals remain unidentified. This is because the residuals of the reduced form regression, which are defined as the observed value of the response variable less the expected value of the observed conditional on exogenous regressors and the individual effects, are functions of unobserved individual effects/unobserved heterogeneity and these individual effects are unobserved. The novelty of our approach lies in integrating out the unobserved time invariant individual effects on which the structural equations are conditioned, and which also appear in the residuals of the reduced form equations. The integration is performed with respect to the posterior distribution of the individual effects obtained from the first stage reduced form estimation. This leaves us with the expected a posteriori (EAP) values of the individual effects, which can then be used to get the residuals. We would also like to mention that our approach of obtaining control functions that are based on (EAP) values of the individual effects is non-standard as far as the econometric literature is concerned. This is because numerical integration with respect to estimated – estimated in the first stage – parametric distribution of the individual effects has to be performed in order to obtain the (EAP) values of the individual effects. This creates additional difficulties for computing the error adjusted covariance matrix of the second stage structural parameters. In Appendix C we show how to compute the error adjusted covariance matrix for the estimates of the structural parameters.

Notwithstanding the computational difficulties, our framework suggests, first, a straightforward and a precise tests of correlation between unobserved heterogeneity and the deemed endogenous covariates, as well as correlation with the unobserved shocks. The number of control variables in the structural equation, as it turns out, is equal to the twice the number of endogenous regressors, one set to control for the endogeneity with respect to individual effect and the other to control for endogeneity with respect to idiosyncratic component. This is in one crucial way that our method differs from PW. Secondly, since the EAP values of the individual effects are functions of exogenous covariates, the correlatedness of the exogenous covariates and the individual effects in the structural equations are accounted for, circumventing the need for a Mundlak (1978) or Chamberlain (1984) type specification of the conditional distribution of the unobserved heterogeneity in the second stage structural equation. This has the added advantage of (a) conserving on the degrees of freedom and (b) allowing us to estimate the structural parameters of interest with more precision, especially when there is not enough variation among the regressors across time.

While we can do all this, our model still retains the attractive features of the PW, namely, no assumptions on the serial dependence in the response variable, and the suspected endogenous explanatory variable is allowed to arbitrarily correlate with unobserved shocks in other time periods.

The rest of the paper is organized as follows. Section 2 introduces the model, Section 2.1 discuss identification for continuous response model, which is a precursor to the discussion of identification of APE for discrete response model in Section 2.2. In Section 3 we provide the concluding remarks. Technical details for computing error adjusted covariance matrix for the second stage parameter estimates are provided in Appendix C. Finally, in Appendix D we provide a note on methodology employed to carry out numerical integration.

2 Model Specification

$$\mathbf{y}^*_{it} = \mathbf{Z}^{y\prime}_{it}\boldsymbol{\varphi} + \mathbf{X}'_{it}\tilde{\boldsymbol{\varphi}} + \boldsymbol{\theta}_i + \boldsymbol{\zeta}_{it}, \tag{2.1}$$

$$\mathbf{x}_{it} = \mathbf{Z}'_{it}\boldsymbol{\beta} + \tilde{\boldsymbol{\alpha}}_i + \boldsymbol{\epsilon}_{it}, \tag{2.2}$$

Equation (2.1) is the system of 'n' structural equations, where $\mathbf{Z}^y_{it} = \text{diag}(\mathbf{z}^y_{1it}, \ldots, \mathbf{z}^y_{nit})$ and each of the $\mathbf{z}^y_{kit}$, $k \in \{1, \ldots, n\}$, is a vector of strictly exogenous variables. Let $\mathbf{z}^y_{it}$ be the union of the exogenous variables appearing in $\mathbf{Z}^y_{it}$. $\mathbf{X}_{it} = \text{diag}(\mathbf{x}_{1it}, \ldots, \mathbf{x}_{nit})$, where each of the $\mathbf{x}_{kit}$, $k \in \{1, \ldots, n\}$, is a vector of endogenous regressors appearing in k^{th} structural equation. $\boldsymbol{\theta}_i = (\theta_{1i}, \ldots, \theta_{ni})'$ is the vector of unobserved time invariant individual effect, while $\boldsymbol{\zeta}_{it} = (\zeta_{1it}, \ldots, \zeta_{nit})'$ is the vector of idiosyncratic error component.

Equation (2.2) is the system of 'm' equations written in a reduced form for the endogenous variables $\mathbf{x}_{it}$, where $\mathbf{x}_{it}$ is continuous and is the union of all endogenous regressors in $(\mathbf{x}'_{1it}, \ldots, \mathbf{x}'_{nit})'$. $\mathbf{Z}_{it} = \text{diag}(\mathbf{z}_{1it}, \ldots, \mathbf{z}_{mit})$ is the matrix of exogenous variables or instruments appearing in each of the m reduced form equation in (2.2) and $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_m)'$. For every $l \in (1, \ldots, m)$, $\mathbf{z}_l = \mathbf{z} = (\mathbf{z}^{y\prime}, \tilde{\mathbf{z}}')'$, where the dimension of $\tilde{\mathbf{z}}$ is greater than or equal to the dimension $\mathbf{x}$. $\tilde{\boldsymbol{\alpha}}_i = (\tilde{\alpha}_{i1}, \ldots, \tilde{\alpha}_m)'$ are the unobserved individual effect for each of the m equation, and $\boldsymbol{\epsilon}_{it} = (\epsilon_{1it}, \ldots, \epsilon_{mit})'$ is the vector of idiosyncratic error terms. Let $\mathcal{Z}_i = (\mathbf{z}'_{i1}, \ldots, \mathbf{z}'_{iT})'$ and $\mathbf{X}_i = \{\mathbf{x}'_{i1}, \ldots, \mathbf{x}'_{iT}\}'$.

We assume that the unobserved individual effects $\boldsymbol{\theta}_i$ and $\tilde{\boldsymbol{\alpha}}_i$, which we model as a random effects, are normally distributed as

$$\begin{pmatrix} \boldsymbol{\theta}_i \\ \tilde{\boldsymbol{\alpha}}_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}\begin{pmatrix} \Sigma_{\theta\theta} & \Sigma_{\theta\alpha} \\ \Sigma_{\alpha\theta} & \Sigma_{\alpha\alpha} \end{pmatrix}\right].$$

$\boldsymbol{\theta}_i$ and $\boldsymbol{\alpha}_i$ are assumed to be independent of $\boldsymbol{\zeta}_{it}$ and $\boldsymbol{\epsilon}_{it}$. The distribution of the idiosyncratic error terms of the system of equations (2.1) (2.2), is given by:

$$\begin{pmatrix} \boldsymbol{\zeta}_{it} \\ \boldsymbol{\epsilon}_{it} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}\begin{pmatrix} \Sigma_{\zeta\zeta} & \Sigma_{\zeta\epsilon} \\ \Sigma_{\epsilon\zeta} & \Sigma_{\epsilon\epsilon} \end{pmatrix}\right].$$

Finally, while $\mathcal{Z}_i$ is correlated with $\boldsymbol{\theta}_i$ and $\boldsymbol{\alpha}_i$, conditional on $\boldsymbol{\theta}_i$ and $\boldsymbol{\alpha}_i$, $\mathcal{Z}_i$ is independent of $\boldsymbol{\zeta}_{it}$ and $\boldsymbol{\epsilon}_{it}$. Beyond assuming the above we do not place any restriction on the serial correlation among $\boldsymbol{\zeta}_{it}$ and $\boldsymbol{\epsilon}_{it}$.

To estimate the structural equations of the above model, given by the equations (2.1) we develop a two stage estimation procedure. In the first stage the system of reduced form equations, equation (2.2), is estimated. In the second stage, given the estimates of equation (2.2), equations in (2.1) are estimated jointly. In the second stage additional correction terms or "control variables", obtained from the first stage reduced form estimates, correct for the bias due to endogeneity of the $\mathbf{x}$. In the subsections to follow, where we study the identification of structural parameters for continuous and discrete response models, we show the construction of correction terms.

2.1 The First Stage: Maximum Likelihood Estimation of Reduced Form Equations

In the first stage of our econometric methodology we estimate the system of reduced form equations

$$\mathbf{x}_{it} = \mathbf{Z}'_{it}\boldsymbol{\beta} + \tilde{\boldsymbol{\alpha}}_i + \boldsymbol{\epsilon}_{it}, \tag{2.2}$$

where $\mathbf{x}_{it}$ is continuous. Since $\boldsymbol{\alpha}_i$ and $\mathcal{Z}_i$ are correlated in order to estimate $\boldsymbol{\delta}$, $\Sigma_{\epsilon\epsilon}$, and $\Sigma_{\alpha\alpha}$ consistently, we use Mundlak's correlated random effects (CRE) formulation. We assume that

$$\tilde{\boldsymbol{\alpha}}_i = \bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \boldsymbol{\alpha}_i, \tag{2.3}$$

where $\bar{\mathbf{Z}}_i = \text{diag}(\bar{\mathbf{z}}_{1i}, \ldots, \bar{\mathbf{z}}_{mi})$ and each of the $\bar{\mathbf{z}}_{li}$, $l \in \{1, \ldots, m\}$, are the mean of time varying variables in $\mathbf{z}_{lit} = \mathbf{z}_{it}$. Given the above, equation (2.2) can now be written as

$$\mathbf{x}_{it} = \mathbf{Z}'_{it}\boldsymbol{\beta} + \bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{it},$$

which to ease notations we can write as

$$\mathbf{x}_{it} = \mathbb{Z}'_{it}\boldsymbol{\delta} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{it}, \tag{2.2a}$$

where $\mathbb{Z}_{it} = \text{diag}((\mathbf{z}'_{it}, \bar{\mathbf{z}}'_i)', \ldots, (\mathbf{z}'_{it}, \bar{\mathbf{z}}'_i)')$, and define $\boldsymbol{\delta} = ((\boldsymbol{\beta}'_1, \bar{\boldsymbol{\delta}}'_1), \ldots, (\boldsymbol{\beta}'_m, \bar{\boldsymbol{\delta}}'_m))'$. In the modified reduced form equation (2.2a), $\boldsymbol{\alpha}_i$ and $\boldsymbol{\epsilon}_{it}$ are mutually uncorrelated, are independent of $\mathbb{Z}_{it}$, and

$$\begin{pmatrix} \boldsymbol{\theta}_i \\ \boldsymbol{\alpha}_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} \Sigma_{\theta\theta} & \Lambda_{\theta\alpha} \\ \Lambda_{\alpha\theta} & \Lambda_{\alpha\alpha} \end{pmatrix}\right].$$

The parameters, $\Theta_1 = \{\boldsymbol{\delta}, \Sigma_{\epsilon\epsilon}, \Lambda_{\alpha\alpha}\}$, of the modified equation (2.2a) can now be estimated by a step-wise maximum likelihood method for seemingly unrelated regression (SUR) developed by Biørn (2004). However, Biørn's paper does not account for any possible serial correlation among $\boldsymbol{\epsilon}_{it}$. For balanced panel Kobel (2004) provides a (SUR) estimator with first order serial correlation among $\boldsymbol{\epsilon}_{it}$. If $m = 1$, then one can employ the methodology in Baltagi and Li (1994) or Baltagi, Song, and Jung (2010). In what follows we will assume that there is no serial dependence among the idiosyncratic component, $\boldsymbol{\epsilon}_{it}$, and employ Biørn's methodology to estimate the reduced form equation (2.2a). In Appendix A we briefly describe the methodology in Biørn (2004).

2.2 Identification for Continuous Response Model

The identification strategy that allows us to construct the control variables that correct for the bias, which arises due to endogeneity of the regressors $\mathbf{x}$, is based on the following conditional mean restriction:

$$\mathrm{E}(\boldsymbol{\theta}_i + \boldsymbol{\zeta}_{it} | \mathbf{X}_i, \mathcal{Z}_i, \tilde{\boldsymbol{\alpha}}_i) = \mathrm{E}(\boldsymbol{\theta}_i + \boldsymbol{\zeta}_{it} | \mathcal{Z}_i, \boldsymbol{\epsilon}_i, \tilde{\boldsymbol{\alpha}}_i) = \mathrm{E}(\boldsymbol{\theta}_i + \boldsymbol{\zeta}_{it} | \boldsymbol{\epsilon}_i, \tilde{\boldsymbol{\alpha}}_i). \tag{2.4}$$

According to the above, the mean dependence of the composite structural error term $\boldsymbol{\theta}_i + \boldsymbol{\zeta}_{it}$ on the vector of regressors $\mathbf{X}_i$, $\mathcal{Z}_i$, and $\tilde{\boldsymbol{\alpha}}_i$ is completely characterized by the reduced form error vectors $\boldsymbol{\epsilon}_i$ and $\tilde{\boldsymbol{\alpha}}_i$. The expectation of $\boldsymbol{\theta} + \boldsymbol{\zeta}_{it}$ given $\tilde{\boldsymbol{\alpha}}_i$ and $\boldsymbol{\epsilon}_i$ is given by

$$\begin{aligned} E(\boldsymbol{\zeta}_{it} + \boldsymbol{\theta}_i | \tilde{\boldsymbol{\alpha}}_i, \boldsymbol{\epsilon}_i) &= E(\boldsymbol{\zeta}_{it} | \boldsymbol{\epsilon}_i) + E(\boldsymbol{\theta}_i | \tilde{\boldsymbol{\alpha}}_i) = E(\boldsymbol{\zeta}_{it} | \boldsymbol{\epsilon}_{it}) + E(\boldsymbol{\theta}_i | \tilde{\boldsymbol{\alpha}}_i) \\ &= \Sigma_{\zeta\epsilon} \Sigma_{\epsilon\epsilon}^{-1} \boldsymbol{\epsilon}_{it} + \Sigma_{\theta\alpha} \Sigma_{\alpha\alpha}^{-1} \tilde{\boldsymbol{\alpha}}_i \\ &= \tilde{\Sigma}_{\zeta\epsilon} \Sigma_{\epsilon} \Sigma_{\epsilon\epsilon}^{-1} \boldsymbol{\epsilon}_{it} + \bar{\Sigma}_{\theta\alpha} \tilde{\boldsymbol{\alpha}}_i \\ &= \tilde{\Sigma}_{\zeta\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \boldsymbol{\epsilon}_{it} + \bar{\Sigma}_{\theta\alpha} \tilde{\boldsymbol{\alpha}}_i, \end{aligned} \tag{2.5}$$

where the first equality follows from the fact that $\boldsymbol{\zeta}_{it}$ is independent of $\boldsymbol{\alpha}_i$ and $\boldsymbol{\theta}_i$ is independent of $\boldsymbol{\epsilon}_{it}$. The second equality follows from the assumption that conditional on ϵ_{it}, $\boldsymbol{\zeta}_{it}$ is independent of $\boldsymbol{\epsilon}_{i_{-t}}$. This assumption has also been made in Wooldridge (1995), Papke and Wooldridge (2008), and Semykina and Wooldridge (2010)[1]. The $(n \times m)$ matrices $\tilde{\Sigma}_{\zeta\epsilon}$ in the

[1]This assumption can be relaxed and one can specify the dependence of $\boldsymbol{\zeta}_{it}$ and $\boldsymbol{\epsilon}_i$ without adding any conceptual difficulties.

fourth equality is

$$\tilde{\Sigma}_{\zeta\epsilon} = \begin{pmatrix} \rho_{\zeta_1\epsilon_1}\sigma_{\zeta_1} & \cdots & \rho_{\zeta_1\epsilon_m}\sigma_{\zeta_1} \\ \vdots & & \vdots \\ \rho_{\zeta_n\epsilon_1}\sigma_{\zeta_n} & \cdots & \rho_{\zeta_n\epsilon_m}\sigma_{\zeta_n} \end{pmatrix}$$

and the $(m \times m)$ matrix Σ_ϵ is $\text{diag}(\sigma_{\epsilon 1}, \ldots, \sigma_{\epsilon m})$, so that $\tilde{\Sigma}_{\zeta\epsilon}\Sigma_\epsilon = \Sigma_{\zeta\epsilon}$. Finally, in the last equality $\tilde{\Sigma}_{\epsilon\epsilon}^{-1} = \Sigma_\epsilon \Sigma_{\epsilon\epsilon}^{-1}$. We prefer to write the above conditional expectation as $\text{E}(\boldsymbol{\zeta}_{it} + \boldsymbol{\theta}_i | \tilde{\boldsymbol{\alpha}}_i, \boldsymbol{\epsilon}_{it}) = \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it} + \tilde{\Sigma}_{\theta\alpha}\tilde{\Sigma}_{\alpha\alpha}^{-1}\tilde{\boldsymbol{\alpha}}_i$ since the elements of $\tilde{\Sigma}_{\epsilon\epsilon}^{-1}$ can be obtained from the estimates of the first stage reduced form estimation and, as we will see, it is the elements of $\tilde{\Sigma}_{\zeta\epsilon}$ and $\bar{\Sigma}_{\theta\alpha}$, which are estimated in the second stage structural estimation, that give us a potential test of the exogeneity of $\mathbf{x}_{it}$ with respect to $\boldsymbol{\zeta}_{it}$ and $\tilde{\boldsymbol{\theta}}_i$.

