

Analysis of interactive fixed effects dynamic linear panel regression with measurement error

Nayoung Lee
Hyungsik Roger Moon
Martin Weidner

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP37/11

Analysis of Interactive Fixed Effects Dynamic Linear Panel Regression with Measurement Error*

Nayoung Lee[†]
CUHK

Hyungsik Roger Moon
USC & U of Maryland

Martin Weidner
UCL and CeMMAP

March 12, 2012

Abstract

This paper studies a simple dynamic linear panel regression model with interactive fixed effects in which the variable of interest is measured with error. To estimate the dynamic coefficient, we consider the least-squares minimum distance (LS-MD) estimation method.

Keywords: dynamic panel, interactive fixed effects, measurement error, LS-MD estimation.

JEL Classification: C23, C26

1 Introduction

This paper studies a *simple* dynamic linear panel regression model with interactive fixed effects in which the variable of interest, say Y_{it}^* , contains measurement error:

$$Y_{it}^* = \alpha_0 Y_{it-1}^* + \lambda_i^0 f_t^0 + \epsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (1)$$

$$Y_{it} = Y_{it}^* + \eta_{it}. \quad (2)$$

*We thank Badi Baltagi and an anonymous referee for helpful comments and suggestions. Moon thanks the NSF for financial support.

[†]Lee (Corresponding Author): Department of Economics, Chinese University of Hong Kong; Shatin, New Territories, Hong Kong; Email: nayoung.lee@cuhk.edu.hk; Telephone: (852) 3943-8004; Fax: (852) 2603-5805. Moon: Department of Economics, University of Maryland, Tydings Hall, Room 3105, College Park, MD 20742; E-mail: moon@econ.umd.edu. Weidner: Department of Economics, University College London, Gower St., London, WC1E 6BT, U.K.; E-mail: m.weidner@ucl.ac.uk.

Here Y_{it} is the observed variable and η_{it} represents measurement error. The term $\lambda_i^0 f_t^0$ describes unobserved interactive fixed effects.^{1,2} The goal of the paper is to estimate α_0 when both the number of individuals N and the number of time periods T are large.³

The dynamics of the observed variable Y_{it} can be written as

$$Y_{it} = \alpha_0 Y_{it-1} + \lambda_i^0 f_t^0 + U_{it}, \quad (3)$$

where $U_{it} = \epsilon_{it} + \eta_{it} - \alpha_0 \eta_{it-1}$. There are two noticeable features in equations (1) and (3) compared to the widely studied dynamic panel regression model. First, the individual effects take an interactive form instead of the time invariant form. Secondly, the variable of interest Y_{it}^* is not observed but measured with error. To our knowledge, combining these two features in dynamic linear panel regression models has not been studied in the large N, T panel literature.

We expect two hurdles in estimating α_0 . One is the presence of the interactive fixed effects $\lambda_i^0 f_t^0$ which might cause a so-called incidental parameter problem in both the cross section and the time dimension. The second one is that the composite error U_{it} in the observed variable equation (3) is correlated with the lagged dependent variable Y_{it-1} and we may therefore need to use instrumental variables (IVs).

The main contribution of the paper is to find a valid estimation method that overcomes these two problems. The proposed estimator is a nested two-step estimator based on least squares minimization in the first step and distance minimization for some of the first step parameter estimates in the second step⁴. Following Moon, Shum and Weidner (2012) (hereafter MSW), we call this method the LS-MD estimation method. This approach was used in estimating endogenous quantile regression models by Chernozhukov and Hansen (2006, 2008) and in estimating the random coefficient logit demand model by MSW.

¹In this paper, we consider a single factor, that is, the dimensions of f_t and λ_i are equal to one. The extension to the multiple factor case is straightforward, but omitted due to space limitation.

²When interpreting λ_i^0 as individual specific fixed effects, the term f_t^0 represents the (time-varying) linear projection coefficient of Y_{it}^* on λ_i^0 (holding Y_{it-1}^* constant). This allows the effect of the unobserved individual characteristic λ_i^0 on Y_{it}^* to be time-varying. Alternatively, one can interpret f_t^0 as a common time specific shock (a common factor) and λ_i^0 then describes reaction to the common shock (a factor loading).

³We consider large N, T approximations to characterize the bias due to the incidental parameters $\lambda_i^0 f_t^0$, see e.g. Bai (2009) and Hahn and Kuersteiner (2004).

⁴An alternative approach would be to use the common correlated effect methods suggested by Harding and Lamarche (2011). Both approaches have their own merits and weaknesses. Comparing these different methods is not our interest in this paper.

2 LS-MD Estimation

The properties of the quasi-maximum likelihood estimator (QMLE), which minimizes the sum of squared residuals, for large N, T linear panel regressions with interactive fixed effects were discussed in Bai (2009), and Moon and Weidner (2010). However, this estimation method cannot be used to estimate model (3) since the regressor Y_{it-1} is endogenous w.r.t. the error U_{it} through the lagged measurement error η_{it-1} . In this case, we may use instrumental variables. Since U_{it} has an $MA(1)$ type serial dependence structure, we have $E(U_{it}Y_{it-1-s}) = 0$ for all $s \geq 1$. This suggests to choose $Z_{it} = (Z_{1,it}, \dots, Z_{L,it})' = (Y_{it-2}, \dots, Y_{it-1-L})'$ for the IVs of the endogenous regressor Y_{it-1} . The question, then, is how to use the instrumental variables Z_{it} to estimate α_0 in the presences of interactive fixed effects $\lambda_i^0 f_t^0$ when both N and T are large.

