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 $Y^\ast_{it}$, contains measurement error:

\begin{equation}
Y^\ast_{it} = \alpha_0 Y^\ast_{it-1} + \lambda f_t^{0} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,
\end{equation}

\begin{equation}
Y_{it} = Y^\ast_{it} + \eta_{it}.
\end{equation}

∗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 $\eta_{it}$ represents measurement error. The term $\lambda^0_i f^0_t$ describes unobserved interactive fixed effects.\footnote{In this paper, we consider a single factor, that is, the dimensions of $f_t$ and $\lambda_i$ are equal to one. The extension to the multiple factor case is straightforward, but omitted due to space limitation.} The goal of the paper is to estimate $\alpha_0$ when both the number of individuals $N$ and the number of time periods $T$ are large.\footnote{When interpreting $\lambda^0_i$ as individual specific fixed effects, the term $f^0_t$ represents the (time-varying) linear projection coefficient of $Y^*_i$ on $\lambda^0_i$ (holding $Y^*_{it-1}$ constant). This allows the effect of the unobserved individual characteristic $\lambda^0_i$ on $Y^*_i$ to be time-varying. Alternatively, one can interpret $f^0_t$ as a common time specific shock (a common factor) and $\lambda^0_i$ then describes reaction to the common shock (a factor loading).}

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

$$Y_{it} = \alpha_0 Y_{it-1} + \lambda^0_i f^0_t + U_{it},$$

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 $Y^*_i$ 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 $\alpha_0$. One is the presence of the interactive fixed effects $\lambda^0_i f^0_t$ 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 step\footnote{We consider large $N, T$ approximations to characterize the bias due to the incidental parameters $\lambda^0_i f^0_t$, see e.g. Bai (2009) and Hahn and Kuersteiner (2004).}. 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.

An 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 $\eta_{it-1}$. 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 $Z_{it} = (Z_{1,it}, ..., Z_{L,it})' = (Y_{it-2}, ..., Y_{it-1-L})'$ for the IVs of the endogenous regressor $Y_{it-1}$. The question, then, is how to use the instrumental variables $Z_{it}$ to estimate $\alpha_0$ in the presence of interactive fixed effects $\lambda_0^i f_t^0$ 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 $\alpha$, 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

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

where $\gamma = (\gamma_1, ..., \gamma_L)', \lambda = (\lambda_1, ..., \lambda_N)'$ and $f = (f_1, ..., f_T)'$. Step 2: For some positive definite weight matrix $W_{NT}^\gamma$, we estimate $\alpha$ 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 $W_{NT}^\gamma$. 

3
3 Asymptotic Results

Assumption 3.1 (i) The unobserved error terms \( \{\epsilon_{it}\} \sim iid (0, \sigma^2) \) and \( \{\eta_{it}\} \sim iid (0, \sigma^2_n) \) 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 \( f^0_t \) are strictly stationary and ergodic with sup, \( |f_t| < \infty \) and \( \frac{1}{T} \sum_{t=1}^{T}(f^0_t)^2 \to_p \Sigma_f > 0 \), and \( \lambda_i \) are iid with sup, \( |\lambda_i| < \infty \) and \( \frac{1}{N} \sum_{i=1}^{N}(\lambda^0)^2 \to_p \Sigma_\lambda > 0 \). Also assume that \( \{f^0_t\}, \{\lambda^0_i\}, \{\epsilon_{it}\}, \{\eta_{it}\} \) are independent. (iii) \( W_{NT} \to_p W^* > 0 \). (vi) \( |\alpha_0| < 1 \) and \( \alpha_0 \neq 0 \).

The iid assumptions of \( \epsilon_{it} \) and \( \eta_{it} \) 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 \( \eta_{it} \) is classical in the sense that \( \eta_{it} \) 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 \( P_A = A (A' A)^{-1} A' \) and \( M_A = I - P_A \). 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 \( \lambda^0 = (\lambda^0_1, ..., \lambda^0_N)' \) and \( f^0 = (f^0_1, ..., f^0_T)' \). 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 \( \frac{1}{NT} y'_{-1} P_z y_{-1} - \max_{\lambda} \frac{1}{NT} y'_{-1} P_{I_T \otimes \lambda} y_{-1} > c \) with probability approaching one as \( N, T \to \infty \), where \( \lambda = (\lambda^0, \lambda) \).