The condition in (2.4) then implies that the conditional distribution of $\mathbf{y}_{it}^*$ given $\mathbf{X}_i$, $\mathcal{Z}_i$, and $\tilde{\boldsymbol{\alpha}}_i$ is given by

$$\begin{aligned} \text{E}(\mathbf{y}_{it}^* | \mathbf{X}_i, \mathcal{Z}_i, \tilde{\boldsymbol{\alpha}}_i) &= \mathbf{Z}_{it}^{y\prime}\boldsymbol{\varphi} + \mathbf{X}_{it}'\tilde{\boldsymbol{\varphi}} + \bar{\Sigma}_{\theta\alpha}\tilde{\boldsymbol{\alpha}}_i + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it} \\ &= \mathbf{Z}_{it}^{y\prime}\boldsymbol{\varphi} + \mathbf{X}_{it}'\tilde{\boldsymbol{\varphi}} + \bar{\Sigma}_{\theta\alpha}(\bar{\mathbf{Z}}_i'\bar{\boldsymbol{\delta}} + \boldsymbol{\alpha}_i) + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it} = \text{E}(\mathbf{y}_{it}^* | \mathbf{X}_i, \mathcal{Z}_i, \boldsymbol{\alpha}_i) \end{aligned} \tag{2.6}$$

Given (2.5), the linear projections of $\boldsymbol{\theta}_i + \boldsymbol{\zeta}_{it}$ in error form, given $\tilde{\boldsymbol{\alpha}}_i$ and $\boldsymbol{\epsilon}_{it}$, can be written as

$$\boldsymbol{\theta}_i + \boldsymbol{\zeta}_{it} = \bar{\Sigma}_{\theta\alpha}(\bar{\mathbf{Z}}_i'\bar{\boldsymbol{\delta}} + \boldsymbol{\alpha}_i) + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it} + \bar{\boldsymbol{\theta}}_i + \bar{\boldsymbol{\zeta}}_{it} \tag{2.7}$$

where $\bar{\boldsymbol{\theta}}_i$ and $\bar{\boldsymbol{\zeta}}_{it}$ are both normally distributed with mean 0 and are independent of $\mathcal{Z}_i$, $\mathbf{X}_i$, $\boldsymbol{\epsilon}_{it}$ and $\boldsymbol{\alpha}_i$. The above then implies that the projections of $\mathbf{y}_{it}^*$ in error form given $\boldsymbol{\alpha}_i$ and $\boldsymbol{\epsilon}_{it}$ is given by

$$\mathbf{y}_{it}^* = \mathbf{Z}_{it}^{y\prime}\boldsymbol{\varphi} + \mathbf{X}_{it}'\tilde{\boldsymbol{\varphi}} + \bar{\Sigma}_{\theta\alpha}(\bar{\mathbf{Z}}_i'\bar{\boldsymbol{\delta}} + \boldsymbol{\alpha}_i) + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it} + \bar{\boldsymbol{\theta}}_i + \bar{\boldsymbol{\zeta}}_{it} \tag{2.8}$$

To estimate the above system of equation, the standard technique is to replace $\boldsymbol{\epsilon}_{it}$ by the residuals from the first stage reduced form regression. However, the residuals, $\mathbf{x}_{it} - \text{E}(\mathbf{x}_{it} | \mathcal{Z}_i, \boldsymbol{\alpha}_i) = \mathbf{x}_{it} - \mathbb{Z}_{it}'\boldsymbol{\delta} - \boldsymbol{\alpha}_i$, are not identified because the $\boldsymbol{\alpha}_i$'s are unobserved, even though $\boldsymbol{\delta}$, $\Lambda_{\alpha\alpha}$ and $\Sigma_{\epsilon\epsilon}$ are consistently estimated in the first stage estimation of the reduced form equation (2.2a). It could be possible to estimate the structural parameters in (2.8) if we could integrate out $\boldsymbol{\alpha}_i$ with respect to its conditional distribution $\mathbf{f}(\boldsymbol{\alpha}_i | \mathbf{X}_i, \mathcal{Z}_i)$. To see this, consider

$\mathrm{E}(y^*_{it}|\mathbf{X}_i, \mathcal{Z}_i, \boldsymbol{\alpha}_i)$ in (2.6)

$$
\begin{aligned}
E(y^*_{it}|\mathbf{X}_i, \mathcal{Z}_i) &= \int \mathrm{E}(y^*_{it}|\mathbf{X}_i, \mathcal{Z}_i, \boldsymbol{\alpha}_i)\mathbf{f}(\boldsymbol{\alpha}_i|\mathbf{X}_i, \mathcal{Z}_i)d\boldsymbol{\alpha}_i \\
&= \mathbf{Z}^{y\prime}_{it}\boldsymbol{\varphi} + \mathbf{X}'_{it}\tilde{\varphi} + \bar{\Sigma}_{\theta\alpha}\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}^{-1}_{\epsilon\epsilon}(\mathbf{x}_{it} - \mathbf{Z}'_{it}\boldsymbol{\delta}) + \int(\bar{\Sigma}_{\theta\alpha}\boldsymbol{\alpha}_i - \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\boldsymbol{\alpha}_i)\mathbf{f}(\boldsymbol{\alpha}_i|\mathbf{X}_i, \mathcal{Z}_i)d\boldsymbol{\alpha}_i \\
&= \mathbf{Z}^{y\prime}_{it}\boldsymbol{\varphi} + \mathbf{X}'_{it}\tilde{\varphi} + \bar{\Sigma}_{\theta\alpha}\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}^{-1}_{\epsilon\epsilon}(\mathbf{x}_{it} - \mathbf{Z}'_{it}\boldsymbol{\delta}) + \int(\bar{\Sigma}_{\theta\alpha}\boldsymbol{\alpha}_i - \bar{\Sigma}_{\zeta\epsilon}\boldsymbol{\alpha}_i)\mathbf{f}(\boldsymbol{\alpha}_i|\mathbf{X}_i)d\boldsymbol{\alpha}_i \\
&= \mathbf{Z}^{y\prime}_{it}\boldsymbol{\varphi} + \mathbf{X}'_{it}\tilde{\varphi} + \bar{\Sigma}_{\theta\alpha}\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}^{-1}_{\epsilon\epsilon}(\mathbf{x}_{it} - \mathbf{Z}'_{it}\boldsymbol{\delta}) + \bar{\Sigma}_{\theta\alpha}\hat{\boldsymbol{\alpha}}_i - \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\alpha}}_i \\
&= \mathbf{Z}^{y\prime}_{it}\boldsymbol{\varphi} + \mathbf{X}'_{it}\tilde{\varphi} + \bar{\Sigma}_{\theta\alpha}(\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i) + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}^{-1}_{\epsilon\epsilon}(\mathbf{x}_{it} - \mathbf{Z}'_{it}\boldsymbol{\delta} - \hat{\boldsymbol{\alpha}}_i) \\
&= \mathbf{Z}^{y\prime}_{it}\boldsymbol{\varphi} + \mathbf{X}'_{it}\tilde{\varphi} + \bar{\Sigma}_{\theta\alpha}(\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i) + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\epsilon}}_{it}, \qquad (2.9)
\end{aligned}
$$

where the second equality follows from the fact that $\mathcal{Z}_i$ and $\boldsymbol{\alpha}_i$ are independent. $\hat{\boldsymbol{\alpha}}_i = \hat{\boldsymbol{\alpha}}_i(\mathbf{X}_i, \mathcal{Z}_i, \Theta_1)$ and $\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\alpha}}_i = \tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\alpha}}_i(\mathbf{X}_i, \mathcal{Z}_i, \Theta_1)$ are the expected a posteriori (EAP) values of the functions of time invariant individual effects $\boldsymbol{\alpha}_i$.

To obtain (2.9), using Bayes rule we can write $\mathbf{f}(\boldsymbol{\alpha}|\mathbf{X}, \mathcal{Z})$ as

$$\mathbf{f}(\boldsymbol{\alpha}|\mathbf{X}) = \frac{\mathbf{f}(\mathbf{X}|\boldsymbol{\alpha})\mathbf{g}(\boldsymbol{\alpha})}{\mathbf{h}(\mathbf{X})} = \frac{\mathbf{f}(\mathbf{X}, \mathcal{Z}|\boldsymbol{\alpha})\mathbf{g}(\boldsymbol{\alpha})}{\mathbf{h}(\mathbf{X}, \mathcal{Z})}, \qquad (2.10)$$

where $\mathbf{g}$ and $\mathbf{h}$ are density functions. The above can be written as

$$\frac{\mathbf{f}(\mathbf{X}, \mathcal{Z}|\boldsymbol{\alpha})\mathbf{g}(\boldsymbol{\alpha})}{\mathbf{h}(\mathbf{X}, \mathcal{Z})} = \frac{\mathbf{f}(\mathbf{X}|\mathcal{Z}, \boldsymbol{\alpha})\mathbf{p}(\mathcal{Z}|\boldsymbol{\alpha})\mathbf{g}(\boldsymbol{\alpha})}{\mathbf{h}(\mathbf{X}|\mathcal{Z})\mathbf{p}(\mathcal{Z})},$$

By our assumption the time invariant individual effects, $\boldsymbol{\alpha}$, are independent of the exogenous variables $\mathcal{Z}$, hence $\mathbf{p}(\mathcal{Z}|\boldsymbol{\alpha}) = \mathbf{p}(\mathcal{Z})$, that is,

$$\mathbf{f}(\boldsymbol{\alpha}|\mathbf{X}) = \frac{\mathbf{f}(\mathbf{X}|\mathcal{Z}, \boldsymbol{\alpha})\mathbf{g}(\boldsymbol{\alpha})}{\mathbf{h}(\mathbf{X}|\mathcal{Z})} = \frac{\mathbf{f}(\mathbf{X}|\mathcal{Z}, \boldsymbol{\alpha})\mathbf{g}(\boldsymbol{\alpha})}{\int \mathbf{f}(\mathbf{X}|\mathcal{Z}, \boldsymbol{\alpha})\mathbf{g}(\boldsymbol{\alpha})d\boldsymbol{\alpha}}, \qquad (2.11)$$

Hence,

$$
\begin{aligned}
\int \tilde{\Sigma}^{-1}_{\epsilon\epsilon}\boldsymbol{\alpha}\mathbf{f}(\boldsymbol{\alpha}|\mathbf{X})d(\boldsymbol{\alpha}) &= \int \frac{\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\boldsymbol{\alpha}\mathbf{f}(\mathbf{X}|\mathcal{Z}, \alpha)\mathbf{g}(\boldsymbol{\alpha})d\boldsymbol{\alpha}}{\int \mathbf{f}(\mathbf{X}|\mathcal{Z}, \boldsymbol{\alpha})\mathbf{g}(\boldsymbol{\alpha})d\boldsymbol{\alpha}} \\
&= \frac{\int \tilde{\Sigma}^{-1}_{\epsilon\epsilon}\boldsymbol{\alpha}_i \prod_{t=1}^{T} \mathbf{f}(\mathbf{x}_t|\mathcal{Z}, \boldsymbol{\alpha})\mathbf{g}(\boldsymbol{\alpha})d\boldsymbol{\alpha}}{\int \prod_{t=1}^{T} \mathbf{f}(\mathbf{x}_t|\mathcal{Z}, \boldsymbol{\alpha})\mathbf{g}(\boldsymbol{\alpha})d\boldsymbol{\alpha}} \\
&= \tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\alpha}}(\mathbf{X}, \mathcal{Z}, \boldsymbol{\delta}, \Sigma_{\epsilon\epsilon}, \Lambda_{\alpha\alpha}) \qquad (2.12)
\end{aligned}
$$

where the second equality follow from the fact that conditional on $\mathcal{Z}$ and $\boldsymbol{\alpha}$, each of the $\mathbf{x}_t$, $\mathbf{x}_t \in \{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$ are independently normally distributed with mean $\mathbf{Z}'_t\boldsymbol{\delta} + \boldsymbol{\alpha}$ and standard deviation $\Sigma_{\epsilon\epsilon}$. $\mathbf{g}(\boldsymbol{\alpha})$ by our assumption is normally distributed with mean zero and variance

$\Lambda_{\alpha\alpha}$. Similarly we can obtain $\hat{\boldsymbol{\alpha}}(\mathbf{X}, \mathcal{Z}, \Theta_1)$. The functional form of $\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}_i(\mathbf{X}_i, \mathcal{Z}_i, \Theta_1)$ is given by:

$$
\begin{aligned}
&\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}_i(\mathbf{X}_i, \mathcal{Z}_i, \Theta_1) \\
&= \frac{\int \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\alpha} \prod_{t=1}^{T} \frac{1}{(2\pi)^{m/2}|\Sigma_{\epsilon\epsilon}|^{1/2}} \exp(-\frac{1}{2}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - \boldsymbol{\alpha})^T \Sigma_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - \boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha}}{\int \prod_{t=1}^{T} \frac{1}{(2\pi)^{m/2}|\Sigma_{\epsilon\epsilon}|^{1/2}} \exp(-\frac{1}{2}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - \boldsymbol{\alpha})^T \Sigma_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - \boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha}} \\
&= \frac{\int \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\alpha} \exp(-\frac{1}{2}\sum_{t=1}^{T}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - \boldsymbol{\alpha})'\Sigma_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - \boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha}}{\int \exp(-\frac{1}{2}\sum_{t=1}^{T}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - \boldsymbol{\alpha})'\Sigma_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - \boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha}} \\
&= \frac{\int \tilde{\Sigma}_{\epsilon\epsilon}^{-1}C\mathfrak{a} \exp(-\frac{1}{2}\sum_{t=1}^{T}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - \boldsymbol{\alpha})'\Sigma_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - C\mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a}}{\int \exp(-\frac{1}{2}\sum_{t=1}^{T}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - C\mathfrak{a})'\Sigma_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - C\mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a}} \qquad (2.13)
\end{aligned}
$$

where $\boldsymbol{\alpha} = C\mathfrak{a}$, CC' being the Cholesky decomposition of the covariance matrix $\Lambda_{\alpha\alpha}$, so that $d\boldsymbol{\alpha} = |C|d\mathfrak{a} = |\Lambda_{\alpha\alpha}|^{1/2}d\mathfrak{a}$. $\hat{\hat{\boldsymbol{\alpha}}}(\mathbf{X}_i, \mathcal{Z}_i.\hat{\Theta}_1)$ and $\hat{\tilde{\Sigma}}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}(\mathbf{X}_i, \mathcal{Z}_i, \hat{\Theta}_1)$, the estimated expected a posteriori value of $\boldsymbol{\alpha}$ and $\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\alpha}$ respectively, can be estimated by employing numerical integration techniques with respect to $\mathfrak{a}$ at the estimated values $\hat{\delta}$, $\hat{\Sigma}_{\epsilon\epsilon}$, and $\hat{\Lambda}_{\alpha\alpha}$. In Appendix D we provide a note on numerical technique employed to estimate $\hat{\hat{\boldsymbol{\alpha}}}()$ and $\hat{\tilde{\Sigma}}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}()$. Also, it can be shown that

Lemma 1 $\hat{\hat{\boldsymbol{\alpha}}}_i(\boldsymbol{X}_i, \mathcal{Z}_i, \hat{\Theta}_1)$ *and* $\hat{\tilde{\Sigma}}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}_i(\boldsymbol{X}_i, \mathcal{Z}_i, \hat{\Theta}_1)$ *converges a.s. to* $\hat{\boldsymbol{\alpha}}_i(\boldsymbol{X}_i, \mathcal{Z}_i, \Theta_1)$ *and* $\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}_i(\boldsymbol{X}_i, \mathcal{Z}_i, \Theta_1)$ *respectively, where* $\hat{\Theta}_1 = \{\hat{\boldsymbol{\delta}}, \hat{\Sigma}_{\epsilon\epsilon}, \hat{\Lambda}_{\alpha\alpha}\}$ *are consistent first stage estimates.*

Proof of Lemma 1 Given in Appendix B .

If population parameters, $\boldsymbol{\delta}$, $\Sigma_{\epsilon\epsilon}$, and $\Lambda_{\alpha\alpha}$, were known, the above implies that we could write the linear predictor of y_{it}^*, given $\mathbf{X}_i$ and $\mathcal{Z}_i$ in error form as

$$
\mathbf{y}_{it}^* = \mathbf{Z}_{it}^{y\prime}\boldsymbol{\varphi} + \mathbf{X}_{it}'\tilde{\varphi} + \bar{\Sigma}_{\theta\alpha}(\bar{\mathbf{Z}}'\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i) + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it} + \tilde{\boldsymbol{\theta}}_i + \tilde{\boldsymbol{\zeta}}_{it}, \qquad (2.14)
$$

where we assume that $\tilde{\boldsymbol{\theta}}_i$ and $\tilde{\boldsymbol{\zeta}}_{it}$ are distributed with mean 0 and with variance $\Sigma_{\tilde{\theta}\tilde{\theta}}$ and $\Sigma_{\tilde{\zeta}\tilde{\zeta}}$ respectively, and are independent of $\mathcal{Z}_i$, $\mathbf{X}_i$. With estimates $\hat{\hat{\boldsymbol{\alpha}}}_i$ and $\hat{\tilde{\Sigma}}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it} = \tilde{\Sigma}_{\epsilon\epsilon}^{-1}(\mathbf{x}_{it} - \mathbf{Z}_{it}'\hat{\boldsymbol{\delta}} - \hat{\hat{\boldsymbol{\alpha}}}_i)$ in place, the system of equations in (2.14) can now be estimated as seemingly unrelated regression (SUR). A panel version of SUR can be employed to gain efficiency.