The estimation method we consider in this paper is a two-step least-squares minimum distance (LS-MD) estimation. This was recently proposed by MSW for estimating the BLP demand model. A similar multi-step estimation idea was also used in Chernozhukov and Hansen (2006, 2008) in estimating endogenous quantile regressions with IVs.

The LS-MD estimation consists of the following two steps: Step 1: For given α , we solve the least squares problem augmented by the instrumental variables Z_{it} , that is, we run the OLS regression of $Y_{it} - \alpha Y_{it-1}$ on Z_{it} with interactive fixed effects $\lambda_i f_t$ and solve

$$\left(\hat{\gamma}(\alpha), \hat{\lambda}(\alpha), \hat{f}(\alpha) \right) = \arg \min_{(\gamma, \lambda, f)} \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \alpha Y_{it-1} - \gamma' Z_{it} - \lambda_i f_t)^2,$$

where $\gamma = (\gamma_1, \dots, \gamma_L)'$, $\lambda = (\lambda_1, \dots, \lambda_N)'$ and $f = (f_1, \dots, f_T)'$. Step 2: For some positive definite weight matrix W_{NT}^γ , we estimate α by minimizing the length of $\hat{\gamma}(\alpha)$ as

$$\hat{\alpha} = \arg \min_{\alpha} \hat{\gamma}(\alpha)' W_{NT}^\gamma \hat{\gamma}(\alpha).$$

The idea of the LS-MD method is that since Z_{it} is excluded in the regression equation (3) the coefficient of Z_{it} should be zero when $\alpha = \alpha_0$. When there is no interactive fixed effect one can show that the LS-MD estimator is equivalent to the conventional 2SLS estimator for an appropriate weight matrix W_{NT}^γ .

3 Asymptotic Results

Assumption 3.1 (i) The unobserved error terms $\{\epsilon_{it}\} \sim iid(0, \sigma_\epsilon^2)$ and $\{\eta_{it}\} \sim iid(0, \sigma_\eta^2)$ across i and over t and $E|\epsilon_{it}|^\kappa, E|\eta_{it}|^\kappa < \infty$ for some $\kappa > 8$. Also, $\{\epsilon_{it}\}$ and $\{\eta_{it}\}$ are independent. (ii) Assume that f_t^0 are strictly stationary and ergodic with $\sup_t |f_t| < \infty$ and $\frac{1}{T} \sum_{t=1}^T (f_t^0)^2 \rightarrow_p \Sigma_f > 0$, and λ_i are iid with $\sup_i |\lambda_i| < \infty$ and $\frac{1}{N} \sum_{i=1}^N (\lambda_i^0)^2 \rightarrow_p \Sigma_\lambda > 0$. Also assume that $\{f_t^0\}, \{\lambda_i^0\}, \{\epsilon_{it}\}, \{\eta_{it}\}$ are independent. (iii) $W_{NT}^\gamma \rightarrow_p W^\gamma > 0$. (vi) $|\alpha_0| < 1$ and $\alpha_0 \neq 0$.

The iid assumptions of ϵ_{it} and η_{it} are made for simplicity of the analysis. Later, an extension to a non-iid case will be discussed. Assumption 3.1(i) also assumes that the measurement error η_{it} is classical in the sense that η_{it} has zero mean and is uncorrelated with Y_{it}^* . Later we discuss how to extend our method to some special cases of non-classical measurement error. Assumption 3.1(ii) assumes that the factors are strong, which is standard in the factor analysis literature. Assumption 3.1(vi) assumes that $\alpha_0 \neq 0$, otherwise the IVs become irrelevant.

Before we present the next assumption, we introduce some further notation. We use $[a_{it}]_{it}$ to denote an $N \times T$ matrix with elements a_{it} . For a full column rank matrix A , let $\mathbb{P}_A = A(A'A)^{-1}A'$ and $\mathbb{M}_A = I - \mathbb{P}_A$. We use notation $Y = [Y_{it}]_{it}$, $Y_{-k} = [Y_{it-k}]_{it}$, $Z = [Z_{it}]_{it}$, $U = [U_{it}]_{it}$, $\epsilon = [\epsilon_{it}]_{it}$, $\eta = [\eta_{it}]_{it}$, and $\eta_{-1} = [\eta_{it-1}]_{it}$. Define $\lambda^0 = (\lambda_1^0, \dots, \lambda_N^0)'$ and $f^0 = (f_1^0, \dots, f_T^0)'$. We also define the NT -vectors $y_{-1} = vec(Y_{-1})$ and $z = vec(Z)$.

Assumption 3.2 Assume that there exists a positive constant $c > 0$ such that $\frac{1}{NT} y_{-1}' \mathbb{P}_z y_{-1} - \max_\lambda \frac{1}{NT} y_{-1}' \mathbb{P}_{I_T \otimes \tilde{\lambda}} y_{-1} > c$ with probability approaching one as $N, T \rightarrow \infty$, where $\tilde{\lambda} = (\lambda^0, \lambda)$.

Assumption 3.2 is a relevance condition on the instruments. It demands that the explanatory power of the instruments Z_{it} for the endogenous regressor Y_{it-1} , given by $\frac{1}{NT} y_{-1}' \mathbb{P}_z y_{-1}$, is larger than the joint explanatory power for Y_{it-1} of the true factor loading λ^0 together with any other factor loading λ , given by $\frac{1}{NT} y_{-1}' \mathbb{P}_{I_T \otimes \tilde{\lambda}} y_{-1}$. If there are no interactive fixed effects included in the model, then the assumption simplifies to the standard relevance condition $\frac{1}{NT} y_{-1}' \mathbb{P}_z y_{-1} > 0$, which is satisfied for $\alpha^0 \neq 0$.