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 \( \frac{1}{NT} y'_{-1} P_z y_{-1} \), is larger than the joint explanatory power for \( Y_{it-1} \) of the true factor loading \( \lambda^0 \) together with any other factor loading \( \lambda \), given by \( \frac{1}{NT} y'_{-1} P_{I_T \otimes \lambda} y_{-1} \). If there are no interactive fixed effects included in the model, then the assumption simplifies to the standard relevance condition \( \frac{1}{NT} y'_{-1} P_z y_{-1} > 0 \), which is satisfied for \( \alpha^0 \neq 0 \).

Suppose that Assumption 3.1 holds, and consider the special case where \( f^0_t \) has mean
zero and is distributed independently over $t$. Then, Assumption 3.2 is equivalent to

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

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 $f^0_t$ 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. Note that the lower bound in (4) goes to zero when $\Sigma_{\lambda} \Sigma_f$ becomes small relative to $\sigma_\epsilon^2$, 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 $\hat{\alpha} \to_p \alpha_0$ as $N,T \to \infty$.

To present the limiting distribution of $\hat{\alpha}$, we need to introduce some further notation. Define the $NT$-vectors $y_{-1}^{\lambda f}$ and $z_{-1}^{\lambda f}$ by $y_{-1}^{\lambda f} = \text{vec} (M_{\lambda 0} Y_{-1} M_{f 0})$, and $z_{-1}^{\lambda f} = \text{vec} (M_{\lambda 0} Z_t M_{f 0})$, where $l = 1, \ldots, L$. Let $u = \text{vec} (U)$ and $z^{\lambda f} = (z_1^{\lambda f}, \ldots, z_L^{\lambda f})$.

Define $G = \text{plim}_{N,T \to \infty} \frac{1}{NT y_{-1}^{\lambda f}'} z_{-1}^{\lambda f} = \sigma_\epsilon^2 \left( \begin{array}{c} 1 \\ \vdots \\ \alpha_{L-1}^{\lambda f} \end{array} \right)$, and

$$
W = \text{plim}_{N,T \to \infty} \left( \frac{1}{NT} z^{\lambda f} z^{\lambda f} \right)^{-1} W_N^T \left( \frac{1}{NT} z^{\lambda f} z^{\lambda f} \right)^{-1} W_N = \left\{ \begin{array}{c} 0 \\ \vdots \\ 1 \end{array} \right\} W_N^T \left( \begin{array}{c} \sigma_\epsilon^2 \\ \vdots \\ \sigma_\eta^2 \end{array} \right) \left( \begin{array}{ccc} 1 & \cdots & \alpha_{L-1}^{\lambda f} \\ \vdots & \ddots & \vdots \\ \alpha_{L-1}^{\lambda f} & \cdots & 1 \end{array} \right) + \sigma_\eta^2 I_L.
$$

---

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 $f^0_t$ can result in situations where Assumption 3.2 is not satisfied for any value of $\alpha_0$. The assumption that $f^0_t$ 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

$$c(2) = \left[ C(2) (Z_t, U) \right]_{t=1,\ldots, L},$$

where

$$C(2) (Z_t, U) = -\frac{1}{\sqrt{NT}} \left[ \text{tr} \left( UMf_0 U' M^0 Z_l f^0 (f^0 f^0)^{-1} (M^0)^{-1} \lambda^0 \right) + \text{tr} \left( U' M^0 U M f_0 Z_l^0 \left( (M^0)^{-1} (f^0 f^0)^{-1} f^0 \right) \right) + \text{tr} \left( U' M^0 Z_l M f_0 U M f_0 Z_l^0 \left( (f^0 f^0)^{-1} f^0 \right) \right) \right].$$

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

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

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

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

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

$$b_t = \text{plim} \frac{1}{N} \text{tr} \left[ P_{f,0} \left[ E (\epsilon' \tilde{\epsilon}_{-l-1}) + E (\lambda \eta \eta_{-l-1})' \eta_{-l-1} \right] \right] + \text{plim} \frac{1}{N} \text{tr} \left[ E (U' U) M f_0 f_0' (f^0 f^0)^{-1} f^0 \right],$$

$$\tilde{\epsilon}_{-l} = [\tilde{\epsilon}_{-l-1}, \tilde{\epsilon}_{-l-1}]_t, \quad \tilde{\epsilon}_{-l-1} = \sum_{s=0}^{\infty} \alpha_s \epsilon_{-l-1-s}, \quad \tilde{f}_{-l-1} = \sum_{s=0}^{\infty} \alpha_s f_{-l-1-s} ,$$

$$\Omega = \left( \sigma^2 + (1 - \alpha_0)^2 \sigma^2_\eta \right) \left\{ \begin{array}{ccc} \sigma^2_\epsilon & \cdots & \sigma^2_\eta L-1 \\ \alpha_0 L-1 & \cdots & 1 \end{array} \right\}.$$
Theorem 3.2 Suppose that Assumptions 3.1 hold. As \( N, T \to \infty \) with \( \frac{N}{T} \to \kappa^2 \) and \( 0 < \kappa < \infty \), we have