We note here that for any k, $k \in \{1, \ldots, n\}$, $\tilde{\Sigma}_{\zeta_k\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}$ and $\bar{\Sigma}_{\theta_k\alpha}\hat{\boldsymbol{\alpha}}_i$ in (2.14) take the form

$$
\rho_{\zeta_k\epsilon_1}\sigma_{\zeta_k}f_1(\Sigma_{\epsilon\epsilon}, \hat{\epsilon}_{1it}, \ldots, \hat{\epsilon}_{mit}) + \ldots + \rho_{\zeta_k\epsilon_m}\sigma_{\zeta_k}f_m(\Sigma_{\epsilon\epsilon}, \hat{\epsilon}_{1it}, \ldots, \hat{\epsilon}_{mit})
$$

and

$$\bar{\rho}_{\theta_k\alpha_1}\hat{\alpha}_{1i} + \ldots + \bar{\rho}_{\theta_k\alpha_m}\sigma_{\theta_k}\hat{\alpha}_{mi}$$

respectively, and where each of the f's above are linear in $\hat{\boldsymbol{\epsilon}}_{it}$. The estimates $\rho_{\zeta_k\epsilon_l}\sigma_{\zeta_k}$, $l \in \{1,\ldots,m\}$, provides us with a test of exogeneity of the regressor x_l with respect to ζ_k and the estimates $\bar{\rho}_{\theta_k\alpha_l}$ provides us with test of exogeniety of x_l with respect to θ_k.

2.3 Identification for Discrete Response Models

Let $n = 1$ and suppose that y_{it} is a binary variable, which takes value 0 or 1. Let y_t^* be the latent variable underlying y_t, whose DGP is given by

$$y_{it}^* = \mathbf{z}_{it}^{y\prime}\varphi + \mathbf{x}_{it}'\tilde{\varphi} + \theta_i + \zeta_{it}. \tag{2.15}$$

To ease notations we let $\mathcal{X} = \{\mathbf{z}^y \cup \mathbf{x}\}$ and $\boldsymbol{\varphi} = \{\varphi \cup \tilde{\varphi}\}$, then given the reduced form population parameters, Θ_1, the linear predictor of y_{it}^*, given $\mathbf{X}_i$ and $\mathcal{Z}_i$, in error form as shown in (2.14) is

$$y_{it}^* = \mathcal{X}_{it}'\boldsymbol{\varphi} + \bar{\Sigma}_{\theta\alpha}(\bar{\mathbf{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i) + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it} + \tilde{\theta}_i + \tilde{\zeta}_{it}, \tag{2.16}$$

where $\tilde{\theta}_i$ and $\tilde{\zeta}_{it}$ are i.i.d. and normally distributed with mean 0 and variance $\sigma_{\tilde{\theta}}^2$ and $\sigma_{\tilde{\zeta}}^2$. From the fact that in probit models the parameters are identified only up to a scale, for an individual i, the probability of $y_t = 1$ given $\mathbf{X}_i$ and $\mathcal{Z}_i$ is given by

$$\begin{aligned}\Pr(y_t = 1|\mathbf{X}, \mathcal{Z}) &= \Pr(y_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}) = \Pr(y_t^* > 0|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}) \\ &= \Phi\bigg(\{\mathcal{X}_t'\boldsymbol{\varphi} + \bar{\Sigma}_{\theta\alpha}(\bar{\mathbf{Z}}'\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}) + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t\}.(\tilde{\sigma}_{\theta}^2 + \tilde{\sigma}_{\zeta}^2)^{-1/2}\bigg),\end{aligned} \tag{2.17}$$

where the first equality follows from the fact that $\hat{\boldsymbol{\epsilon}}_t$ and $\hat{\boldsymbol{\alpha}}$ is a function of $\mathbf{X}$ and $\mathcal{Z}$. $\Phi()$ is the cumulative standard normal density function. However, $\Pr(y_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}) = \Pr(y_t = 1|\mathbf{X}, \mathcal{Z}, \mathrm{E}(\boldsymbol{\epsilon}_t|\mathbf{X}, \mathcal{Z}), \mathrm{E}(\boldsymbol{\alpha}|\mathbf{X}, \mathcal{Z}))$ is generally not equal to $\Pr(y_t = 1|\mathbf{X}, \mathcal{Z}, \boldsymbol{\epsilon}_t, \boldsymbol{\alpha})$. Our measure of interest, however, is $\int \Pr(y_t = 1|\mathbf{X}, \mathcal{Z}, \boldsymbol{\epsilon}_t, \boldsymbol{\alpha})dg(\boldsymbol{\epsilon}_t, \boldsymbol{\alpha})$, the average structural function, (ASF), and the average partial effect (APE) of changing a variable, say z_k, in time period t from z_{kt} to $z_{kt} + \Delta_{z_k}$, given by

$$\begin{aligned}\frac{\Delta \Pr(y_t = 1)}{\Delta_{z_k}} =& \bigg[\int \Pr(y_t = 1|\mathbf{X}, \mathcal{Z}_{-1}, z_{k_{-t}}, (z_{kt} + \Delta_{z_k}), \boldsymbol{\epsilon}_t, \boldsymbol{\alpha})dg(\boldsymbol{\epsilon}_t, \boldsymbol{\alpha}) \\ &- \int \Pr(y_t = 1|\mathbf{X}, \mathcal{Z}_{-1}, z_{k_{-t}}, z_{kt}, \boldsymbol{\epsilon}_t, \boldsymbol{\alpha})dg(\boldsymbol{\epsilon}_t, \boldsymbol{\alpha})\bigg]/\Delta_{z_k},\end{aligned} \tag{2.18}$$

where $g(\boldsymbol{\epsilon}_t, \boldsymbol{\alpha})$ is the joint distribution function of $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\alpha}$. To recover the above measure in (2.18), like Chamberlain (1984), we make an assumption about the conditional distribution for $\tilde{\boldsymbol{\alpha}}_i$. We assume that

$$\begin{aligned}\tilde{\boldsymbol{\alpha}}_i = E(\tilde{\boldsymbol{\alpha}}_i|\mathbf{X}_i, \mathcal{Z}_i) + \check{\boldsymbol{\alpha}}_i = E(\bar{\mathbf{Z}}_i'\boldsymbol{\delta} + \boldsymbol{\alpha}_i|\mathbf{X}_i, \mathcal{Z}_i) + \check{\boldsymbol{\alpha}}_i &= \bar{\mathbf{Z}}_i'\bar{\boldsymbol{\delta}} + E(\boldsymbol{\alpha}_i|\mathbf{X}_i, \mathcal{Z}_i) + \check{\boldsymbol{\alpha}}_i \\ &= \bar{\mathbf{Z}}_i'\boldsymbol{\delta} + \hat{\boldsymbol{\alpha}}_i + \check{\boldsymbol{\alpha}}_i, \end{aligned} \tag{2.19}$$

where $\check{\boldsymbol{\alpha}}_i$ is normally distributed with mean 0, variance $\Sigma_{\check{\alpha}\check{\alpha}}$, is independent of everything else, and $\hat{\boldsymbol{\alpha}}_i$, as we have shown above, is

$$\mathrm{E}(\boldsymbol{\alpha}|\mathbf{X}) = \int \boldsymbol{\alpha} f(\boldsymbol{\alpha}|\mathbf{X}) d(\boldsymbol{\alpha}) = \hat{\boldsymbol{\alpha}}(\mathbf{X}, \mathcal{Z}, \Theta_1).$$

The above implies that, conditional on $\mathbf{X}$ and $\mathcal{Z}$, $\boldsymbol{\epsilon}_t$ is distributed as

$$\boldsymbol{\epsilon}_t = \mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - \hat{\boldsymbol{\alpha}} - \check{\boldsymbol{\alpha}} = \hat{\boldsymbol{\epsilon}}_t - \check{\boldsymbol{\alpha}},$$

Hence, under the assumption about the conditional distribution of $\boldsymbol{\alpha}_i$, we can write (2.16) as

$$\begin{aligned} y_t^* &= \mathcal{X}_t'\boldsymbol{\varphi} + \bar{\Sigma}_{\theta\alpha}(\bar{\mathbf{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}) + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + (\bar{\Sigma}_{\theta\alpha} - \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1})\check{\boldsymbol{\alpha}} + \bar{\zeta}_t \\ &= \mathcal{X}_t'\boldsymbol{\varphi} + \bar{\Sigma}_{\theta\alpha}(\bar{\mathbf{Z}}_i'\boldsymbol{\delta} + \hat{\boldsymbol{\alpha}}) + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + \bar{\alpha} + \bar{\zeta}_t, \end{aligned} \tag{2.20}$$

where $\bar{\alpha} = (\bar{\Sigma}_{\theta\alpha} - \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1})\check{\boldsymbol{\alpha}}$ is uncorrelated with any of the covariates. Since $\bar{\alpha}$ is a linear combination of the elements in $\check{\boldsymbol{\alpha}}$, $\bar{\alpha}$ is also normally distributed with a variance, say, $\sigma_{\bar{\alpha}}^2$, and $\sigma_{\bar{\zeta}}^2$ the variance of $\bar{\zeta}$, which is normally distributed with mean 0 and uncorrelated with any of the covariates.

Now, having assumed the conditional distribution of $\boldsymbol{\alpha}_i$, for any individual i, we now have

$$\Pr(y_t = 1|\mathbf{X}, \mathcal{Z}, \boldsymbol{\epsilon}_t, \boldsymbol{\alpha}) = \Pr(y_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}, \bar{\alpha})$$

and

$$\int \Pr(y_t = 1|\mathbf{X}, \mathcal{Z}, \boldsymbol{\epsilon}_t, \boldsymbol{\alpha}) dg(\boldsymbol{\epsilon}_t, \boldsymbol{\alpha}) = \int \Pr(y_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}, \bar{\alpha}) dF(\hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}, \bar{\alpha}),$$

where $F(\hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}, \bar{\alpha})$ is the joint distribution function of the arguments. Now

$$\begin{aligned} \int \Pr(y_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\alpha}}, \bar{\alpha}) dF(\hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}, \bar{\alpha}) &= \int\int \Pr(y_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}, \bar{\alpha}) h(\bar{\alpha}|\hat{\boldsymbol{\alpha}}) d\bar{\alpha} dG(\hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}) \\ &= \int\int \Pr(y_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}, \bar{\alpha},) h(\bar{\alpha}) d\bar{\alpha} dG(\hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}) \\ &= \int \Pr(y_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}) dG(\hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}), \end{aligned} \tag{2.21}$$

where $G(\hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}})$ is the distribution of $\hat{\boldsymbol{\epsilon}}_t$ and $\hat{\boldsymbol{\alpha}}$, and $h(\bar{\alpha}|\hat{\boldsymbol{\alpha}})$ is the conditional distribution of $\bar{\alpha}$ given $\boldsymbol{\epsilon}_t$ and $\hat{\boldsymbol{\alpha}}$. The second equality above follows from the fact that $\hat{\boldsymbol{\epsilon}}_t$ and $\hat{\boldsymbol{\alpha}}$ is independent of $\bar{\alpha}$. Thus we have shown that

$$\begin{aligned}
&\int \Pr(y_t = 1|\mathbf{X}, \mathcal{Z}, \boldsymbol{\epsilon}_t, \boldsymbol{\alpha}) dg(\boldsymbol{\epsilon}_t, \boldsymbol{\alpha}) = \int \Pr(y_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}) dG(\hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}) \\
&= \int \Big[\int \Phi\bigg(\Big\{ \mathcal{X}_t' \boldsymbol{\varphi} + \bar{\Sigma}_{\theta\alpha} (\bar{\mathbf{Z}}' \bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}) + \tilde{\Sigma}_{\zeta\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_t + \bar{\alpha} \Big\} \frac{1}{\sigma_{\bar{\zeta}}} \bigg) h(\bar{\alpha}) d\bar{\alpha} \Big] dG(\hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}) \\
&= \int \Phi\bigg(\Big\{ \mathcal{X}_t' \boldsymbol{\varphi} + \bar{\Sigma}_{\theta\alpha} (\bar{\mathbf{Z}}' \bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}) + \tilde{\Sigma}_{\zeta\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_t \Big\} \frac{1}{(\sigma_{\bar{\zeta}}^2 + \sigma_{\bar{\alpha}}^2)^{1/2}} \bigg) dG(\hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}) \\
&= \int \Phi\bigg(\mathcal{X}_t' \varphi + \bar{\Sigma}_{\theta\alpha} (\bar{\mathbf{Z}}' \bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}) + \tilde{\Sigma}_{\zeta\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_t \bigg) dG(\hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}) \qquad (2.22)
\end{aligned}$$

where, with a slight abuse of notation, we denote $\boldsymbol{\varphi}$, $\bar{\Sigma}_{\theta\alpha}$, and $\tilde{\Sigma}_{\zeta\epsilon}$ as the scaled vector of coefficients, the scaling factor being $\frac{1}{(\sigma_{\bar{\zeta}}^2 + \sigma_{\bar{\alpha}}^2)^{1/2}}$. The coefficients $\boldsymbol{\varphi}$, $\bar{\Sigma}_{\theta\alpha}$, and $\tilde{\Sigma}_{\zeta\epsilon}$ can be obtained by simply running a pooled probit regression. While pooled probit consistently estimates the scaled vector of coefficients, it is likely to be inefficient. It is possible to estimate the parameters more efficiently than pooled probit that is still consistent under the same set of assumptions. One possibility is minimum distance estimation. That is, estimate a separate models for each t, and then impose the restrictions using minimum distance methods.

To obtain the sample analog of RHS of (2.22) for any fixed $\mathcal{X}_{it} = \bar{\mathcal{X}}$ we can compute

$$\frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \Pr(y_{it} = 1|\bar{\mathcal{X}}, \hat{\hat{\boldsymbol{\epsilon}}}_{it}, \hat{\hat{\alpha}}_i) = \frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \Phi\bigg(\bar{\mathcal{X}}' \hat{\boldsymbol{\varphi}} + \hat{\bar{\Sigma}}_{\theta\alpha} (\bar{\mathbf{Z}}_i' \hat{\bar{\boldsymbol{\delta}}} + \hat{\hat{\boldsymbol{\alpha}}}_i) + \hat{\tilde{\Sigma}}_{\zeta\epsilon} \hat{\tilde{\Sigma}}_{\epsilon\epsilon}^{-1} \hat{\hat{\boldsymbol{\epsilon}}}_{it} \bigg), \qquad (2.23)$$

which will converge to $\int \Pr(y_{it} = 1|\bar{\mathcal{X}}, \hat{\boldsymbol{\epsilon}}_{it}, \hat{\boldsymbol{\alpha}}_i) dG(\hat{\boldsymbol{\epsilon}}_{it}, \hat{\boldsymbol{\alpha}}_i)$ in probability as $\sum_{i=1}^N T_i = \mathcal{N} \to \infty$. The coefficients indicated above are the optimal first stage reduced form and second stage structural estimates. With (2.23) we can now compute (2.18), the mean effect or the average partial effect (APE), of changing a variable, say, w_t, where w_t is an element of either $\mathbf{x}_t$ or $\mathbf{z}_t^y$, from w_t to $w_t + \Delta_w$. In the limit when Δ_w tends to zero, and since the integrand is a smooth function of its arguments we can change the order of differentiation and integration in (2.18) to get

$$\frac{\partial \Pr(y_t = 1)}{\partial w} = \int \varphi_w \phi\bigg(\bar{\mathcal{X}}' \boldsymbol{\varphi} + \bar{\Sigma}_{\theta\alpha} (\bar{\mathbf{Z}}' \bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}) + \tilde{\Sigma}_{\zeta\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_t \bigg) dG(\hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}), \qquad (2.24)$$

where ϕ is the density function of a standard normal. Then, for any fixed $\mathcal{X}_{it} = \bar{\mathcal{X}}$, an estimate of the APE of w, the sample analog of the RHS in (2.24), can be computed as follows:

$$\frac{\partial \widehat{\Pr}(y_{it}=1)}{\partial w} = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{\varphi}_w \phi\bigg(\bar{\mathcal{X}}' \hat{\boldsymbol{\varphi}} + \hat{\tilde{\Sigma}}_{\theta\alpha} (\bar{\mathbf{Z}}_i' \hat{\tilde{\boldsymbol{\delta}}} + \hat{\hat{\boldsymbol{\alpha}}}_i) + \hat{\tilde{\Sigma}}_{\zeta\epsilon} \hat{\tilde{\Sigma}}_{\epsilon\epsilon}^{-1} \hat{\hat{\boldsymbol{\epsilon}}}_{it} \bigg), \tag{2.25}$$

which converges in probability to its true value in (2.24) as $\sum_{i=1}^{N} T_i = \mathcal{N} \to \infty$.