Suppose that Assumption 3.1 holds, and consider the special case where f_t^0 has mean

zero and is distributed independently over t . Then, Assumption 3.2 is equivalent to⁵

$$\alpha_0^2 > \frac{1 + \frac{\sigma_\varepsilon^2}{\Sigma_\lambda \Sigma_f} + \frac{\sigma_\eta^2}{\Sigma_\lambda \Sigma_f}}{\left(1 + \frac{\sigma_\varepsilon^2}{\Sigma_\lambda \Sigma_f}\right)^2 + \frac{\sigma_\eta^2}{\Sigma_\lambda \Sigma_f}}. \quad (4)$$

Thus, by imposing an appropriate lower bound on $|\alpha_0|$, one can guarantee that the lagged values of Y_{it} are sufficiently relevant instruments. The conclusion that an appropriate lower bound on $|\alpha_0|$ is sufficient for the relevance assumption Assumption 3.2 can be extended to cases where f_t^0 is correlated across t , but in general it is not possible to give such a convenient analytic expression as in (4) for the lower bound.⁶ Note that the lower bound in (4) goes to zero when $\Sigma_\lambda \Sigma_f$ becomes small relative to σ_ε^2 , i.e., the bound is not restrictive when the relative influence of the factors on Y_{it} is small.

Theorem 3.1 *Under Assumption 3.1 and 3.2 we have $\hat{\alpha} \rightarrow_p \alpha_0$ as $N, T \rightarrow \infty$.*^{7,8}

To present the limiting distribution of $\hat{\alpha}$, we need to introduce some further notation. Define the NT -vectors $y_{-1}^{\lambda f}$ and $z_l^{\lambda f}$ by $y_{-1}^{\lambda f} = \text{vec}(\mathbb{M}_{\lambda_0} Y_{-1} \mathbb{M}_{f_0})$, and $z_l^{\lambda f} = \text{vec}(\mathbb{M}_{\lambda_0} Z_l \mathbb{M}_{f_0})$, where $l = 1, \dots, L$. Let $u = \text{vec}(U)$ and $z^{\lambda f} = (z_1^{\lambda f}, \dots, z_L^{\lambda f})'$.

Define $G = \text{plim}_{N, T \rightarrow \infty} \frac{1}{NT} y_{-1}^{\lambda f'} z^{\lambda f} = \frac{\sigma_\varepsilon^2}{1 - \alpha_0^2} (\alpha_0, \alpha_0^2, \dots, \alpha_0^L)'$, and

$$\begin{aligned} W &= \text{plim}_{N, T \rightarrow \infty} \left(\frac{1}{NT} z^{\lambda f'} z^{\lambda f} \right)^{-1} W_{NT}^\gamma \left(\frac{1}{NT} z^{\lambda f'} z^{\lambda f} \right)^{-1} \\ &= \left\{ \frac{\sigma_\varepsilon^2}{1 - \alpha_0^2} \begin{bmatrix} 1 & \cdots & \alpha_0^{L-1} \\ \vdots & \ddots & \vdots \\ \alpha_0^{L-1} & \cdots & 1 \end{bmatrix} + \sigma_\eta^2 I_L \right\} W^\gamma \left\{ \frac{\sigma_\varepsilon^2}{1 - \alpha_0^2} \begin{bmatrix} 1 & \cdots & \alpha_0^{L-1} \\ \vdots & \ddots & \vdots \\ \alpha_0^{L-1} & \cdots & 1 \end{bmatrix} + \sigma_\eta^2 I_L \right\}. \end{aligned}$$

⁵For the proof of this, we refer to the supplementary appendix which is available at <http://www.cemmap.ac.uk/publications.php>.

⁶A non-zero mean of f_t^0 can result in situations where Assumption 3.2 is not satisfied for any value of α_0 . The assumption that f_t^0 is mean zero would not be restrictive if we would include a conventional individual specific fixed effect in the model, in addition to the interactive fixed effect — or equivalently (from an asymptotic perspective), one can demean Y_{it} separately for each i before estimating the model with only interactive effects.

⁷The proof is omitted due to space limitation. It is a special case of MSW where their $\delta(\alpha) = Y - \alpha Y_{-1}$ and the conditions in Assumptions 3.1 and 3.2 are sufficient for the consistency conditions in MSW (see the supplementary appendix available at <http://www.cemmap.ac.uk/publications.php>.)

⁸Note that Assumption 3.2 is a sufficient condition for the relevance of the instruments, but nothing is known about the necessity of this assumption. The LS-MD estimator may also give consistent parameter estimates in some situations where the assumption is violated.

Notice that under Assumption 3.1, the limits G and W are well defined. Also, notice that under Assumption 3.1, we have $GWG' > 0$.