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

Notice that the bias \( b \) in the limit distribution is due to the incidental parameters \( \lambda_i f_t^0 \) 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

\[
Y_{it}^* = \alpha_0 Y_{it-1}^* + \lambda_i f_t^0 + \epsilon_{it},
\]

\[
Y_{it} = Y_{it}^* + \eta_{it},
\]

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>
<thead>
<tr>
<th>N,T</th>
<th>( \alpha_0 = 0.2 )</th>
<th>( \alpha_0 = 0.5 )</th>
<th>( \alpha_0 = 0.8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>bias</td>
<td>s.d.</td>
<td>rmse</td>
<td>bias</td>
</tr>
<tr>
<td>20,20</td>
<td>-0.173</td>
<td>0.694</td>
<td>0.715</td>
</tr>
<tr>
<td>20,50</td>
<td>-0.061</td>
<td>0.292</td>
<td>0.299</td>
</tr>
<tr>
<td>20,100</td>
<td>-0.005</td>
<td>0.168</td>
<td>0.168</td>
</tr>
<tr>
<td>50,20</td>
<td>-0.129</td>
<td>0.440</td>
<td>0.458</td>
</tr>
<tr>
<td>50,50</td>
<td>-0.012</td>
<td>0.158</td>
<td>0.158</td>
</tr>
<tr>
<td>50,100</td>
<td>-0.007</td>
<td>0.102</td>
<td>0.102</td>
</tr>
<tr>
<td>100,20</td>
<td>-0.092</td>
<td>0.303</td>
<td>0.316</td>
</tr>
<tr>
<td>100,50</td>
<td>-0.008</td>
<td>0.105</td>
<td>0.105</td>
</tr>
<tr>
<td>100,100</td>
<td>0.001</td>
<td>0.067</td>
<td>0.067</td>
</tr>
</tbody>
</table>
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 exiting 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 $\lambda_if_t$ for fixed $\alpha$. In the second step, minimize $\hat{\gamma}(\alpha)'W_{NT}\hat{\gamma}(\alpha)$ w.r.t. $\alpha$.

2. **Heteroskedastic error:** Until now, we assume that the errors $\epsilon_{it}$ and $\eta_{it}$ 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).

\(^9\)We 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 $\alpha_0$ 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

$$Y_{it} = \gamma_0 + \gamma_1 Y_{it}^* + v_{it},$$

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 $\gamma_1$ 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 $\delta_i = (\gamma_1 \lambda_i, (1 - \alpha) \gamma_0_i)'$, $h_t = (f_t, 1)'$ 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 $Z_{it} = (Z_{1, it}, ..., Z_{L, it})' = (Y_{it-2}, ..., Y_{it-1-L})'$ still remains uncorrelated with $U_{it}$.

---

\(^{10}\)A special case of model (7) is $\gamma_0_i = 0$ and $\gamma_{1it} = 1$, in which case $v_{it}$ is classical.

\(^{11}\)Bollinger 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


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, $\|A\|^2_2$ denotes the largest eigenvalue of $A'A$ and $\|A\|^2_F$ denotes the trace of $A'A$.

• Assumption 1(i) holds since uniformly in $\alpha$ outside of any neighborhood of $\alpha_0$ we have

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

Also, it follows that

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

• Assumption 1(ii) is satisfied because $\|U\| = \|\epsilon - \alpha_0 \eta_{-1}\| \leq \|\epsilon\| + \|\eta\| + |\alpha_0| \|\eta_{-1}\| = O_p\left(\sqrt{\max(J,T)}\right)$ 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 \( \alpha \) and by the conditions in Assumption 3.1.