Suppose, w is dummy variable taking values 0 and 1, then the APE of change of w_{it} from 0 to 1, at population parameters, on the probability of $y_{it} = 1$, given other covariates, is given by

$$\begin{aligned} \int \bigg[\Phi\bigg(\bar{\mathcal{X}}_{-w}^{f\prime} \boldsymbol{\varphi}_{-w} + \varphi_w + \bar{\Sigma}_{\theta\alpha} (\bar{\mathbf{Z}}' \bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}) + \tilde{\Sigma}_{\zeta\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_t \bigg) \\ - \Phi\bigg(\bar{\mathcal{X}}_{-w}^{f\prime} \boldsymbol{\varphi}_{-w} + \bar{\Sigma}_{\theta\alpha} (\bar{\mathbf{Z}}' \bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}) + \tilde{\Sigma}_{\zeta\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_t \bigg) \bigg] dG(\hat{\boldsymbol{\epsilon}}_t, \hat{\boldsymbol{\alpha}}), \end{aligned} \tag{2.26}$$

whose sample analog, given $\mathcal{X}_{it_{-w}} = \bar{\mathcal{X}}_{_{-w}}$, the estimated first and second stage coefficients, $\hat{\hat{\boldsymbol{\epsilon}}}_{it}$, and $\hat{\hat{\boldsymbol{\alpha}}}_i$, can be computed by employing (2.23) for $w = 1$ and $w = 0$.

Finally, we would like to add that though we have elucidate our methodology for binary response model, it can be applied to other nonlinear models such as tobit or bivariate probit.

3 Concluding Remarks

The primary objective of the paper has been to come up with an estimator for nonlinear models of panel data that takes the unobserved heterogeneity, its correlation with the regressors, and the endogeneity of a subset of regressors that are correlated with unobserved heterogeneity and the idiosyncratic component into account. "Average partial effects", (APE), a measure that is important to measure the effectiveness of policy initiative has also been constructed. To achieve the above mentioned end we combined the methodology of "correlated random effect" (CRE) with the "control function" (CF) approach to come up with control functions that correct for bias that can arise due some regressors being correlated with the unobserved heterogeneity components as well as the idiosyncratic component. The control functions that we construct are based on "expected a posteriori" (EAP) values of unobserved heterogeneity/individual effects that appear as conditioning variables in the structural equation and which are correlated with the exogenous as well as endogenous variables. To compute the EAP values of the individual effects numerical integration with respect to the estimated – estimated using the first stage reduced form estimates – posterior distribution is performed.

Since the EAP values are functions of the endogenous as well as exogenous regressors, with EAP values of individual effects substituted for the individual effects, the correlation between the covariates and the unobserved individual effects are accounted for.

While we have developed an estimator for nonlinear panel data and constructed a measure of APE, the main contribution of this paper has been to suggest two set of tests for endogeneity: one for endogeneity with respect to unobserved heterogeneity, and another with respect to idiosyncratic component for a subset of regressors that are deemed potentially endogenous. Moreover, our methodology circumvents the need for specifying a Mundlak (1978) or Chamberlain (1984) type specification for the conditional distribution of the unobserved heterogeneity in the structural equation. This allows us to conserves on degrees of freedom and to estimate the structural parameters of interest with much more precision when there is not enough variation among the regressors across time.

Our structural model, however, is static. Our next step is to incorporate dynamics in our estimation methodology. Besides, our methodology takes care of only of endogenous regressors that are continuous. Our next endeavor is to develop an estimator that accounts for both continuous and discrete endogenous regressors along the methodology proposed here. Thirdly, we would like to mention that our estimate of the EAP values of the unobserved individual effects are based on the parametric specification. For future research it would be worthwhile to investigate nonparametric methods to estimate the EAPs, which could then possibly lead to semiparametric estimation of the structural equation similar to that suggested by Blundell and Powell (2003).

References

Amemiya, T. (1971): "The Estimation of the Variances in a Variance-Components Model," *International Economic Review*, 2, 1–13.

Arellano, M. (2003): "Discrete Choices with Panel Data," *Investigaciones Econmicas*, 27, 423–458.

Arellano, M., and S. Bonhomme (2011): "Nonlinear Panel Data Analysis," *Annual Review of Economics*, forthcoming.

Arellano, M., and J. Hahn (2006): "A Likelihood-based Approximate Solution to the Incidental Parameter Problem in Dynamic Nonlinear Models with Multiple Effects," .

——— (2007): "Understanding Bias in Nonlinear Panel Models: Some Recent Developments," in *Advances in Economics and Econonometrics: Theory and Applications, Ninth World Congress*, ed. by W. N. R. Blundell, and T. Persson. Cambridge University Press, Cambridge.

Arellano, M., and B. Honore (2001): "Panel Data Models: Some Recent Developments," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 5, pp. 3229–3296. Elsevier.

Baltagi, B. H., and Q. Li (1994): "Estimating Error Component Models with General MA(q) Disturbances," *Journal of Econometrics*, 10(2), 396–408.

Baltagi, B. H., S. H. Song, and B. C. Jung (2010): "Testing for Heteroskedasticity and Serial Correlation in a Random Effects Panel Data Model," *Journal of Econometrics*, 154, 122–124.

Bester, A., and C. Hansen (2007): "Flexible Correlated Random Effects Estimation in Panel Models with Unobserved Heterogeneity," Technical report, mimeo.

Bester, C. A., and C. Hansen (2009): "A Penalty Function Approach to Bias Reduction in Non-linear Panel Models with Fixed Effects," *Journal of Business and Economic Statistics*, 27(2), 131–148.

Biørn, E. (2004): "Regression Systems for Unbalanced Panel Data: A Stepwise Maximum Likelihood Procedure," *Journal of Econometrics*, 122, 281–291.

Blundell, R., and J. Powell (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econonometrics: Theory and Applications, Eighth World Congress, Vol. II*, ed. by L. H. M. Dewatripont, and S. Turnovsky. Cambridge University Press, Cambridge.

Carro, J. (2007): "Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects," *Journal of Econometrics*, 140(2), 503–528.

Chamberlain, G. (1984): "Panel Data," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. 2. Elsevier.

——— (2010): "Binary Response Models for Panel Data: Identification and Information," *Econometrica*, 78, 159–168.

Chernozhukov, V., I. Fernandez-val, J. Hahn, and W. Newey (2009): "Identification and Estimation of Marginal Effects in Nonlinear Panel Models," Unpublished Manuscript.

Chernozhukov, V., I. Fernandez-val, and W. Newey (2009): "Quantile and Average Effects in Nonseparable Panel Models," CeMMAP working papers.

Chernozhukov, V., J. Hahn, and W. Newey (2005): "Bound analysis in panel models with correlated random effects," Technical report, MIT, UCLA.

Cools, R., and A. Haegemans (1994): "An Imbedded Family of Cubature Formulae for n-dimensional Product Regions," *Journal of Computational and Applied Mathematics*, 51, 251–260.

Fernandez-Val, I. (2009): "Fixed effects estimation of structural parameters and marginal effects in panel probit models," *Journal of Econometrics*, 150(1), 71–85.

Genz, A., and B. Keister (1996): "Fully Symmetric Interpolatory Rules for Multiple Integrals over Infinite Regions with Gaussian Weight," *Journal of Computation and Applied Mathematics*, 71, 299–309.

Hahn, J., and G. Kuersteiner (2011): "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects," *Econometric Theory*.

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

Honore, B., and E. Tamer (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74(3), 611–629.

Kobel, B. M. (2004): "First-order serial correlation in seemingly unrelated regressions," *Economics Letters*, 82(1), 1–7.

KROMMER, A. R., AND C. W. UEBERHUBER (1994): *Numerical Integration: on Advanced Computer Systems (Lecture Notes in Computer Science)*. Springer-Verlag, 1 edn.

LANCASTER, E. (2000): "The incidental parameter problem since 1948," *Journal of Econometrics*, 95(2), 391–413.

LUTKEPOHL, H. (1996): *Handbook of Matrices*. Wiley, Chichester.

MAGNUS, J.R., H. N. (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, Chichester.

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

NEWEY, W. K. (1984): "A Method of Moment Interpretation of Sequential Estimators," *Economics Letters*, 14, 201–206.

NEYMAN, J., AND E. SCOTT (1948): "Consistent estimates based on partially consistent observations," *Econometrica*, 16(1), 1–32.

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.

SEMYKINA, A., AND J. WOOLDRIDGE (2010): "Estimating panel data models in the presence of endogeneity and selection," *Journal of Econometrics*, 157, 375–380.

WEIDNER, M. (2011): "Semiparametric Estimation of Nonlinear Panel Data Models with Generalized Random Effects," Working Paper, University College London.

WOOLDRIDGE, J. M. (1995): "Selection Corrections for Panel Data Models Under Conditional Mean Independence Assumptions," *Journal of Econometrics*, 68, 115–132.

——— (2009): "Correlated Random Effects Models with Unbalanced Panels," Working Paper, Michigan State University, Department of Economics.

WOUTERSEN, T. (2002): "Robustness against Incidental Parameters," Unpublished Manuscript.

Appendix A: Maximum Likelihood Estimation of the Reduced form Equations

In this section we briefly describe Biørn (2004) step wise maximum likelihood procedure to estimate the reduced form system of equation

$$\mathbf{x}_{it} = \mathbb{Z}'_{it}\boldsymbol{\delta} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{it}. \tag{A-1}$$

While Biørn (2004) deals with unbalanced panel, here we assume that our panel is balanced. Let N be the total number of individuals. Let $\mathcal{N}$ be the total number of observations, i.e., $\mathcal{N} = NT$. Let $\mathbf{x}_{i(T)} = (\mathbf{x}'_{i1}, \dots \mathbf{x}'_{ip})'$, $\mathbb{Z}_{i(T)} = (\mathbb{Z}'_{i1}, \dots \mathbb{Z}'_{iT})'$ and $\boldsymbol{\epsilon}_{i(T)} = (\boldsymbol{\epsilon}'_{i1}, \dots \boldsymbol{\epsilon}'_{iT})'$ and write the model as

$$\mathbf{x}_{i(T)} = \mathbb{Z}'_{i(T)}\boldsymbol{\delta} + (e_p \otimes \boldsymbol{\alpha}_i) + \boldsymbol{\epsilon}_{i(T)} = \mathbb{Z}'_{i(T)}\boldsymbol{\delta} + \boldsymbol{u}_{i(T)}, \tag{A-2}$$

$$\mathrm{E}(\boldsymbol{u}_{i(T)}\boldsymbol{u}'_{i(T)}) = I_T \otimes \Sigma_{\epsilon\epsilon} + E_T \otimes \Lambda_{\alpha\alpha} = K_T \otimes \Sigma_{\epsilon\epsilon} + J_T \otimes \Sigma_{(T)} = \Omega_{u(T)} \tag{A-3}$$

where

$$\Sigma_{(T)} = \Sigma_{\epsilon\epsilon} + T\Lambda_{\alpha\alpha}, \tag{A-4}$$

and I_T is the T dimensional identity matrix, e_T is the $(T \times 1)$ vector of ones, $E_T = e_T e'_T$, $J_T = (1/T)E_T$, and $K_T = I_T - J_T$. The latter two matrices are symmetric and idempotent and have orthogonal columns, which facilitates inversion of $\Omega_{u(T)}$.

A.1 GLS estimation

Before addressing the maximum likelihood problem, we consider the GLS problem for $\boldsymbol{\delta}$ when Λ_α and $\Sigma_{\epsilon\epsilon}$ are known. Define $Q_{i(T)} = \boldsymbol{u}'_{i(T)}\Omega^{-1}_{u(T)}\boldsymbol{u}_{i(T)}$, then GLS estimation is the problem of minimizing $Q = \sum_{i=1}^{N} Q_{i(T)}$ with respect to $\boldsymbol{\delta}$. Since $\Omega^{-1}_{u(T)} = K_T \otimes \Sigma^{-1}_{\epsilon\epsilon} + J_T \otimes (\Sigma_{\epsilon\epsilon} + T\Lambda_{\alpha\alpha})^{-1}$, we can rewrite Q as

$$Q = \sum_{i=1}^{N} \boldsymbol{u}'_{i(T)}[K_p \otimes \Sigma^{-1}_{\epsilon\epsilon}]\boldsymbol{u}_{i(T)} + \sum_{i=1}^{N} \boldsymbol{u}'_{i(T)}[J_T \otimes (\Sigma_{\epsilon\epsilon} + T\Lambda_{\alpha\alpha})^{-1}]\boldsymbol{u}_{i(T)}. \tag{A-5}$$

GLS estimator of $\boldsymbol{\delta}$ when $\Lambda_{\alpha\alpha}$ and $\Sigma_{\epsilon\epsilon}$ are known is obtained from $\partial Q/\partial\boldsymbol{\delta} = 0$, and is given by

$$\hat{\boldsymbol{\delta}}_{GLS} = \Bigg[\sum_{i=1}^{N} \mathbb{Z}'_{i(T)}[K_T \otimes \Sigma_{\epsilon\epsilon}^{-1}]\mathbb{Z}_{i(T)} + \sum_{i=1}^{N} \mathbb{Z}'_{i(T)}[J_T \otimes (\Sigma_{\epsilon\epsilon} + T\Lambda_{\alpha\alpha})^{-1}]\mathbb{Z}_{i(T)}\Bigg]^{-1} \times \Bigg[\sum_{i=1}^{N} \mathbb{Z}'_{i(T)}[K_T \otimes \Sigma_{\epsilon\epsilon}^{-1}]\boldsymbol{x}_{i(T)} + \sum_{i=1}^{N} \mathbb{Z}'_{i(T)}[J_T \otimes (\Sigma_{\epsilon\epsilon} + T\Lambda_{\alpha\alpha})^{-1}]\boldsymbol{x}_{i(T)}\Bigg]. \quad \text{(A-6)}$$

A.1.1 Maximum Likelihood Estimation

Now consider ML estimation of $\boldsymbol{\delta}$, $\Sigma_{\epsilon\epsilon}$, and $\Lambda_{\alpha\alpha}$. Assuming normality of the individual effects and the disturbances, i.e., $\boldsymbol{\alpha}_i \sim \text{IIN}(0, \Lambda_{\alpha\alpha})$ and $\boldsymbol{\epsilon}_{it} \sim \text{IIN}(0, \Sigma_{\epsilon\epsilon})$, then $\boldsymbol{u}_{i(T)} = (e_T \otimes \boldsymbol{\alpha}_i) + \boldsymbol{\epsilon}_{i(T)} \sim \text{IIN}(0_{mT,1}, \Omega_{u(T)})$. The log-likelihood functions of all $\mathbf{x}$'s conditional on all $\mathbf{Z}$'s for an individual and for all individuals in the data set then become, respectively,

$$\mathcal{L}_{i(T)} = \frac{-mT}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega_{u(T)}| - \frac{1}{2}Q_{i(T)}(\boldsymbol{\delta}, \Sigma_{\epsilon\epsilon}, \Lambda_{\alpha\alpha}), \quad \text{(A-7)}$$

$$\mathcal{L} = \sum_{i=1}^{N} \mathcal{L}_{i(T)} = \frac{-mNT}{2}\ln(2\pi) - \frac{1}{2}N\ln|\Omega_{u(T)}| - \frac{1}{2}\sum_{i=1}^{N} Q_{i(T)}(\boldsymbol{\delta}, \Sigma_{\epsilon\epsilon}, \Lambda_{\alpha\alpha}), \quad \text{(A-8)}$$

where

$$Q_{i(T)}(\boldsymbol{\delta}, \Sigma_{\epsilon\epsilon}, \Lambda_{\alpha\alpha}) = [\mathbf{x}_{i(T)} - \mathbb{Z}'_{i(T)}\boldsymbol{\delta}]'[K_p \otimes \Sigma_{\epsilon\epsilon}^{-1} + J_p \otimes (\Sigma_{\epsilon\epsilon} + p\Lambda_{\alpha\alpha})^{-1}][\mathbf{x}_{i(T)} - \mathbb{Z}'_{i(T)}\boldsymbol{\delta}], \quad \text{(A-9)}$$

and $|\Omega_{u(T)}| = |\Sigma_{(T)}||\Sigma_{\epsilon\epsilon}|^{T-1}$.

Biørn (2004) splits the problem of estimation into: (A) *Maximization of* $\mathcal{L}$ *with respect to* $\boldsymbol{\delta}$ *for given* $\Sigma_{\epsilon\epsilon}$ *and* $\Lambda_{\alpha\alpha}$ and (B) *Maximization of* $\mathcal{L}$ *with respect to* $\Sigma_{\epsilon\epsilon}$ *and* $\Lambda_{\alpha\alpha}$ *for given* $\boldsymbol{\delta}$. *Subproblem* (A) is identical with the GLS problem, since maximization of $\mathcal{L}$ with respect to $\boldsymbol{\delta}$ for given $\Sigma_{\epsilon\epsilon}$ and $\Lambda_{\alpha\alpha}$ is equivalent to minimization of $\sum_{p=1}^{P}\sum_{i\in I_{(p)}} Q_{(p)}(\boldsymbol{\delta}, \Sigma_{\epsilon\epsilon}, \Lambda_{\alpha\alpha})$, which gives (7). To solve *subproblem*(B) Biørn (2004) derives expressions for the derivatives of both $\mathcal{L}_p$ and $\mathcal{L}$ with respect to $\Sigma_{\epsilon\epsilon}$ and $\Lambda_{\alpha\alpha}$. The complete stepwise algorithm for solving jointly subproblems (A) and (B) then consists in switching between (A-6) and minimizing (A-8) with respect to $\Sigma_{\epsilon\epsilon}$ and $\Lambda_{\alpha\alpha}$ to obtain $\Sigma_{\epsilon\epsilon}$ and $\Lambda_{\alpha\alpha}$ and iterating until convergence.