Define

$$c^{(2)} = \left[C^{(2)}(Z_l, U) \right]_{l=1, \dots, L},$$

where

$$\begin{aligned} & C^{(2)}(Z_l, U) \\ &= -\frac{1}{\sqrt{NT}} \left[\text{tr} \left(U \mathbb{M}_{f^0} U' \mathbb{M}_{\lambda^0} Z_l f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} \right) + \text{tr} \left(U' \mathbb{M}_{\lambda^0} U \mathbb{M}_{f^0} Z_l' \lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'} \right) \right. \\ & \quad \left. + \text{tr} \left(U' \mathbb{M}_{\lambda^0} Z_l \mathbb{M}_{f^0} U' \lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'} \right) \right]. \end{aligned}$$

MSW showed that under Assumption 3.1, as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow \kappa^2$, where $0 < \kappa < \infty$, we can approximate

$$\sqrt{NT}(\hat{\alpha} - \alpha_0) = (GWG')^{-1} GW \left[\frac{1}{\sqrt{NT}} (z^{\lambda f})' u + c^{(2)} \right] + o_p(1). \quad (5)$$

Notice that as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow \kappa^2$, where $0 < \kappa < \infty$, under Assumptions 3.1 we can show that

$$\frac{1}{\sqrt{NT}} (z^{\lambda f})' u + c^{(2)} \Rightarrow N(-\kappa b, \Omega), \quad (6)$$

where $b = (b_1, \dots, b_L)'$, and

$$\begin{aligned} b_l &= \text{plim}_{N, T \rightarrow \infty} \frac{1}{N} \text{tr} \left[\mathbb{P}_{f^0} \left[E(\epsilon' \tilde{\epsilon}_{-l-1}) + E((\eta - \alpha_0 \eta_{-1})' \eta_{-l-1}) \right] \right] \\ & \quad + \text{plim}_{N, T \rightarrow \infty} \frac{1}{N} \text{tr} \left[\mathbb{E}(U' U) \mathbb{M}_{f^0} \tilde{f}_{-l-1}^0 (f^{0'} f^0)^{-1} f^{0'} \right], \\ \tilde{\epsilon}_{-l} &= [\tilde{\epsilon}_{it-l}]_{it}, \quad \tilde{\epsilon}_{it-l} = \sum_{s=0}^{\infty} \alpha_0^s \epsilon_{it-l-s}, \quad \tilde{f}_{-l-1}^0 = \left[\sum_{s=0}^{\infty} \alpha_0^s f_{t-1-l-s} \right]_t, \\ \Omega &= \left(\sigma_\epsilon^2 + (1 - \alpha_0)^2 \sigma_\eta^2 \right) \left\{ \frac{\sigma_\epsilon^2}{1 - \alpha_0^2} \begin{bmatrix} 1 & \cdots & \alpha_0^{L-1} \\ \vdots & \ddots & \vdots \\ \alpha_0^{L-1} & \cdots & 1 \end{bmatrix} + \sigma_\eta^2 I_L \right\}. \end{aligned}$$

Combining (5) and (6), we have the following theorem.

Theorem 3.2 *Suppose that Assumptions 3.1 hold. As $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow \kappa^2$ and $0 < \kappa < \infty$, we have*

$$\sqrt{NT}(\hat{\alpha} - \alpha_0) \Rightarrow \mathcal{N}\left(-\kappa(GWG')^{-1}GWb, (GWG')^{-1}GW\Omega WG'(GWG')^{-1}\right).$$

Notice that the bias b in the limit distribution is due to the incidental parameters $\lambda_i^0 f_t^0$ and the lagged dependent variables as IVs, which is similar to the bias in Moon and Weidner (2010). This bias can be consistently estimated and is correctable, for details we refer to Moon and Weidner (2010) and Moon, Shum, and Weidner (2011).

4 Monte Carlo Simulations

In this section we investigate the finite sample properties of the LS-MD estimator $\hat{\alpha}$ through small scale Monte Carlo simulations. The data generating process is

$$\begin{aligned} Y_{it}^* &= \alpha_0 Y_{it-1}^* + \lambda_i^0 f_t^0 + \epsilon_{it}, \\ Y_{it} &= Y_{it}^* + \eta_{it}, \end{aligned}$$

where $\alpha_0 \in \{0.2, 0.5, 0.8\}$, $\{\lambda_i\}, \{f_t\}, \{\eta_{it}\} \sim iid N(0, 0.4)$ and $\{\epsilon_{it}\} \sim iid N(0, 1)$. We consider various combinations of $N \in \{20, 50, 100\}$ and $T \in \{20, 50, 100\}$. We use $Z_{it} = Y_{it-2}$ as an instrument. Notice that $\alpha_0 = 0.2$ violates the sufficient identification (4).

Table 1. Monte Carlo Simulation Results

N, T	$\alpha_0 = 0.2$			$\alpha_0 = 0.5$			$\alpha_0 = 0.8$		
	bias	s.d.	rmse	bias	s.d.	rmse	bias	s.d.	rmse
20,20	-0.173	0.694	0.715	-0.052	0.171	0.179	-0.030	0.090	0.095
20,50	-0.061	0.292	0.299	-0.004	0.077	0.077	-0.005	0.031	0.032
20,100	-0.005	0.168	0.168	-0.0004	0.055	0.055	-0.001	0.021	0.021
50,20	-0.129	0.440	0.458	-0.022	0.098	0.100	-0.015	0.061	0.063
50,50	-0.012	0.158	0.158	-0.003	0.048	0.048	-0.001	0.020	0.020
50,100	-0.007	0.102	0.102	-0.001	0.033	0.033	-0.0005	0.013	0.013
100,20	-0.092	0.303	0.316	-0.014	0.068	0.069	-0.014	0.057	0.059
100,50	-0.008	0.105	0.105	-0.003	0.034	0.034	-0.001	0.014	0.014
100,100	0.001	0.067	0.067	-0.0001	0.023	0.023	-0.0003	0.009	0.009

The finite sample properties of $\hat{\alpha}$, obtained in simulations with 1000 repetitions, are reported in Table 1. Except for the case of $\alpha_0 = 0.2$ with small samples, the LS-MD estimator $\hat{\alpha}$ performs well in finite samples.⁹ When $\alpha_0 = 0.2$, the finite sample properties improve as either N and T increases.