The covariance matrix of $\boldsymbol{\delta}$, vech$(\Lambda_{\alpha\alpha})$ and vech$(\Sigma_{\epsilon\epsilon})$ is given by

$$V\begin{bmatrix}\boldsymbol{\delta}\\ \text{vech}(\Lambda_{\alpha\alpha})\\ \text{vech}(\Sigma_{\epsilon\epsilon})\end{bmatrix} = \left[\sum_{i=1}^{N}\begin{bmatrix}\frac{\partial\mathcal{L}_{i(T)}(\hat{\boldsymbol{\delta}},\hat{\Sigma}_{\epsilon\epsilon},\hat{\Lambda}_{\alpha\alpha})}{\partial\boldsymbol{\delta}}\\ \frac{\partial\mathcal{L}_{i(T)}(\hat{\boldsymbol{\delta}},\hat{\Sigma}_{\epsilon\epsilon},\hat{\Lambda}_{\alpha\alpha})}{\partial\text{vech}(\Lambda_{\alpha\alpha})}\\ \frac{\partial\mathcal{L}_{i(T)}(\hat{\boldsymbol{\delta}},\hat{\Sigma}_{\epsilon\epsilon},\hat{\Lambda}_{\alpha\alpha})}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})}\end{bmatrix}\begin{bmatrix}\frac{\partial\mathcal{L}_{i(T)}(\hat{\boldsymbol{\delta}},\hat{\Sigma}_{\epsilon\epsilon},\hat{\Lambda}_{\alpha\alpha})}{\partial\boldsymbol{\delta}'}\\ \frac{\partial\mathcal{L}_{i(T)}(\hat{\boldsymbol{\delta}},\hat{\Sigma}_{\epsilon\epsilon},\hat{\Lambda}_{\alpha\alpha})}{\partial\text{vech}(\Lambda_{\alpha\alpha})'}\\ \frac{\partial\mathcal{L}_{i(T)}(\hat{\boldsymbol{\delta}},\hat{\Sigma}_{\epsilon\epsilon},\hat{\Lambda}_{\alpha\alpha})}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'}\end{bmatrix}\right]^{-1} \tag{A-10}$$

where each of the above is computed at the estimated values $\hat{\Sigma}_{\epsilon\epsilon}$, $\hat{\Lambda}_{\alpha\alpha}$, and $\hat{\boldsymbol{\delta}}$, and vech$(\Sigma_{\epsilon\epsilon})$ and vech$(\Lambda_{\alpha\alpha})$ are column-wise vectorization of the lower triangle of the symmetric matrix $\Sigma_{\epsilon\epsilon}$ and $\Lambda_{\alpha\alpha}$ and each are $\frac{m(m+1)}{2}$ column matrices. Biørn (2004) has derived the first order conditions with respect to $\Sigma_{\epsilon\epsilon}$ and $\Lambda_{\alpha\alpha}$. The first order condition with respect to $\boldsymbol{\delta}$ can be easily obtained. Here we state the first order conditions for the likelihood function for an individual i with respect to vech$(\Sigma_{\epsilon\epsilon})$ and vech$(\Lambda_{\alpha\alpha})$, which can then be used to compute the covariance matrix of the first stage reduced form estimates given in (A-10). It can shown that

$$\frac{\partial\mathcal{L}_{i(T)}}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})} = -\frac{1}{2}L_m\text{vec}\left[\Sigma_{(T)}^{-1} + (T-1)\Sigma_{\epsilon\epsilon}^{-1} - \Sigma_{(T)}^{-1}B_{ui(T)}\Sigma_{(T)}^{-1} - \Sigma_{\epsilon\epsilon}^{-1}W_{ui(T)}\Sigma_{\epsilon\epsilon}^{-1}\right],$$

and

$$\frac{\partial\mathcal{L}_{i(T)}}{\partial\text{vech}(\Lambda_{\alpha\alpha})} = -\frac{1}{2}L_m\text{vec}\left[T\Sigma_{(T)}^{-1} - T\Sigma_{(T)}^{-1}B_{ui(T)}\Sigma_{(T)}^{-1}\right],$$

where L_m is an elimination matrix. $W_{ui(T)}$ and $B_{ui(T)}$ respectively are defined as follows

$$W_{ui(T)} = \tilde{E}_{i(T)}K_T\tilde{E}'_{i(T)} \text{ and } B_{ui(T)} = \tilde{E}_{i(T)}J_T\tilde{E}'_{i(T)}, \tag{A-11}$$

where the disturbances defined in (A-2) for an individual i, have been arranged in a $(m\times T)$ matrix $\tilde{E}_{i(T)} = [\boldsymbol{u}_{i1},\dots,\boldsymbol{u}_{iT}]$ so that $\boldsymbol{u}_{i(T)} = \text{vec}(E_{i(T)})$, 'vec' being the vectorization operator.

Appendix B: Proofs

B.1 Lemma 1

$\hat{\hat{\boldsymbol{\alpha}}}_i(\boldsymbol{X}_i,\mathcal{Z}_i,\hat{\Theta}_1)$ and $\hat{\tilde{\Sigma}}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}_i(\boldsymbol{X}_i,\mathcal{Z}_i,\hat{\Theta}_1)$ converges a.s. to $\hat{\boldsymbol{\alpha}}_i(\boldsymbol{X}_i,\mathcal{Z}_i,\Theta_1^*)$ and $\tilde{\Sigma}_{\epsilon\epsilon}^{*-1}\hat{\boldsymbol{\alpha}}_i(\boldsymbol{X}_i,\mathcal{Z}_i,\Theta_1^*)$ respectively, where $\hat{\Theta}_1 = \{\hat{\boldsymbol{\delta}}', vech(\hat{\Sigma}_{\epsilon\epsilon})', vech(\hat{\Lambda}_{\alpha\alpha})'\}'$ is consistent first stage estimates and Θ_1^* is the true population parameter.

Proof: Now for an individual i

$$\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}(\mathbf{X},\mathcal{Z},\Theta_1) = \frac{\int\Sigma_\epsilon\Sigma_{\epsilon\epsilon}^{-1}C\mathfrak{a}\exp(-\frac{1}{2}\sum_{t=1}^{T}(\mathbf{x}_t-\mathbb{Z}_t'\boldsymbol{\delta}-\boldsymbol{\alpha})'\Sigma_{\epsilon\epsilon}^{-1}(\mathbf{x}_t-\mathbb{Z}_t'\boldsymbol{\delta}-C\mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a}}{\int\exp(-\frac{1}{2}\sum_{t=1}^{T}(\mathbf{x}_t-\mathbb{Z}_t'\boldsymbol{\delta}-C\mathfrak{a})'\Sigma_{\epsilon\epsilon}^{-1}(\mathbf{x}_t-\mathbb{Z}_t'\boldsymbol{\delta}-C\mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a}}$$

$$= \frac{\int\Sigma(\Theta_1,\mathfrak{a})\exp(-\frac{1}{2}r(\Theta_1,\mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a}}{\int\exp(-\frac{1}{2}r(\Theta_1,\mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a}}, \tag{B-1}$$

where $\boldsymbol{\alpha} = C\mathfrak{a}$, CC' being the Cholesky decomposition of the covariance matrix $\Lambda_{\alpha\alpha}$, so that $d\boldsymbol{\alpha} = |C|d\mathfrak{a} = |\Lambda_{\alpha\alpha}|^{1/2}d\mathfrak{a}$, $\Sigma(\Theta_1, \mathfrak{a}) = \Sigma_{\epsilon\epsilon}\Sigma_{\epsilon\epsilon}^{-1}C\mathfrak{a}$, and finally $r(\Theta_1, \mathfrak{a}) = \sum_{t=1}^{T}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - C\mathfrak{a})'\Sigma_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - C\mathfrak{a})$.

First consider the expression in the numerator $\int \Sigma(\Theta_1, \mathfrak{a})\exp(-\frac{1}{2}r(\Theta_1, \mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a}$. Now, $\Sigma(\Theta_1, \mathfrak{a}) = \Sigma_{\epsilon}\Sigma_{\epsilon\epsilon}^{-1}C\mathfrak{a}$ is an $m \times 1$ matrix and continuous in Θ_1 and $\mathfrak{a}$. Let $\Sigma_l(\Theta_1, \mathfrak{a})$ be the l^{th} element of $\Sigma(\Theta_1, \mathfrak{a})$. Now, by the assumptions of MLE we know that $\boldsymbol{\Theta_1}$ is a compact set, where $\Theta_1 \in \boldsymbol{\Theta_1}$, and also for a given $\mathfrak{a}$, $|\Sigma_l(\Theta_1, \mathfrak{a})|$, $|.|$ being the absolute value of its argument, is continuous in Θ_1. Therefore $|\Sigma_l(\Theta_1, \mathfrak{a})|$ attains its supremum on $\boldsymbol{\Theta_1}$. Let

$$\Theta_{l1}^{\mathfrak{a}} = \underset{\Theta_1 \in \boldsymbol{\Theta_1}}{\operatorname{argmax}} |\Sigma_l(\Theta_1, \mathfrak{a})|, \tag{B-2}$$

then by an application of the Maximum Theorem we can conclude that $|\Sigma_l(\Theta_{l1}^{\mathfrak{a}}, \mathfrak{a})|$ is continuous in $\mathfrak{a}$. The above then implies that $|\Sigma_l(\Theta_{l1}^{\mathfrak{a}}, \mathfrak{a})| \geq \Sigma_l(\Theta_1, \mathfrak{a})\exp(-\frac{1}{2}r(\Theta_1, \mathfrak{a}))$ $\forall\Theta_1 \in \boldsymbol{\Theta_1}$. We also know that $\hat{\Theta}_1 \xrightarrow{a.s.} \Theta_1^*$, and since each of the $\Sigma_l(\Theta_1, \mathfrak{a})\exp(-\frac{1}{2}r(\Theta_1, \mathfrak{a}))$, $l \in \{1, \ldots, m\}$, is continuous in Θ_1 and $\mathfrak{a}$, $\Sigma_l(\hat{\Theta}_1, \mathfrak{a})\exp(-\frac{1}{2}r(\hat{\Theta}_1, \mathfrak{a})) \xrightarrow{a.s.} \Sigma_l(\Theta_1^*, \mathfrak{a})\exp(-\frac{1}{2}r(\Theta_1^*, \mathfrak{a}))$ for any given $\mathfrak{a}$. Thus by an application of Lebesque Dominated Convergence Theorem we can conclude that $\int \Sigma_l(\hat{\Theta}_1, \mathfrak{a})\exp(-\frac{1}{2}r(\hat{\Theta}_1, \mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a} \xrightarrow{a.s.} \int \Sigma_l(\Theta_1^*, \mathfrak{a})\exp(-\frac{1}{2}r(\Theta_1^*, \mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a}$.

Define $\Sigma(\Theta_1^{\mathfrak{a}}, \mathfrak{a}) = \{|\Sigma_1(\Theta_{11}^{\mathfrak{a}}, \mathfrak{a})|, \ldots, |\Sigma_m(\Theta_{m1}^{\mathfrak{a}}, \mathfrak{a})|\}'$, then $\Sigma(\Theta_1^{\mathfrak{a}}, \mathfrak{a}) \geq \Sigma(\Theta_1, \mathfrak{a})\exp(-\frac{1}{2}r(\Theta_1, \mathfrak{a}))$ $\forall\Theta_1 \in \boldsymbol{\Theta_1}$, and Lebesque Dominated Convergence Theorem implies that

$$\int \Sigma(\hat{\Theta}_1, \mathfrak{a})\exp(-\frac{1}{2}r(\hat{\Theta}_1, \mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a} \xrightarrow{a.s.} \int \Sigma(\Theta_1^*, \mathfrak{a})\exp(-\frac{1}{2}r(\Theta_1^*, \mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a}.$$

Also, since $1 \geq \exp(-\frac{1}{2}r(\Theta_1, \mathfrak{a}))$, we can conclude that

$$\int \exp(-\frac{1}{2}r(\hat{\Theta}_1, \mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a} \xrightarrow{a.s.} \int \exp(-\frac{1}{2}r(\Theta_1^*, \mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a}.$$

Given that both the numerator and the denominator in (B-1) defined at $\hat{\Theta}_1$ converge almost surly to the same defined at Θ_1^*, it can now be easily shown that

$$\hat{\tilde{\Sigma}}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}(\mathbf{X}, \mathcal{Z}, \hat{\Theta}_1) \xrightarrow{a.s.} \tilde{\Sigma}_{\epsilon\epsilon}^{*-1}\hat{\boldsymbol{\alpha}}(\mathbf{X}, \mathcal{Z}, \Theta_1^*).$$

Appendix C: Asymptotic Covariance Matrix of the Second Stage Structural Estimates

Newey (1984) has shown that sequential estimators can be interpreted as members of a class of Method of Moments (MM) estimators and that this interpretation facilitates derivation of asymptotic covariance matrices for multi-step estimators. Let $\Theta = \{\Theta_1', \Theta_2'\}'$, where Θ_1 and Θ_2 are respectively the parameters to be estimated in the first and second step estimation of the sequential estimator. Following Newey (1984) we write the first and second step estimation as an MM estimation based on the following population moment conditions:

$$\mathrm{E}(\mathcal{L}_{i(T)\Theta_1}) = E\frac{\partial \ln L_{i(T)}(\Theta_1)}{\partial \Theta_1} = 0 \tag{C-1}$$

$$\mathrm{E}(H_{i(T)\Theta_2}(\Theta_1, \Theta_2)) = 0 \tag{C-2}$$

and where $L_{i(T)}(\Theta_1)$ is the likelihood function for individual i belonging to group p, for the first step system of reduced form equations and $\mathrm{E}(H_{i(T)\Theta_2}(\Theta_1, \Theta_2))$ is the population moment condition for estimating Θ_2. If in the second stage we are to use likelihood technique to estimate the second stage parameters Θ_2 then $\mathrm{E}(H_{i(T)\Theta_2}(\Theta_1, \Theta_2)) = \mathrm{E}(\mathcal{L}_{i(T)2\Theta_2}(\Theta_1, \Theta_2)) = E\frac{\partial \ln L_{i(T)2}(\Theta_1, \Theta_2)}{\partial \Theta_2} = 0$ where $L_{i(T)2}(\Theta_1, \Theta_2)$ is the likelihood function for an individual for the second step estimation.

The estimates for Θ_1 and Θ_2 are obtained by solving the sample analog of the above population moment conditions. The sample analog of moment conditions for the first step estimation is given by

$$\frac{1}{N}\mathcal{L}_{\Theta_1}(\hat{\Theta}_1) = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial \ln L_{i(T)}(\hat{\Theta}_1)}{\partial \Theta_1} \tag{C-3}$$

where $\mathcal{L}_i(\Theta_1)$ is given by equation (A-7) in Appendix A. $\Theta_1 = \{\boldsymbol{\delta}', \mathrm{vech}(\Lambda_{\alpha\alpha})', \mathrm{vech}(\Sigma_{\epsilon\epsilon})'\}'$ and N is the total number of individuals/firms.

Under standard regularity conditions, which our assumptions satisfy, the first step, reduced form ML estimate, $\hat{\Theta}_1$, obtained by solving $\frac{1}{N}\mathcal{L}_{\Theta_1}(\hat{\Theta}_1) = 0$ is consistent, and that $\sqrt{N}(\hat{\Theta}_1 - \Theta_1^*)$ is asymptotically distributed as $\mathrm{N}(0, \Sigma_{\Theta_1})$. $\Sigma_{\Theta_1} = \mathcal{I}(\Theta_1^*)^{-1} = \lim \frac{1}{N}V_1(\hat{\Theta}_1)$, where $\mathcal{I}(\Theta_1^*)$ is the information matrix, Θ_1^* is the true value of Θ_1, and $V_1(\hat{\Theta}_1)$ is the estimated asymptotic covariance matrix of $\hat{\Theta}_1$ given in (A-10). Amemiya (1971) discusses the estimation and asymptotic properties of the variance of the error components obtained by ML method for two way random effect model for the case of a single equation. While the analysis in Amemiya (1971) can be extended to a multiple equation setting, here we do not work out the details of Σ_{Θ_1}, but only state that, given the regularity conditions of MLE, $\frac{1}{N}V_1(\hat{\Theta}_1)$ converges to Σ_{Θ_1}, a positive definite matrix.

The sample analog of population moment condition for the second step estimation is given by

$$\frac{1}{N}H_{\Theta_2}(\hat{\Theta}_1, \hat{\Theta}_2) = \frac{1}{N}\sum_{i=1}^{N} H_{i(T)\Theta_2}(\hat{\Theta}_1, \hat{\Theta}_2) \tag{C-4}$$

where $\Theta_2 = \{\boldsymbol{\varphi}', \tilde{\boldsymbol{\varphi}}', \mathrm{vec}(\bar{\Sigma}_{\theta\alpha})', \mathrm{vec}(\tilde{\Sigma}_{\zeta\epsilon})'\}'$, which with a abuse of notation we write as $\Theta_2 = \{\boldsymbol{\varphi}', \mathrm{vec}(\bar{\Sigma}_{\theta\alpha})', \mathrm{vec}(\tilde{\Sigma}_{\zeta\epsilon})'\}'$. If we pool all the observations together as, for example, in the probit model discussed earlier, we have

$$\frac{1}{N}H_{\Theta_2}(\hat{\Theta}_1, \hat{\Theta}_2) = \frac{1}{N}\sum_{i=1}^{N} H_{i(T)\Theta_2}(\hat{\Theta}_1, \hat{\Theta}_2) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} H_{it\Theta_2}(\hat{\Theta}_1, \hat{\Theta}_2). \tag{C-5}$$

We have shown that with EAP values $\hat{\boldsymbol{\alpha}}_i(\mathbf{X}_i, \mathcal{Z}_i, \Theta_1)$ substituted for $\boldsymbol{\alpha}_i$ still leads to the identification of Θ_2. Let Θ_2^* be the true values of Θ_2. Under the assumptions we make, solving $\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T_i} H_{it\Theta_2}(\hat{\Theta}_1, \Theta_2) = 0$ is asymptotically equivalent to solving $\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T_i} H_{it\Theta_2}(\Theta_1^*, \Theta_2) = 0$, where $\hat{\Theta}_1$ is a consistent first step estimate of Θ_1. Hence $\hat{\Theta}_2$ obtained by solving $\frac{1}{N}H_{\Theta_2}(\hat{\Theta}_1, \hat{\Theta}_2) = 0$ is a consistent estimate of Θ_2. Newey (1984) has derived the asymptotic distribution of the second step estimates of a two step sequential estimator.

To derive the asymptotic distribution of the second step estimates $\hat{\Theta}_2$, consider the stacked up sample moment conditions

$$\frac{1}{N}\begin{bmatrix} \mathcal{L}_{\Theta_1}(\hat{\Theta}_1) \\ H_{\Theta_2}(\hat{\Theta}_1, \hat{\Theta}_2) \end{bmatrix} = 0, \tag{C-6}$$

A series of Taylor's expansion of $\mathcal{L}_{\Theta_1}(\hat{\Theta}_1)$, $H_{\Theta_2}(\hat{\Theta}_1, \hat{\Theta}_2)$ and around Θ^* gives

$$\frac{1}{N}\begin{bmatrix} \mathcal{L}_{\Theta_1\Theta_1} & 0 \\ H_{\Theta_2\Theta_1} & H_{\Theta_2\Theta_2} \end{bmatrix} \begin{bmatrix} \sqrt{N}(\hat{\Theta}_1 - \Theta_1^*) \\ \sqrt{N}(\hat{\Theta}_2 - \Theta_2^*)) \end{bmatrix} = -\frac{1}{\sqrt{N}}\begin{bmatrix} \mathcal{L}_{\Theta_1} \\ H_{\Theta_2}. \end{bmatrix} \tag{C-7}$$

In matrix notation the above can be written as

$$B_{\Theta\Theta_N}\sqrt{N}(\hat{\Theta} - \Theta) = -\frac{1}{\sqrt{N}}\Lambda_{\Theta_N},$$

where Λ_{Θ_N} is evaluated at Θ^* and $B_{\Theta\Theta_N}$ is evaluated at points somewhere between $\hat{\Theta}$ and Θ^*. Under the standard regularity conditions for Generalized Method of Moments (GMM), see Newey (1984), $B_{\Theta\Theta_N}$ converges in probability to the lower block triangular matrix $B_* = \lim \mathrm{E}(B_{\Theta\Theta_N})$. B_* is given by

$$B_* = \begin{bmatrix} \mathbb{L}_{\Theta_1\Theta_1} & 0 \\ \mathbb{H}_{\Theta_2\Theta_1} & \mathbb{H}_{\Theta_2\Theta_2} \end{bmatrix}$$

where $\mathbb{L}_{\Theta_1\Theta_1} = \mathrm{E}(\mathcal{L}_{i(T)\Theta_1\Theta_1})$, $\mathbb{H}_{\Theta_2\Theta_1} = \mathrm{E}(H_{i(T)\Theta_2\Theta_1})$. $\frac{1}{\sqrt{N}}\Lambda_N$ converges asymptotically in distribution to a normal random variable with mean zero and a covariance matrix $A_* = \lim \mathrm{E}\frac{1}{N}\Lambda_N\Lambda_N'$, where A_* is given by

$$A_* = \begin{bmatrix} V_{LL} & V_{LH} \\ V_{HL} & V_{HH} \end{bmatrix},$$

and a typical element of A_*, say V_{LH}, is given by $V_{LH} = \mathrm{E}[\mathcal{L}_{i(T)\Theta_1}(\Theta_1)H_{i(T)\Theta_2}(\Theta_1,\Theta_2)']$. Under the regularity conditions $\sqrt{N}(\hat{\Theta} - \Theta^*)$ is asymptotically normal with zero mean and covariance matrix given by $B_*^{-1}A_*B_*^{-1\prime}$.

$$\sqrt{N}(\hat{\Theta} - \Theta^*) \overset{a}{\sim} \mathrm{N}[(0), (B_*^{-1}A_*B_*^{-1\prime})] \tag{C-8}$$

Now, since at Θ_1^*, to which $\hat{\Theta}_1$ converges,

$$\mathbb{L}_{\Theta_1\Theta_1} = \mathrm{E}\left[\frac{\partial\mathcal{L}_{i(T)}(\Theta_1)}{\partial\Theta_1\Theta_1'}\right] = -\mathrm{E}\left[\frac{\partial\mathcal{L}_{i(T)}(\Theta_1)}{\partial\Theta_1}\frac{\partial\mathcal{L}_{i(T)}(\Theta_1)}{\partial\Theta_1'}\right] = -\mathrm{E}[\mathcal{L}_{i(T)\Theta_1}(\Theta_1)\mathcal{L}_{i(T)\Theta_1}(\Theta_1)'],$$

we can employ the derivative $\mathcal{L}_{i(T)}(\Theta_1)$ of with respect to Θ_1 to compute $\mathcal{L}_{i(T)\Theta_1\Theta_1}$. In (C-8) by an application of the partitioned inverse formula we get

$$\mathbb{L}_{\Theta_1\Theta_1}^{-1}V_{LL}\mathbb{L}_{\Theta_1\Theta_1}^{-1\prime} = V_1^*, \tag{C-9}$$

where $V_1^* = \Sigma_{\Theta_1}$ is the asymptotic covariance matrix of $\hat{\Theta}_1$ based on maximization of $\mathcal{L}(\Theta_1)$.

To derive the asymptotic distribution of $\sqrt{N}(\hat{\Theta}_2 - \Theta_2^*)$, again an application of partitioned inverse formula and some matrix manipulation we get the asymptotic covariance matrix of $\sqrt{N}(\hat{\Theta}_2 - \Theta_2^*)$, V_2^*, where

$$\begin{aligned} V_2^* = {} & \mathbb{H}_{\Theta_2\Theta_2}^{-1}V_{HH}\mathbb{H}_{\Theta_2\Theta_2}^{-1\prime} + \mathbb{H}_{\Theta_2\Theta_2}^{-1}\mathbb{H}_{\Theta_2\Theta_1}\{\mathbb{L}_{\Theta_1\Theta_1}^{-1}V_{LL}\mathbb{L}_{\Theta_1\Theta_1}^{-1\prime}\}\mathbb{H}_{\Theta_2\Theta_1}'\mathbb{H}_{\Theta_2\Theta_2}^{-1\prime} \\ & -\mathbb{H}_{\Theta_2\Theta_2}^{-1}\{\mathbb{H}_{\Theta_2\Theta_1}\mathbb{L}_{\Theta_1\Theta_1}^{-1}V_{LH} + V_{HL}\mathbb{L}_{\Theta_1\Theta_1}^{-1\prime}\mathbb{H}_{\Theta_2\Theta_1}'\}\mathbb{H}_{\Theta_2\Theta_2}^{-1\prime}. \end{aligned} \tag{C-10}$$

To estimate the asymptotic covariance matrix V_2^*, sample analog of the B_* and A_*, B_N and $A_N = \frac{1}{N}\Lambda_N\Lambda_N'$ respectively, given in (C-7) can be computed. A typical element of A_N, say V_{LH_N}, is given by $V_{LH_N} = \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}_{i(T)\Theta_1}(\hat{\Theta}_1)H_{i(T)\Theta_2}(\hat{\Theta}_1,\hat{\Theta}_2)'$.

Now, while computation of $H_{\Theta_2\Theta_2}$ in (C-7) is straight forward, computation of $H_{\Theta_2\Theta_1} = \sum_{i=1}^{N}\sum_{t=1}^{T}H_{it\Theta_2\Theta_1} = \sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial H_{it\Theta_2}(\Theta_1,\Theta_2)}{\partial\Theta_1'}$ is challenging because Θ_1 enters the second stage of the sequential estimator through $\bar{\mathbf{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i(\Theta_1)$ and $\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}(\Theta_1)$. In what follows, we assume that the second stage structural estimation involves estimating a binary response model as in our discussion on identification of discrete response models. For binary response model we have

$$H_{it\Theta_2}(\Theta_1,\Theta_2) = y_{it}\frac{\phi(\mathbb{X}'_{it}\Theta_2)\mathbb{X}_{it}}{\Phi(\mathbb{X}'_{it}\Theta_2)} - (1-y_{it})\frac{\phi(\mathbb{X}'_{it}\Theta_2)\mathbb{X}_{it}}{1-\Phi(\mathbb{X}'_{it}\Theta_2)} = q_{it}\lambda_{it}\mathbb{X}_{it}, \tag{C-11}$$

where $\mathbb{X}_{it} = \{\mathcal{X}'_{it}, (\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i(\Theta_1))', (\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\epsilon}}_{it}(\Theta_1))'\}'$, $q_{it} = 2y_{it}-1$, and $\lambda_{it} = \frac{\phi(q_{it}\mathbb{X}'_{it}\Theta_2)}{\Phi(q_{it}\mathbb{X}'_{it}\Theta_2)}$. Hence we have

$$\frac{\partial H_{it\Theta_2}(\Theta_1,\Theta_2)}{\partial\Theta'_1} = q_{it}\Bigg[\lambda_{it}\bigg(- q_{it}\frac{\partial\mathbb{X}'_{it}}{\partial\Theta'_1}\Theta_2\mathbb{X}_{it} + \frac{\partial\mathbb{X}_{it}}{\partial\Theta'_1}\bigg) - \lambda^2_{it}\mathbb{X}_{it}q_{it}\frac{\partial\mathbb{X}'_{it}}{\partial\Theta'_1}\Theta_2\Bigg], \tag{C-12}$$

where

$$\frac{\partial\mathbb{X}_{it}}{\partial\Theta'_1} = \begin{bmatrix} \frac{\partial\mathcal{X}_{it}}{\partial\boldsymbol{\delta}'} & \frac{\partial\mathcal{X}_{it}}{\partial\text{vech}(\Lambda_{\alpha\alpha})'} & \frac{\partial\mathcal{X}_{it}}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'} \\ \frac{\partial(\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}}+\hat{\boldsymbol{\alpha}}_i)}{\partial\boldsymbol{\delta}'} & \frac{\partial(\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}}+\hat{\boldsymbol{\alpha}}_i)}{\partial\text{vech}(\Lambda_{\alpha\alpha})'} & \frac{\partial(\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}}+\hat{\boldsymbol{\alpha}}_i)}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'} \\ \frac{\partial\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\epsilon}}_{it}}{\partial\boldsymbol{\delta}'} & \frac{\partial\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\epsilon}}_{it}}{\partial\text{vech}(\Lambda_{\alpha\alpha})'} & \frac{\partial\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\epsilon}}_{it}}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'} \end{bmatrix}.$$

Since $\mathcal{X}_{it}$ above is not a function of Θ_1, $\frac{\partial\mathcal{X}_{it}}{\partial\Theta'_1} = \mathbf{0}_{\mathcal{X}}$, where $\mathbf{0}_{\mathcal{X}}$ is a null matrix with row dimension that of column vector $\mathcal{X}_{it}$ and column dimension that of column vector Θ_1. In subsection (C.1) we derive the derivative of $\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i(\Theta_1)$ and $\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\epsilon}}_{it}(\Theta_1)$ with respect to $\Theta_1 = \{\boldsymbol{\delta}', \text{vech}(\Lambda_{\alpha\alpha})', \text{vech}(\Sigma_{\epsilon\epsilon})'\}'$. We show that

$$\frac{\partial(\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i)}{\partial\boldsymbol{\delta}'} = \mathbb{O}'_{\mathbb{Z}i} - \frac{1}{U^2_{dr}}\sum_{t=1}^{T}\Big[U_{nr}U'_{nr} - U_{dr}F_{dr}\Big]\Sigma^{-1}_{\epsilon\epsilon}\mathbb{Z}'_{it},$$

$$\frac{\partial\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\epsilon}}_{it}}{\partial\boldsymbol{\delta}'} = -\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\mathbb{Z}'_{it} + \frac{\tilde{\Sigma}^{-1}_{\epsilon\epsilon}}{U^2_{dr}}\sum_{t=1}^{T}\Big[U_{nr}U'_{nr} - U_{dr}F_{dr}\Big]\Sigma^{-1}_{\epsilon\epsilon}\mathbb{Z}'_{it},$$

$$\frac{\partial(\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i)}{\partial\text{vech}(\Lambda_{\alpha\alpha})'} = \frac{1}{2U^2_{dr}}[U_{dr}F_{nr} - U_{nr}\text{vec}(F_{dr})'](\Lambda^{-1}_{\alpha\alpha}\otimes\Lambda^{-1}_{\alpha\alpha})'L'_m,$$

$$\frac{\partial\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\epsilon}}_{it}}{\partial\text{vech}(\Lambda_{\alpha\alpha})'} = \frac{-\tilde{\Sigma}^{-1}_{\epsilon\epsilon}}{2U^2_{dr}}[U_{dr}F_{nr} - U_{nr}\text{vec}(F_{dr})'](\Lambda^{-1}_{\alpha\alpha}\otimes\Lambda^{-1}_{\alpha\alpha})'L'_m,$$

$$\begin{aligned}\frac{\partial(\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i)}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'} = \frac{1}{2U^2_{dr}}\sum_{t=1}^{T}\Big[&U_{dr}(-\mathbf{r}'_{it}\otimes F_{dr} - F_{dr}\otimes\mathbf{r}'_{it} + F_{nr}) \\ &- U_{nr}\text{vec}(-U_{nr}\mathbf{r}'_{it} - \mathbf{r}_{it}U'_{nr} + F_{dr})'\Big](\Sigma^{-1}_{\epsilon\epsilon}\otimes\Sigma^{-1}_{\epsilon\epsilon})'L'_m,\end{aligned}$$

and

$$\begin{aligned}\frac{\partial\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\epsilon}}_{it}}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'} = &\frac{(U_{nr}\Sigma^{-1}_{\epsilon\epsilon}\otimes I_m)}{2U_{dr}}\text{vec}((\text{dg}(\Sigma_{\epsilon\epsilon}))^{-1/2})'L'_m + \Bigg[\frac{(U_{nr}\otimes\Sigma'_{\epsilon})'}{U_{dr}} \\ &- \frac{\tilde{\Sigma}^{-1}_{\epsilon\epsilon}}{2(U_{dr})^2}\sum_{t=1}^{T}\bigg(U_{dr}(-\mathbf{r}'_{it}\otimes F_{dr} - F_{dr}\otimes\mathbf{r}'_{it} + F_{nr}) - U_{nr}\text{vec}(-U_{nr}\mathbf{r}'_{it} - \mathbf{r}_{it}U'_{nr} + F_{dr})'\bigg)\Bigg](\Sigma^{-1}_{\epsilon\epsilon}\otimes\Sigma^{-1}_{\epsilon\epsilon})'L'_m,\end{aligned}$$

where

$$U_{nr} = \int I_m \boldsymbol{\alpha} \exp(-\frac{1}{2} r(\Theta_1, \boldsymbol{\alpha}))\phi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}, \quad F_{nr} = \int I_m \boldsymbol{\alpha} \text{vec}(\boldsymbol{\alpha}\boldsymbol{\alpha}')' \exp(-\frac{1}{2} r(\Theta_1, \boldsymbol{\alpha}))\phi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}$$
$$U_{dr} = \int \exp(-\frac{1}{2} r(\Theta_1, \boldsymbol{\alpha}))\phi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}, \quad F_{dr} = \int \boldsymbol{\alpha}\boldsymbol{\alpha}' \exp(-\frac{1}{2} r(\Theta_1, \boldsymbol{\alpha}))\phi(\boldsymbol{\alpha}) d\boldsymbol{\alpha},$$
$$\mathbf{r}_{it} = \mathbf{x}_{it} - \mathbb{Z}_i \boldsymbol{\delta},$$

$\mathbb{O}_{\mathbb{Z}i} = \text{diag}((0_z', \bar{\mathbf{z}}_i')', \ldots, (0_z', \bar{\mathbf{z}}_i')')$, and $0_z'$ is a vector of zeros of having the dimension of $\mathbf{z}_{it}$, which has been defined in Section 2 in the main text, and L_m is an elimination matrix. Now, while numerical integration technique, at the estimated value $\hat{\Theta}_1$, was employed for the computation of $\hat{U}_{nr}$ and $\hat{U}_{dr}$ to obtain $\hat{\hat{\boldsymbol{\alpha}}}_i$ for estimating the structural parameters of interest, in order to obtain the error adjusted standard errors of the structural estimates, $\hat{F}_{nr}$ and $\hat{F}_{dr}$ will also have to be numerically computed at the estimated reduced form parameters $\hat{\Theta}_1$.