5 Discussions

Choice of Instrumental Variables: It is well known in the GMM literature that the choice of moment conditions — the choice of the lag length (L) in our setup — is one of the important factors that affect the finite sample properties of the GMM estimator. Various moment condition selection procedures have been proposed in the literature. These include, for example, the minimization of the (higher order) approximated mean squared error (e.g., Donald and Newey (2001), Okui (2009), and Kuersteiner (2010)) or of the asymptotic coverage error (e.g., Okui (2009)). However, it is not straightforward to apply these procedures to the LS-MD estimator. First, the LS-MD estimator has a bias even in the first order approximation. Secondly, the key approximation techniques used in the literature (e.g., Nagar’s expansion and the Edgeworth expansion) are not available in the existing literature for the LS-MD estimator. Developing a procedure for selection of L is therefore beyond the scope of this paper.

Extensions: Our LS-MD estimation can be used for more sophisticated cases. We briefly discuss how to extend our simple model.

1. **Inclusion of covariates:** The LS-MD estimation procedure can be easily extended to include a model with other exogenous regressors, say X_{it} . For example, in the first step one can regress $Y_{it} - \alpha Y_{it-1}$ on X_{it} , Z_{it} with interactive fixed effects $\lambda_i f_t$ for fixed α . In the second step, minimize $\hat{\gamma}(\alpha)' W_{NT}^\gamma \hat{\gamma}(\alpha)$ w.r.t. α .
2. **Heteroskedastic error:** Until now, we assume that the errors ϵ_{it} and η_{it} are homoskedastic for simplicity. If the errors are heteroskedastic, then the term $c^{(2)}$ contributes additional bias terms to the limit distribution of $\hat{\alpha}$. These biases are correctable (see e.g. Bai, 2009, and Moon and Weidner, 2010).

⁹We also investigated the finite sample properties of the bias corrected estimator and found that analytical bias correction simultaneously reduces the bias and the standard deviation of the estimator, except when both α^0 and T are small. We omit the detailed results due to space limitation, and since the biases in Table 1 without bias correction are already quite small relative to the corresponding standard deviations.

3. **Non-classical measurement error:** Measurement error so far is assumed to be classical. In many applications, however, measurement error can be correlated with the unobserved latent variable and the covariates. Our estimation method is still valid under more general measurement error models. For example, suppose that people tend to report income Y_{it} proportionally to Y_{it}^* as

$$Y_{it} = \gamma_{0i} + \gamma_{1it}Y_{it}^* + v_{it}, \quad (7)$$

where v_{it} is an unobserved error. Note that the measurement error in model (7) is non-classical since the measurement error, $\eta_{it} = Y_{it} - Y_{it}^*$, could be correlated with Y_{it}^* and the mean of the measurement error is not necessarily zero.¹⁰ Model (7) is a modified version of a linear measurement error model that allows for a heterogeneous relationship between Y_{it} and Y_{it}^* across cross-section and over time.¹¹ When the coefficient γ_{1it} is random satisfying $\gamma_{1it} = \gamma_1 + w_{it}$, where $\{w_{it}\}$ and $\{v_{it}\}$ are *iid* across i and over t with zero mean, and $\{w_{it}\}, \{v_{it}\}, \{\epsilon_{it}\}$ are independent of each other, then we have the following dynamic equation with two factors (or one factor and a time invariant fixed effect) as

$$Y_{it} = \alpha Y_{it-1} + \delta_i' h_t + U_{it},$$

where $\delta_i = (\gamma_1 \lambda_i, (1 - \alpha) \gamma_{0i})'$, $h_t = (f_t, 1)'$ and

$$U_{it} = \gamma_1 \epsilon_{it} + v_{it} - \alpha v_{it-1} + Y_{it}^* w_{it} - \alpha Y_{it-1}^* w_{it-1}. \quad (9)$$

Note that the composite error U_{it} in (9) has serial dependence structure similar to an $MA(1)$ process, and in this case $Z_{it} = (Z_{1,it}, \dots, Z_{L,it})' = (Y_{it-2}, \dots, Y_{it-1-L})'$ still remains uncorrelated with U_{it} .

¹⁰A special case of model (7) is $\gamma_{0it} = 0$ and $\gamma_{1it} = 1$, in which case v_{it} is classical.

¹¹Bollinger and Chandra (2005) and Kim and Solon (2005) developed a model allowing for a constant linear relationship between Y_{it} and Y_{it}^* , based on the evidence in surveyed income; i.e., those who earn higher than average tend to report their earning less, while those who earn lower than average tend to report higher. See also Bound, Brown and Mathiowetz (2001).