In Lemma 1 we showed that $\hat{U}_{nr}(\hat{\Theta}_1)$ and $\hat{U}_{dr}(\hat{\Theta}_1)$ converge almost surely to $U_{nr}(\Theta_1^*)$, $U_{dr}(\Theta_1^*)$. By application of Lemma 1 it can be also shown that $\hat{F}_{nr}(\hat{\Theta}_1)$, and $\hat{F}_{dr}(\hat{\Theta}_1)$ converge almost surely to $F_{nr}(\Theta_1^*)$, and $F_{dr}(\Theta_1^*)$ respectively. This would imply that $H_{i(T)\Theta_2\Theta_1}(\hat{\Theta}_1, \hat{\Theta}_2)$ $= \sum_{t=1}^{T} \frac{\partial H_{it\Theta_2}(\hat{\Theta}_1, \hat{\Theta}_2)}{\partial \Theta_1'}$ converge almost surly to $\sum_{t=1}^{T} \frac{\partial H_{it\Theta_2}(\Theta_1^*, \Theta_2^*)}{\partial \Theta_1'} = \sum_{t=1}^{T} H_{it\Theta_2\Theta_1}(\Theta_1^*, \Theta_2^*) =$ $H_{i(T)\Theta_2\Theta_1}(\Theta_1^*, \Theta_2^*)$ and by the weak LLN $\frac{1}{N} H_{i(T)\Theta_2\Theta_1}(\hat{\Theta}_1, \hat{\Theta}_2)$ will converge in probability to $\text{E}(H_{i(T)\Theta_2\Theta_1}(\Theta_1^*, \Theta_2^*)) = \mathbb{H}_{\Theta_2\Theta_1}$

C.1 Derivative of $\bar{\mathbf{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i$ and $\tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it}$ with respect to Θ_1

First consider the derivative of $\bar{\mathbf{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i$ with respect to $\text{vech}(\Lambda_{\alpha\alpha})$. We have

$$\begin{aligned}
\frac{\partial(\bar{\mathbf{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i)}{\partial \text{vech}(\Lambda_{\alpha\alpha})'} &= \frac{\partial \hat{\boldsymbol{\alpha}}_i}{\partial \text{vech}(\Lambda_{\alpha\alpha})'} \\
&= \frac{\partial}{\partial \text{vech}(\Lambda_{\alpha\alpha})'} \left[\frac{\int \boldsymbol{\alpha} \exp(-\frac{1}{2} r(\Theta_1, \boldsymbol{\alpha}))\phi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}}{\int \exp(-\frac{1}{2} r(\Theta_1, \boldsymbol{\alpha}))\phi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}} \right] = \frac{\partial}{\partial \text{vech}(\Lambda_{\alpha\alpha})'} \left[\frac{\int f_{nr}(., \boldsymbol{\alpha})\phi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}}{\int f_{dr}(., \boldsymbol{\alpha})\phi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}} \right] \\
&= \frac{[\int f_{nr}(., \boldsymbol{\alpha}) \frac{\partial \phi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}}{\partial \text{vech}(\Lambda_{\alpha\alpha})'}][\int f_{dr}(., \boldsymbol{\alpha})\phi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}] - [\int f_{nr}(., \boldsymbol{\alpha})\phi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}][\int f_{dr}(., \boldsymbol{\alpha}) \frac{\partial \phi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}}{\partial \text{vech}(\Lambda_{\alpha\alpha})'}]}{[\int f_{dr}(., \boldsymbol{\alpha})\phi(\boldsymbol{\alpha}) d\boldsymbol{\alpha}]^2}
\end{aligned} \tag{C-13}$$

Now, since $\phi(\boldsymbol{\alpha}) = \frac{1}{(2\pi)^{m/2}|\Lambda_{\alpha\alpha}|^{1/2}} \exp(-\frac{1}{2}\boldsymbol{\alpha}'\Lambda_{\alpha\alpha}^{-1}\boldsymbol{\alpha})$ we have

$$\begin{aligned}
\frac{\partial\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha}}{\partial\text{vech}(\Lambda_{\alpha\alpha})'} &= \frac{-\exp(-\frac{1}{2}\boldsymbol{\alpha}'\Lambda_{\alpha\alpha}^{-1}\boldsymbol{\alpha})}{2(2\pi)^{m/2}|\Lambda_{\alpha\alpha}|^{3/2}}\frac{\partial|\Lambda_{\alpha\alpha}|}{\partial\text{vech}(\Lambda_{\alpha\alpha})'} + \frac{\exp(-\frac{1}{2}\boldsymbol{\alpha}'\Lambda_{\alpha\alpha}^{-1}\boldsymbol{\alpha})}{(2\pi)^{m/2}|\Lambda_{\alpha\alpha}|^{1/2}}\frac{\partial(-\frac{1}{2}\boldsymbol{\alpha}'\Lambda_{\alpha\alpha}^{-1}\boldsymbol{\alpha})}{\partial\text{vech}(\Lambda_{\alpha\alpha})'} \\
&= -\frac{1}{2}\phi(\boldsymbol{\alpha})\left(\frac{1}{|\Lambda_{\alpha\alpha}|}\frac{\partial|\Lambda_{\alpha\alpha}|}{\partial\text{vech}(\Lambda_{\alpha\alpha})'} + \frac{\partial(\boldsymbol{\alpha}'\Lambda_{\alpha\alpha}^{-1}\boldsymbol{\alpha})}{\partial\text{vech}(\Lambda_{\alpha\alpha})'}\right) \\
&= -\frac{1}{2}\phi(\boldsymbol{\alpha})\left(\text{vec}(\Lambda_{\alpha\alpha}^{-1})' + \text{vec}(-(\Lambda_{\alpha\alpha}^{-1})'\boldsymbol{\alpha}\boldsymbol{\alpha}'(\Lambda_{\alpha\alpha}^{-1})')'\right)\frac{\partial\text{vec}(d(\Lambda_{\alpha\alpha}))}{\partial\text{vech}(\Lambda_{\alpha\alpha})'} \\
&= -\frac{1}{2}\phi(\boldsymbol{\alpha})\left(\text{vec}(\Lambda_{\alpha\alpha}^{-1})' + \text{vec}(-(\Lambda_{\alpha\alpha}^{-1})'\boldsymbol{\alpha}\boldsymbol{\alpha}'(\Lambda_{\alpha\alpha}^{-1})')'\right)L_m', \qquad \text{(C-14)}
\end{aligned}$$

where L_m' is an elimination matrix. Given (C-14), (C-13) can be simplified as

$$\begin{aligned}
\frac{\partial(\bar{\mathbf{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i)}{\partial\text{vech}(\Lambda_{\alpha\alpha})'} &= -\frac{1}{2U_{dr}^2}\Big[[U_{nr}\text{vec}(\Lambda_{\alpha\alpha}^{-1})' - F_{nr}(\Lambda_{\alpha\alpha}^{-1}\otimes\Lambda_{\alpha\alpha}^{-1})']L_m'U_{dr} \\
&\qquad - U_{nr}[U_{dr}\text{vec}(\Lambda_{\alpha\alpha}^{-1})' - F_{dr}(\Lambda_{\alpha\alpha}^{-1}\otimes\Lambda_{\alpha\alpha}^{-1})']L_m'\Big] \\
&= \frac{1}{2U_{dr}^2}[U_{dr}F_{nr} - U_{nr}\text{vec}(F_{dr})'](\Lambda_{\alpha\alpha}^{-1}\otimes\Lambda_{\alpha\alpha}^{-1})'L_m', \qquad \text{(C-15)}
\end{aligned}$$

where

$$\begin{aligned}
U_{nr} &= \int I_m\boldsymbol{\alpha}\exp(-\frac{1}{2}r(\Theta_1,\boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha}, \quad F_{nr} = \int I_m\boldsymbol{\alpha}\text{vec}(\boldsymbol{\alpha}\boldsymbol{\alpha}')'\exp(-\frac{1}{2}r(\Theta_1,\boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha} \\
U_{dr} &= \int \exp(-\frac{1}{2}r(\Theta_1,\boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha}, \quad F_{dr} = \int \boldsymbol{\alpha}\boldsymbol{\alpha}'\exp(-\frac{1}{2}r(\Theta_1,\boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha}. \qquad \text{(C-16)}
\end{aligned}$$

Also, from (C-15) we gather that

$$\frac{\partial\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial\text{vech}(\Lambda_{\alpha\alpha})'} = \frac{-\partial\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}_i}{\partial\text{vech}(\Lambda_{\alpha\alpha})'} = \frac{-\tilde{\Sigma}_{\epsilon\epsilon}^{-1}}{2U_{dr}^2}[U_{dr}F_{nr} - U_{nr}\text{vec}(F_{dr})'](\Lambda_{\alpha\alpha}^{-1}\otimes\Lambda_{\alpha\alpha}^{-1})'L_m'. \qquad \text{(C-17)}$$

Now consider the derivative of $\bar{\mathbf{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i$ with respect to vech$(\Sigma_{\epsilon\epsilon})$. We have

$$\begin{aligned}
\frac{\partial(\bar{\mathbf{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i)}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'} &= \frac{\partial\hat{\boldsymbol{\alpha}}_i}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'} = \frac{\partial}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'}\left[\frac{\int\boldsymbol{\alpha}\exp(-\frac{1}{2}\sum_{t=1}^T\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha}}{\int\exp(-\frac{1}{2}\sum_{t=1}^T\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha}}\right] \\
&= -\frac{1}{2}\left[\frac{\int\boldsymbol{\alpha}\psi(\boldsymbol{\alpha})\frac{\partial\sum_{t=1}^T\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it}}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'}d\boldsymbol{\alpha}\int\psi(\boldsymbol{\alpha})d\boldsymbol{\alpha} - \int\boldsymbol{\alpha}\psi(\boldsymbol{\alpha})d\boldsymbol{\alpha}\int\psi(\boldsymbol{\alpha})\frac{\partial\sum_{t=1}^T\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it}}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'}d\boldsymbol{\alpha}}{(\int\psi(\boldsymbol{\alpha})d\boldsymbol{\alpha})^2}\right],
\end{aligned}$$

where $\psi(\boldsymbol{\alpha}) = \exp(-\frac{1}{2}\sum_{t=1}^T\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})\phi(\boldsymbol{\alpha})$. With $\frac{\partial\sum_{t=1}^T\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it}}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'} = \sum_{t=1}^T\text{vec}(-(\Sigma_{\epsilon\epsilon}^{-1})'\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}'(\Sigma_{\epsilon\epsilon}^{-1})')'L_m'$

the above can be written as

$$
\begin{aligned}
\frac{\partial\bar{\boldsymbol{\alpha}}_i}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'} = \frac{1}{2(\int\psi(\boldsymbol{\alpha})d\boldsymbol{\alpha})^2}\sum_{t=1}^{T}\Big[&\int\boldsymbol{\alpha}\psi(\boldsymbol{\alpha})\text{vec}((\Sigma_{\epsilon\epsilon}^{-1})'\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}'(\Sigma_{\epsilon\epsilon}^{-1})')'L_m'd\boldsymbol{\alpha}\int\psi(\boldsymbol{\alpha})d\boldsymbol{\alpha} \\
&-\int\psi(\boldsymbol{\alpha})\text{vec}((\Sigma_{\epsilon\epsilon}^{-1})'\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}'(\Sigma_{\epsilon\epsilon}^{-1})')'L_m'd\boldsymbol{\alpha}\int\boldsymbol{\alpha}\psi(\boldsymbol{\alpha})d\boldsymbol{\alpha}\Big] \\
= \frac{1}{2U_{dr}^2}\sum_{t=1}^{T}\Big[&\int\boldsymbol{\alpha}\psi(\boldsymbol{\alpha})\text{vec}(\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}')'(\Sigma_{\epsilon\epsilon}^{-1}\otimes\Sigma_{\epsilon\epsilon}^{-1})'L_m'd\boldsymbol{\alpha}U_{dr} \\
&-U_{nr}\int\psi(\boldsymbol{\alpha})\text{vec}(\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}')'(\Sigma_{\epsilon\epsilon}^{-1}\otimes\Sigma_{\epsilon\epsilon}^{-1})'L_m'd\boldsymbol{\alpha}\Big] \\
= \frac{1}{2U_{dr}^2}\sum_{t=1}^{T}\Big[&\int(U_{dr}\boldsymbol{\alpha}\text{vec}(\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}')'-U_{nr}\text{vec}(\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}')')\psi(\boldsymbol{\alpha})d\boldsymbol{\alpha}\Big](\Sigma_{\epsilon\epsilon}^{-1}\otimes\Sigma_{\epsilon\epsilon}^{-1})'L_m'
\end{aligned}
\tag{C-18}
$$

To simply (C-18) further, write $\boldsymbol{\epsilon}_{it}$ as $\boldsymbol{\epsilon}_{it} = \mathbf{x}_{it} - \mathbb{Z}_i\boldsymbol{\delta} - \boldsymbol{\alpha} = \mathbf{r}_{it} - \boldsymbol{\alpha}$, where $\mathbf{r}_{it} = \mathbf{x}_{it} - \mathbb{Z}_i\boldsymbol{\delta}$. Then $\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}' = \mathbf{r}_{it}\mathbf{r}_{it}' - \boldsymbol{\alpha}\mathbf{r}_{it}' - \mathbf{r}_{it}\boldsymbol{\alpha}' + \boldsymbol{\alpha}\boldsymbol{\alpha}'$, and (C-18) after some simplification can be written as

$$
\begin{aligned}
\frac{\partial\hat{\boldsymbol{\alpha}}_i}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'} = \frac{1}{2U_{dr}^2}\sum_{t=1}^{T}\Big[&U_{dr}(-\mathbf{r}_{it}'\otimes F_{dr} - F_{dr}\otimes\mathbf{r}_{it}' + F_{nr}) \\
&-U_{nr}\text{vec}(-U_{nr}\mathbf{r}_{it}' - \mathbf{r}_{it}U_{nr}' + F_{dr})'\Big](\Sigma_{\epsilon\epsilon}^{-1}\otimes\Sigma_{\epsilon\epsilon}^{-1})'L_m'.
\end{aligned}
\tag{C-19}
$$

The expressions in the parenthesis in (C-19) are

$$
\int I_m\boldsymbol{\alpha}\text{vec}(\boldsymbol{\alpha}\mathbf{r}_{it}')'\exp(-\frac{1}{2}r(\Theta_1,\boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha} = \int I_m\mathbf{r}_{it}'\otimes(\boldsymbol{\alpha}\boldsymbol{\alpha}')\exp(-\frac{1}{2}r(\Theta_1,\boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha} = \mathbf{r}_{it}'\otimes F_{dr}
$$

$$
\int I_m\boldsymbol{\alpha}\text{vec}(\mathbf{r}_{it}\boldsymbol{\alpha}')'\exp(-\frac{1}{2}r(\Theta_1,\boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha} = \int I_m(\boldsymbol{\alpha}\boldsymbol{\alpha}')\otimes\mathbf{r}_{it}'\exp(-\frac{1}{2}r(\Theta_1,\boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha} = F_{dr}\otimes\mathbf{r}_{it}'
$$

$$
\int\text{vec}(\boldsymbol{\alpha}\mathbf{r}_{it}')'\exp(-\frac{1}{2}r(\Theta_1,\boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha} = \text{vec}(\boldsymbol{\alpha}\exp(-\frac{1}{2}r(\Theta_1,\boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha}\mathbf{r}_{it}')' = \text{vec}(U_{nr}\mathbf{r}_{it}')'
$$

$$
\int\text{vec}(\mathbf{r}_{it}\boldsymbol{\alpha}')'\exp(-\frac{1}{2}r(\Theta_1,\boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha} = \text{vec}(\mathbf{r}_{it}\boldsymbol{\alpha}'\exp(-\frac{1}{2}r(\Theta_1,\boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha})' = \text{vec}(\mathbf{r}_{it}U_{nr}')',
$$

where U_{nr}, U_{dr}, F_{nr}, and F_{dr} have been defined in (C-16).

Let us now consider the derivative $\frac{\partial\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}_i}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'} = \frac{\partial(\Sigma_\epsilon\Sigma_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}_i)}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'}$. The total differential of $\Sigma_\epsilon\Sigma_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}_i$ is given by:

$$
d(\Sigma_\epsilon\Sigma_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}_i) = d(\Sigma_\epsilon)\Sigma_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}_i + \Sigma_\epsilon d(\Sigma_{\epsilon\epsilon}^{-1})\hat{\boldsymbol{\alpha}}_i + \Sigma_\epsilon\Sigma_{\epsilon\epsilon}^{-1}d(\hat{\boldsymbol{\alpha}}_i). \tag{C-20}
$$

Now, as defined earlier, $\Sigma_\epsilon = (\text{dg}(\Sigma_{\epsilon\epsilon}))^{1/2}$, hence

$$
\begin{aligned}
\frac{\partial(\Sigma_\epsilon)\Sigma_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}_i}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'} &= \frac{1}{2}(\hat{\boldsymbol{\alpha}}_i'\Sigma_{\epsilon\epsilon}^{-1}\otimes I_m)\text{vec}((\text{dg}(\Sigma_{\epsilon\epsilon}))^{-1/2})\frac{\partial\text{vec}(\Sigma_{\epsilon\epsilon})}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'} \\
&= \frac{1}{2}(\hat{\boldsymbol{\alpha}}_i'\Sigma_{\epsilon\epsilon}^{-1}\otimes I_m)\text{vec}((\text{dg}(\Sigma_{\epsilon\epsilon}))^{-1/2})'L_m'.
\end{aligned}
\tag{C-21}
$$

Now, consider the second term of the differential given in (C-20). It can be shown that

$$\frac{\Sigma_\epsilon \partial(\Sigma_{\epsilon\epsilon}^{-1})\hat{\boldsymbol{\alpha}}_i}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} = -(\hat{\boldsymbol{\alpha}}_i \otimes \Sigma_\epsilon')'(\Sigma_{\epsilon\epsilon}^{-1} \otimes \Sigma_{\epsilon\epsilon}^{-1})\frac{\partial \text{vec}(\Sigma_{\epsilon\epsilon})}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} = -(\hat{\boldsymbol{\alpha}}_i \otimes \Sigma_\epsilon')'(\Sigma_{\epsilon\epsilon}^{-1} \otimes \Sigma_{\epsilon\epsilon}^{-1})L_m'. \quad \text{(C-22)}$$

Finally, consider the third term in the total differential in (C-20). From (C-19) we can conclude that

$$\frac{\Sigma_\epsilon \Sigma_{\epsilon\epsilon}^{-1}\partial(\hat{\boldsymbol{\alpha}}_i)}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} = \frac{\Sigma_\epsilon \Sigma_{\epsilon\epsilon}^{-1}}{2U_{dr}^2}\sum_{t=1}^{T}\Big[U_{dr}(-\mathbf{r}_{it}' \otimes F_{dr} - F_{dr} \otimes \mathbf{r}_{it}' + F_{nr}) - U_{nr}\text{vec}(-U_{nr}\mathbf{r}_{it}' - \mathbf{r}_{it}U_{nr}' + F_{dr})'\Big](\Sigma_{\epsilon\epsilon}^{-1} \otimes \Sigma_{\epsilon\epsilon}^{-1})'L_m'. \quad \text{(C-23)}$$

Combining (C-21), (C-22), and (C-23) we obtain

$$\frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} = \frac{(U_{nr}\Sigma_{\epsilon\epsilon}^{-1} \otimes I_m)}{2U_{dr}}\text{vec}((\text{dg}(\Sigma_{\epsilon\epsilon}))^{-1/2})'L_m' + \Bigg[\frac{(U_{nr} \otimes \Sigma_\epsilon')'}{U_{dr}} - \frac{\Sigma_\epsilon \Sigma_{\epsilon\epsilon}^{-1}}{2(U_{dr})^2}\sum_{t=1}^{T}\Big(U_{dr}(-\mathbf{r}_{it}' \otimes F_{dr} - F_{dr} \otimes \mathbf{r}_{it}' + F_{nr}) - U_{nr}\text{vec}(-U_{nr}\mathbf{r}_{it}' - \mathbf{r}_{it}U_{nr}' + F_{dr})'\Big)\Bigg](\Sigma_{\epsilon\epsilon}^{-1} \otimes \Sigma_{\epsilon\epsilon}^{-1})'L_m'. \quad \text{(C-24)}$$

We note here that $\frac{\partial \hat{\boldsymbol{\alpha}}_i}{\partial \text{vech}(\Lambda_{\alpha\alpha})'}$ and $\frac{\partial \hat{\boldsymbol{\alpha}}_i}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'}$, respectively, for an individual i are same for all time periods.

Finally, let us now consider the derivative of $\bar{\mathbf{Z}}'\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i$ and $\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}$ with respect to $\boldsymbol{\delta}'$. We have

$$\begin{aligned}\frac{\partial(\bar{\mathbf{Z}}'\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i)}{\partial \boldsymbol{\delta}'} &= \frac{\partial \bar{\mathbf{Z}}_i'\bar{\boldsymbol{\delta}}}{\partial \boldsymbol{\delta}'} + \frac{\partial \hat{\boldsymbol{\alpha}}_i}{\partial \boldsymbol{\delta}'} = \mathbb{O}_{\mathbb{Z}i}' + \frac{\partial}{\partial \boldsymbol{\delta}'}\left[\frac{\int \boldsymbol{\alpha}\exp(-\frac{1}{2}\sum_{t=1}^{T}\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha}}{\int \exp(-\frac{1}{2}\sum_{t=1}^{T}\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha}}\right] \\ &= \mathbb{O}_{\mathbb{Z}i}' - \frac{1}{(\int \exp(.)\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha})^2}\sum_{t=1}^{T}\Bigg[\int \boldsymbol{\alpha}\exp(.)\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\mathbb{Z}_{it}'\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha}\int \exp(.)\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha} \\ &\quad - \int \boldsymbol{\alpha}\exp(.)\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha}\int \exp(.)\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\mathbb{Z}_{it}'\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha}\Bigg], \quad \text{(C-25)}\end{aligned}$$

where $\mathbb{O}_{\mathbb{Z}i} = \text{diag}((0_z', \bar{\mathbf{z}}_i')', \ldots, (0_z', \bar{\mathbf{z}}_i')')$, and $0_z'$ is a vector of zeros of having the dimension of $\mathbf{z}_{it}$, which has been defined in Section 2 in the main text. To derive the above result in (C-25) we used the fact that

$$\frac{\partial(\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})}{\partial \boldsymbol{\delta}'} = 2\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\frac{\partial(\boldsymbol{\epsilon}_{it})}{\partial \boldsymbol{\delta}'} = -2\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\mathbb{Z}_{it}'.$$

With some of the results stated above it can be shown that $\frac{\partial \hat{\boldsymbol{\alpha}}_i}{\partial \boldsymbol{\delta}'} = \frac{1}{U_{dr}^2}\sum_{t=1}^{T}\Big[U_{nr}U_{nr}' - U_{dr}F_{dr}\Big]\Sigma_{\epsilon\epsilon}^{-1}\mathbb{Z}_{it}'$. Hence we have

$$\frac{\partial(\bar{\mathbf{Z}}'\bar{\boldsymbol{\delta}} + \hat{\boldsymbol{\alpha}}_i)}{\partial \boldsymbol{\delta}'} = \mathbb{O}_{\mathbb{Z}i}' - \frac{1}{U_{dr}^2}\sum_{t=1}^{T}\Big[U_{nr}U_{nr}' - U_{dr}F_{dr}\Big]\Sigma_{\epsilon\epsilon}^{-1}\mathbb{Z}_{it}', \quad \text{(C-26)}$$

and

$$\frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial \boldsymbol{\delta}'} = \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}(\mathbf{x}_{it} - \mathbb{Z}_{it}'\boldsymbol{\delta})}{\partial \boldsymbol{\delta}'} - \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\alpha}}_i}{\partial \boldsymbol{\delta}'} = -\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\mathbb{Z}_{it}' + \frac{\tilde{\Sigma}_{\epsilon\epsilon}^{-1}}{U_{dr}^2}\sum_{t=1}^{T}\Big[U_{nr}U_{nr}' - U_{dr}F_{dr}\Big]\Sigma_{\epsilon\epsilon}^{-1}\mathbb{Z}_{it}'. \tag{C-27}$$

From (C-26) and (C-27) we can see that while $\frac{\partial(\bar{\mathbf{Z}}'\bar{\boldsymbol{\delta}}+\hat{\boldsymbol{\alpha}}_i)}{\partial \boldsymbol{\delta}'}$ for an individual i remains the same for all time periods, $\frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial \boldsymbol{\delta}'}$ varies with time.

C.2 Hypothesis Testing of Average Partial Effects

In section 2.3 we showed how to compute the average partial effect (APE) of a variable w belonging to $\mathcal{X}'$. To test various hypothesis in order to draw inferences about the APE's we need to compute the standard errors of their estimates. From (2.24) and (2.25) in the main text we know that estimated APE of w on the probability of $y_{it} = 1$ is given by

$$\frac{\partial \widehat{\Pr}(y_{it} = 1)}{\partial w} = \frac{1}{\sum_{i=1}^{N} T_i}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\hat{\varphi}_w\hat{\phi}(\bar{\mathbb{X}}_{it}'\hat{\Theta}_2),$$

where $\bar{\mathbb{X}}_{it} = \{\bar{\mathcal{X}}', (\bar{\mathbf{Z}}_i'\hat{\bar{\boldsymbol{\delta}}}+\hat{\boldsymbol{\alpha}}_i)', (\hat{\bar{\Sigma}}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it})'\}'$ and $\Theta_2 = \{\boldsymbol{\varphi}', \text{vec}(\bar{\Sigma}_{\theta\alpha})', \text{vec}(\tilde{\Sigma}_{\zeta\epsilon})'\}'$. Since each of the $\hat{\varphi}_w\hat{\phi}(\bar{\mathbb{X}}_{it}'\hat{\Theta}_2)$ is a function of $\hat{\Theta}_2$ the variance of $\frac{\partial \widehat{\Pr}(y_{it}=1)}{\partial w}$ will be a function of the variance of the estimate of Θ_2. Now, we know that by the linear approximation approach (delta method), the asymptotic covariance matrix of $\frac{\partial \widehat{\Pr}(y_{it}=1)}{\partial w}$ is given by

$$\text{Asy. Var}[\frac{\partial \widehat{\Pr}(y_{it} = 1)}{\partial w}] = \Bigg[\frac{1}{\sum_{i=1}^{N} T_i}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\frac{\partial \hat{\varphi}_w\hat{\phi}(\bar{\mathbb{X}}_{it}'\hat{\Theta}_2)}{\partial \hat{\Theta}_2'}\Bigg]V_2^*\Bigg[\frac{1}{\sum_{i=1}^{N} T_i}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\frac{\partial \hat{\varphi}_w\hat{\phi}(\bar{\mathbb{X}}_{it}'\hat{\Theta}_2)}{\partial \hat{\Theta}_2'}\Bigg]', \tag{C-28}$$

where V_2^* is the second stage error adjusted covariance matrix, shown above, of $\underline{\hat{\boldsymbol{\varphi}}}$. The RHS of (C-28) turns out to be

$$\Bigg[\frac{1}{\sum_{i=1}^{N} T_i}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\hat{\phi}(\bar{\mathbb{X}}_{it})[e_w - (\hat{\Theta}_2'\bar{\mathbb{X}}_{it})\hat{\varphi}_w\bar{\mathbb{X}}_{it}']\Bigg]V_2^*\Bigg[\frac{1}{\sum_{i=1}^{N} T_i}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\hat{\phi}(\bar{\mathbb{X}}_{it})[e_w - (\hat{\Theta}_2'\bar{\mathbb{X}}_{it})\hat{\varphi}_w\bar{\mathbb{X}}_{it}']\Bigg]', \tag{C-29}$$

where and e_w is a row vector having the dimension of Θ_2' and with 1 at the position of φ_w in Θ_2 and zeros elsewhere. The estimated asymptotic covariance matrix of the APE of all the continuous variables in $\mathcal{X}^f$ on the probability of being financially constrained can be obtained as above.

If w is a dummy variable then from (2.26) we know that the estimated APE of w is given by

$$\Delta_w \Pr(y_{it}=1) = \frac{1}{\sum_{i=1}^N T_i}\sum_{i=1}^N\sum_{t=1}^{T_i} \hat{\Phi}(\bar{\mathcal{X}}_{-w}, w=1, \hat{\hat{\alpha}}_i, \hat{\hat{\boldsymbol{\epsilon}}}_{it}) - \hat{\Phi}(\bar{\mathcal{X}}_{-w}, w=0, \hat{\hat{\alpha}}_i, \hat{\hat{\boldsymbol{\epsilon}}}_{it})$$
$$= \frac{1}{\sum_{i=1}^N T_i}\sum_{i=1}^N\sum_{t=1}^{T_i} \Delta_w\hat{\Phi}_{it}()$$

To obtain the variance of the above, again by the delta method we have

$$\text{Asy. Var}\Delta_w \Pr(f_{it}=1) = \left[\frac{1}{\sum_{i=1}^N T_i}\sum_{i=1}^N\sum_{t=1}^{T_i}\frac{\partial\Delta\hat{\Phi}_{it}(.)}{\partial\Theta_2'}\right]' V_2^* \left[\frac{1}{\sum_{i=1}^N T_i}\sum_{i=1}^N\sum_{t=1}^{T_i}\frac{\partial\Delta\hat{\Phi}_{it}(.)}{\partial\Theta_2'}\right], \tag{C-30}$$

where

$$\frac{\partial\Delta\hat{\Phi}_{it}(.)}{\partial\Theta_2} = \frac{\partial\hat{\Phi}_{it}(w=1)}{\partial\Theta_2} - \frac{\partial\hat{\Phi}_{it}(w=0)}{\partial\Theta_2} = \phi_{it}(w=1)\begin{bmatrix}\bar{\mathbb{X}}_{it-w}\\ 1\end{bmatrix} - \hat{\phi}_{it}(w=0)\begin{bmatrix}\bar{\mathbb{X}}_{it-w}\\ 0\end{bmatrix}.$$

Appendix D: Note on Numerical Integration

In order to obtain the structural estimates we have to compute the expected a posteriori values of the time invariant individual effects given by:

$$\hat{\boldsymbol{\alpha}}(\mathbf{X}, \mathcal{Z}, \Theta_1) = \frac{\int C\mathfrak{a}\exp(-\frac{1}{2}\sum_{t=1}^T(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - \boldsymbol{\alpha})'\Sigma_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - C\mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a}}{\int \exp(-\frac{1}{2}\sum_{t=1}^T(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - C\mathfrak{a})'\Sigma_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - C\mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a}}$$
$$= \frac{\int C\exp(-\frac{1}{2}r(\Theta_1,\mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a}}{\int \exp(-\frac{1}{2}r(\Theta_1,\mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a}} = \frac{U_{nr}}{U_{dr}}, \tag{D-1}$$

where $\boldsymbol{\alpha} = C\mathfrak{a}$, CC' being the Cholesky decomposition of the $(m \times m)$ covariance matrix $\Lambda_{\alpha\alpha}$, so that $d\boldsymbol{\alpha} = |C|d\mathfrak{a} = |\Lambda_{\alpha\alpha}|^{1/2}d\mathfrak{a}$, and $r(\Theta_1,\mathfrak{a}) = \sum_{t=1}^T(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - C\mathfrak{a})'\Sigma_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbb{Z}_t'\boldsymbol{\delta} - C\mathfrak{a})$. And to obtain error adjusted covariance matrix in addition to U_{nr} and U_{dr} we have to estimate F_{nr} and F_{dr} given by

$$F_{nr} = \int I_m\boldsymbol{\alpha}\text{vec}(\boldsymbol{\alpha}\boldsymbol{\alpha}')'\exp(-\frac{1}{2}r(\Theta_1,\boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha} \text{ and } F_{dr} = \int \boldsymbol{\alpha}\boldsymbol{\alpha}'\exp(-\frac{1}{2}r(\Theta_1,\boldsymbol{\alpha}))\phi(\boldsymbol{\alpha})d\boldsymbol{\alpha} \tag{D-2}$$

respectively.

Here we discuss how to compute U_{nr}, U_{dr}, F_{nr}, and F_{dr}. Take, for example, U_{nr}, which can be written as

$$\int C\mathfrak{a}\exp(-\frac{1}{2}r(\Theta_1,\mathfrak{a}))\phi(\mathfrak{a})d\mathfrak{a} = \int C\mathfrak{a}\exp(-\frac{1}{2}r(\Theta_1,\mathfrak{a}))e^{-\frac{\mathfrak{a}'\mathfrak{a}}{2}}d\mathfrak{a} = \int f(\mathfrak{a})e^{-\frac{\mathfrak{a}'\mathfrak{a}}{2}}d\mathfrak{a},$$

where $\int f(\mathfrak{a})e^{-\frac{\mathfrak{a}'\mathfrak{a}}{2}}d\mathfrak{a} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(\mathfrak{a})e^{-\frac{\mathfrak{a}'\mathfrak{a}}{2}}d\mathfrak{a}_1 \dots d\mathfrak{a}_m$.

A general treatment for numerically computing multidimensional integrals can be found in Krommer and Ueberhuber (1994). More recently Cools and Haegemans (1994) have developed integration rules for multidimensional integrals over infinite integration regions with a Gaussian weight function to evaluate integrals of the type stated above, and Genz and Keister (1996) have provided more efficient rules of the same. The integration rules consist of constructing $\mathfrak{N}$ weights, w_j, and points $\mathfrak{a}_j$, $\mathfrak{a}_j \in \mathbb{R}^m$, such that $Q(f)$,

$$Q(f) = \sum_{j=1}^{\mathfrak{N}} w_j f(\mathfrak{a}_j), \tag{D-3}$$

approximate the integral $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(\mathfrak{a})e^{-\frac{\mathfrak{a}'\mathfrak{a}}{2}}d\mathfrak{a}_1 \dots d\mathfrak{a}_m$. Fortran routines for computing $Q(f)$, developed in Genz and Keister (1996), can be obtained from Alan Genz's webpage.