References

- [1] Bai, J. (2009): Panel Data Models with Interactive Fixed Effects, *Econometrica*, 77, 1229-1279.
- [2] Bollinger, C and A. Chandra (2005): Iatrogenic specification error: a cautionary tale of cleaning data, *Journal of Labor Economics* 23(2), 235-257.
- [3] Bound, J, C. C. Brown, and N. Mathiowetz (2001): Measurement Error in Survey Data. *Handbook of Labor Economics*, 5, 3705-3843.
- [4] Chernozhukov, V. and C. Hansen (2006): Instrumental Quantile Regression Inference for Structural and Treatment Effect Models, *Journal of Econometrics*, 132, 491-525.
- [5] Chernozhukov, V. and C. Hansen (2008): Instrumental Variable Quantile Regression: A Robust Inference Approach, *Journal of Econometrics*, 142, 379-398.
- [6] Donald, S. and W. Newey (2001): Choosing the Number of Instruments, *Econometrica*, 69, 1161-1191.
- [7] Hahn, J. and G. Kuersteiner (2002): Asymptotically Unbiased Inference of a Dynamic Panel Model with Fixed Effects When Both n and T Are Large, *Econometrica*, 70, 1639-1657.
- [8] Harding, M. and C. Lamarche (2011): Least Squares Estimation of a Panel Data Model with Multifactor Error Structure and Endogenous Covariates, *Economic Letters*, 111, 197-199.
- [9] Inoue, A. (2006): A Bootstrap Approach to Moment Selection, *Econometrics Journal*, 9, 48-75.
- [10] Kim, B and G. Solon (2005): Implications of mean-reverting measurement error for longitudinal studies of wages and employment. *Review of Economics and Statistics* 87 (1), 193-196.
- [11] Kuersteiner, G. (2010): Kernel Weighted GMM Estimators for Linear Time Series Models, Working Paper, Georgetown University.
- [12] Moon, H.R. and M. Shum, M. Weidner (2012): Interactive Fixed Effects in the BLP Random Coefficients Demand Model, *Working Paper*, Univ. of Southern California.

- [13] Moon, H.R. and M. Weidner (2010): Dynamic Panel Linear Regression Models with Interactive Fixed Effects, *Working Paper*, Univ. of Southern California.
- [14] Okui, R. (2009): Optimal Choice of Moments in Dynamic Panel Data Models, 151, 1-16.

6 Supplementary Appendix (Not for Publication)

6.1 Proof of Consistency

We show that Assumptions 3.1 and 3.2 in the current model are sufficient for Assumption 1 of Moon, Shum, and Weidner (2012) (MSW hereafter) with $\delta(\alpha)$ in MSW replaced by $Y - \alpha Y_{-1}$, and X_k in MSW replaced by 0.

Notation: When A is a matrix, $\|A\|^2$ denotes the largest eigenvalue of $A'A$ and $\|A\|_F^2$ denotes the trace of $A'A$.

- Assumption 1(i) holds since uniformly in α outside of any neighborhood of α_0 we have

$$\frac{\|\delta(\alpha) - \delta(\alpha_0)\|_F}{\|\alpha - \alpha_0\|} = \|Y_{-1}\|_F = \sqrt{\sum_{i=1}^N \sum_{t=1}^T Y_{it-1}^2} = O_p(\sqrt{NT}).$$

Also, it follows that

$$\|Z_l\|_F = \sqrt{\sum_{i=1}^N \sum_{t=1}^T Y_{it-1-l}^2} = O_p(\sqrt{NT}).$$

- Assumption 1(ii) is satisfied because $\|U\| = \|\epsilon + \eta - \alpha_0 \eta_{-1}\| \leq \|\epsilon\| + \|\eta\| + |\alpha_0| \|\eta_{-1}\| = O_p(\sqrt{\max(J, T)})$ because $\{\epsilon_{it}\}, \{\eta_{it}\} \sim iid$ with mean zero and finite moments higher than 4 (See Moon and Weidner (2010)).

- Assumption 1(iii)

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Y_{it-1-l} U_{it} = o_p(1)$$

follows for $l \geq 1$ since $E(Y_{it-1-l} U_{it}) = 0$ if $l \geq 1$.

- Assumption 1(iv) follows since any (nontrivial) linear combinations of Z'_l s have rank higher than two under Assumption 3.1.
- Assumption 1(v) holds by Assumption 3.2 with $\Delta \xi_{\alpha, \beta} = -(\alpha - \alpha_0) y_{-1}$.
- Assumption 1(vi) holds by Assumption 3.1 (iii).

6.2 Asymptotic Normality

- Assumptions 2 and 3 in MSW follow immediately under Assumption 3.1.
- Assumptions 4(i),(ii) and 5 in MSW follow since in this paper $\delta(\alpha) = Y - \alpha Y_{-1}$ is linear in α and by the conditions in Assumption 3.1.
- Assumption 4(iv) of MSW is satisfied with

$$Z_{l,it}^{\text{str}} = \lambda_i \sum_{s=0}^{\infty} \alpha_0^s f_{t-1-l-s}$$

$$Z_{l,it}^{\text{weak}} = \sum_{s=0}^{\infty} \alpha_0^s \epsilon_{it-1-l-s} + \eta_{it-1-l}.$$

- Notice that the conditions in Assumption 4(iii) of MSW are satisfied except for that U_{it} is an MA(1) type error over time, that is, U_{it} and U_{it-1} are dependent, while U_{it} and U_{it-s} are independent for $s \geq 2$. Because of this, we need to modify the proof of Theorem 5.2 of MSW and in what follows we give a sketch.
- Step 1: First we show that

$$\sqrt{N}(\hat{\alpha} - \alpha_0) = O_p(1).$$

- Step 2: Using the asymptotic likelihood expansion derived in Moon and Weidner (2010), we can approximate $\sqrt{NT}\hat{\gamma}(\alpha)$ as a linear function of $\sqrt{NT}(\alpha - \alpha_0)$;

$$\sqrt{NT}\hat{\gamma}(\alpha) = \left(\frac{1}{NT} z^{\lambda f'} z^{\lambda f} \right)^{-1} \left[\frac{1}{\sqrt{NT}} z^{\lambda f'} u + c^{(2)} - z^{\lambda f'} y_{-1}^{\lambda f} \sqrt{NT}(\alpha - \alpha_0) \right] + o_p(1)$$

where $o_p(1)$ holds uniformly in α with $\sqrt{N}|\alpha - \alpha_0| < c$ for all c .

- Step 3: We then approximate the second step objective function as a quadratic function of $\sqrt{NT}(\alpha - \alpha_0)$ by plugging the linear approximation of $\sqrt{NT}\hat{\gamma}(\alpha)$. Then,

we deduce that

$$\begin{aligned}
& \sqrt{NT}(\hat{\alpha} - \alpha_0) \\
&= \left[\left(\frac{1}{NT} y_{-1}^{\lambda f'} z^{\lambda f} \right) \left(\frac{1}{NT} z^{\lambda f'} z^{\lambda f} \right)^{-1} W_{NT}^\gamma \left(\frac{1}{NT} z^{\lambda f'} z^{\lambda f} \right)^{-1} \left(\frac{1}{NT} z^{\lambda f'} y_{-1}^{\lambda f} \right) \right]^{-1} \\
&\times \left(\frac{1}{NT} y_{-1}^{\lambda f'} z^{\lambda f} \right) \left(\frac{1}{NT} z^{\lambda f'} z^{\lambda f} \right)^{-1} W_{NT}^\gamma \left(\frac{1}{NT} z^{\lambda f'} z^{\lambda f} \right)^{-1} \left[\frac{1}{\sqrt{NT}} z^{\lambda f'} u + c^{(2)} \right] + o_p(1) \\
&= (GWG')^{-1} GW \left(\frac{1}{\sqrt{NT}} z^{\lambda f'} u + c^{(2)} \right) + o_p(1),
\end{aligned}$$

as required for (5). (Notice that Steps 1,2,3 are not affected by the MA(1) type dependence of U_{it} .)

- Step 4: By definition

$$\begin{aligned}
\frac{1}{\sqrt{NT}} z^{\lambda f'} u &= \begin{bmatrix} \frac{1}{\sqrt{NT}} \text{tr}(\mathbb{M}_{f_0} U' \mathbb{M}_{\lambda_0} Z_1) \\ \vdots \\ \frac{1}{\sqrt{NT}} \text{tr}(\mathbb{M}_{f_0} U' \mathbb{M}_{\lambda_0} Z_L) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{NT}} \text{tr}(\mathbb{M}_{f_0} U' \mathbb{M}_{\lambda_0} Z_1^{\text{weak}}) \\ \vdots \\ \frac{1}{\sqrt{NT}} \text{tr}(\mathbb{M}_{f_0} U' \mathbb{M}_{\lambda_0} Z_L^{\text{weak}}) \end{bmatrix} \\
&= \frac{1}{\sqrt{NT}} \sum_{i=1}^T \sum_{t=1}^T U_{it} Z_{it}^{\text{weak}} - \left[\frac{1}{\sqrt{NT}} \text{tr}(\mathbb{P}_{f_0} E(U' Z_l^{\text{weak}})) \right]_{l=1, \dots, L} \\
&\quad - \left[\frac{1}{\sqrt{NT}} \text{tr}(\mathbb{P}_{f_0} (U' Z_l^{\text{weak}} - E(U' Z_l^{\text{weak}}))) \right]_{l=1, \dots, L} \\
&\quad + \left[\frac{1}{\sqrt{NT}} \text{tr}(U' \mathbb{P}_{\lambda_0} Z_l^{\text{weak}}) \right]_{l=1, \dots, L} + \left[\frac{1}{\sqrt{NT}} \text{tr}(\mathbb{P}_{f_0} U' \mathbb{P}_{\lambda_0} Z_l^{\text{weak}}) \right]_{l=1, \dots, L} \\
&= I + II + III + IV + V, \text{ say.}
\end{aligned}$$

Then, by the CLT (e.g., Moon and Phillips (1999)) we have

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^T \sum_{t=1}^T U_{it} Z_{it}^{\text{weak}} \Rightarrow N(0, \Omega),$$

where

$$\begin{aligned}\Omega &= E \left(U_{it}^2 \left(Z_{it}^{\text{weak}} \right) \left(Z_{it}^{\text{weak}} \right)' \right) + E \left(U_{it} U_{it-1} \left(Z_{it}^{\text{weak}} \right) \left(Z_{it-1}^{\text{weak}} \right)' \right) \\ &\quad + E \left(U_{it} U_{it+1} \left(Z_{it}^{\text{weak}} \right) \left(Z_{it+1}^{\text{weak}} \right)' \right).\end{aligned}$$

A direct calculation shows that

$$\begin{aligned}& E \left(U_{it}^2 \left(Z_{it}^{\text{weak}} \right) \left(Z_{it}^{\text{weak}} \right)' \right) \\ &= E \left(U_{it}^2 \right) E \left[\left(Z_{it}^{\text{weak}} \right) \left(Z_{it}^{\text{weak}} \right)' \right] \\ &= \left(\sigma_\epsilon^2 + (1 + \alpha_0^2) \sigma_\eta^2 \right) \left\{ \frac{\sigma_\epsilon^2}{1 - \alpha_0^2} \begin{bmatrix} 1 & \cdots & \alpha_0^{L-1} \\ \vdots & \ddots & \vdots \\ \alpha_0^{L-1} & \cdots & 1 \end{bmatrix} + \sigma_\eta^2 I_L \right\}\end{aligned}$$

and

$$\begin{aligned}& E \left(U_{it} U_{it-1} \left(Z_{it}^{\text{weak}} \right) \left(Z_{it-1}^{\text{weak}} \right)' \right) + E \left(U_{it} U_{it+1} \left(Z_{it}^{\text{weak}} \right) \left(Z_{it+1}^{\text{weak}} \right)' \right) \\ &= -2\alpha_0 \sigma_\eta^2 \left\{ \frac{\sigma_\epsilon^2}{1 - \alpha_0^2} \begin{bmatrix} 1 & \cdots & \alpha_0^{L-1} \\ \vdots & \ddots & \vdots \\ \alpha_0^{L-1} & \cdots & 1 \end{bmatrix} + \sigma_\eta^2 I_L \right\},\end{aligned}$$

which leads

$$\Omega = \left(\sigma_\epsilon^2 + (1 - \alpha_0)^2 \sigma_\eta^2 \right) \left\{ \frac{\sigma_\epsilon^2}{1 - \alpha_0^2} \begin{bmatrix} 1 & \cdots & \alpha_0^{L-1} \\ \vdots & \ddots & \vdots \\ \alpha_0^{L-1} & \cdots & 1 \end{bmatrix} + \sigma_\eta^2 I_L \right\}$$

Also, a direct calculation shows that

$$\begin{aligned}& - \text{plim}_{N, T \rightarrow \infty} \frac{1}{\sqrt{NT}} \text{tr} \left(\mathbb{P}_{f_0} E \left(U' Z_l^{\text{weak}} \right) \right) \\ &= -\kappa \text{plim}_{N, T \rightarrow \infty} \frac{1}{N} \text{tr} \left[\mathbb{P}_{f_0} \left[E \left(\epsilon' \tilde{\epsilon}_{-l-1} \right) + E \left((\eta - \alpha_0 \eta_{-1})' \eta_{-l-1} \right) \right] \right].\end{aligned}$$

By modifying Lemma C.2 (f), (j), and (m) in Moon and Weidner (2010), we can show that

$$III, IV, V = o_p(1).$$

- Step 5: By modifying Lemma C.2 (c), (d), (e), (g), (h), (i), (k), and (l) in Moon and Weidner (2010), we can show that

$$\text{plim}_{N, T \rightarrow \infty} c^{(2)} = -\kappa \left[\text{plim}_{N, T \rightarrow \infty} \frac{1}{N} \text{tr} \left[\mathbb{E}(U'U) \mathbb{M}_{f^0} \tilde{f}_{-l-1}^0 (f^{0l} f^0)^{-1} f^{0l} \right] \right]_{l=1, \dots, L}.$$

- Step 6: Combining the limits in Steps 4 and 5 yields the desired result in (6).

6.3 Sufficient Conditions for Assumption 3.2

In matrix notation we can write (3) as

$$Y = \alpha^0 Y_{-1} + \lambda^0 f^{0l} + U.$$

By recursively applying the model we find

$$Y = \lambda^0 F^{0l} + E + Y^{\text{init}},$$

where F is the $T \times 1$ vector with entries $F_t = \sum_{\tau=0}^{t-1} \alpha_0^\tau f_{t-\tau}^0$, and E and Y^{init} are the $T \times N$ matrices with entries $E_{it} = \eta_{it} + \sum_{\tau=0}^{t-1} \alpha_0^\tau \epsilon_{t-\tau}$, and $Y_{it}^{\text{init}} = \alpha_0^t Y_{i0}$. We denote lagged versions of F^0 and E by F_{-1}^0 and E_{-1} , etc.

In the following we assume $L = 1$. In that case Assumption 3.2 is satisfied if

$$\frac{\left(\text{plim}_{N, T \rightarrow \infty} \frac{1}{NT} y'_{-1} z \right)^2}{\text{plim}_{N, T \rightarrow \infty} \frac{1}{NT} z' z} - \text{plim}_{N, T \rightarrow \infty} \left(\max_{\lambda} \frac{1}{NT} y'_{-1} \mathbb{P}_{I_T \otimes \tilde{\lambda}} y_{-1} \right) > 0, \quad (10)$$

where $\tilde{\lambda} = (\lambda^0, \lambda)$. If f_t is mean zero and independent across t we have

$$\begin{aligned}
\frac{1}{NT} y'_{-1} z &= \frac{1}{NT} \text{tr}(Y_{-1} Y_{-2}) \\
&= \frac{1}{NT} \text{tr}(E_{-1} E_{-2}) + \frac{1}{NT} \|\lambda\|^2 (F'_{-1} F_{-2}) + o_p(1) \\
&= \frac{\alpha_0}{1 - \alpha_0^2} (\sigma_\epsilon^2 + \Sigma_\lambda \Sigma_f) + o_p(1), \\
\frac{1}{NT} z' z &= \frac{1}{NT} \text{tr}(Y_{-1} Y_{-1}) \\
&= \frac{1}{1 - \alpha_0^2} (\sigma_\epsilon^2 + \Sigma_\lambda \Sigma_f) + \sigma_\eta^2 + o_p(1),
\end{aligned}$$

and

$$\begin{aligned}
\max_{\lambda} \frac{1}{NT} y'_{-1} \mathbb{P}_{I_T \otimes \tilde{\lambda}} y_{-1} &= \max_{\lambda} \frac{1}{NT} \text{tr}(Y'_{-1} \mathbb{P}_{\tilde{\lambda}} Y_{-1}) \\
&= \max_{\lambda} \frac{1}{NT} \text{tr}(F_{-1} \lambda' \mathbb{P}_{\tilde{\lambda}} \lambda F'_{-1}) + o_p(1) \\
&= \frac{1}{NT} \|\lambda\|^2 \|F_{-1}\|^2 + o_p(1) \\
&= \frac{1}{1 - \alpha_0^2} \Sigma_\lambda \Sigma_f + o_p(1).
\end{aligned}$$

Plugging these results into condition (10) yields condition (4).