- Assumption 4(iv) of MSW is satisfied with

\[
Z_{l,it}^{\text{str}} = \lambda_i \sum_{s=0}^{\infty} \alpha_s f_{t-1-l-s} \\
Z_{l,it}^{\text{weak}} = \sum_{s=0}^{\infty} \alpha_s e_{t-1-l-s} + \eta_{it-1-l}.
\]

- 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 \( \sqrt{NT} (\alpha - \alpha_0) \);

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

where \( o_p(1) \) holds uniformly in \( \alpha \) 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

\[
\sqrt{NT} (\hat{\alpha} - \alpha_0) = \left[ \left( \frac{1}{NT} y^{-1}_z \lambda_f z \right)^{-1} W_{NT}^\gamma \left( \frac{1}{NT} z \lambda_f z \right)^{-1} \left( \frac{1}{NT} z \lambda_f y \right)^{-1} \right]^{-1} \\
\times \left( \frac{1}{NT} y^{-1}_z \lambda_f z \right)^{-1} W_{NT}^\gamma \left( \frac{1}{NT} z \lambda_f z \right)^{-1} \left[ \frac{1}{\sqrt{NT}} z \lambda_f u + c^{(2)} \right] + o_p (1)
\]

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

\[
\frac{1}{\sqrt{NT}} z \lambda_f u = \left[ \frac{1}{\sqrt{NT}} \text{tr} \left( M_{f0} U' M_{\lambda 0} Z_1 \right) \right] = \left[ \frac{1}{\sqrt{NT}} \text{tr} \left( M_{f0} U' M_{\lambda 0} Z_L \right) \right] = \frac{1}{\sqrt{NT}} \sum_{i=1}^{T} \sum_{t=1}^{T} U_{it} Z_{it}^{\text{weak}} - \left[ \frac{1}{\sqrt{NT}} \text{tr} \left( P_{f0} E \left( U' Z_{i}^{\text{weak}} \right) \right) \right]_{l=1, \ldots, L} \\
- \left[ \frac{1}{\sqrt{NT}} \text{tr} \left( P_{f0} \left( U' Z_{i}^{\text{weak}} - E \left( U' Z_{i}^{\text{weak}} \right) \right) \right) \right]_{l=1, \ldots, L} \\
+ \left[ \frac{1}{\sqrt{NT}} \text{tr} \left( U' P_{\lambda 0} Z_{i}^{\text{weak}} \right) \right]_{l=1, \ldots, L} + \left[ \frac{1}{\sqrt{NT}} \text{tr} \left( P_{f0} U' P_{\lambda 0} Z_{i}^{\text{weak}} \right) \right]_{l=1, \ldots, L} = I + II + III + IV + V, \ \text{say.}
\]

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

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

\[ \Omega = E \left( U_{it}^2 \left( Z_{it}^{weak} \right) \left( Z_{it}^{weak} \right)' \right) + E \left( U_{it} U_{it-1} \left( Z_{it}^{weak} \right) \left( Z_{it-1}^{weak} \right)' \right) + E \left( U_{it} U_{it+1} \left( Z_{it}^{weak} \right) \left( Z_{it+1}^{weak} \right)' \right) \]

A direct calculation shows that

\[ E \left( U_{it}^2 \left( Z_{it}^{weak} \right) \left( Z_{it}^{weak} \right)' \right) = E \left( U_{it}^2 \left( Z_{it}^{weak} \right) \left( Z_{it}^{weak} \right)' \right) \]

\[ = E \left( \sigma_\varepsilon^2 \right) \left( \begin{array}{c} 1 \\ \vdots \\ \alpha_0^{L-1} \\ \vdots \\ \alpha_0^{L-1} \end{array} \right) + \sigma_\eta^2 I_L \]

and

\[ E \left( U_{it} U_{it-1} \left( Z_{it}^{weak} \right) \left( Z_{it-1}^{weak} \right)' \right) + E \left( U_{it} U_{it+1} \left( Z_{it}^{weak} \right) \left( Z_{it+1}^{weak} \right)' \right) \]

\[ = -2\alpha_0 \sigma_\eta^2 \left( \begin{array}{c} 1 \\ \vdots \\ \alpha_0^{L-1} \\ \vdots \\ \alpha_0^{L-1} \end{array} \right) + \sigma_\eta^2 I_L \]

which leads

\[ \Omega = \left( \sigma_\varepsilon^2 + (1 - \alpha_0)^2 \sigma_\eta^2 \right) \left( \begin{array}{c} 1 \\ \vdots \\ \alpha_0^{L-1} \\ \vdots \\ \alpha_0^{L-1} \end{array} \right) + \sigma_\eta^2 I_L \]

Also, a direct calculation shows that

\[ -\lim_{N,T \to \infty} \frac{1}{\sqrt{NT}} \text{tr} \left( \mathbb{P}_0 \mathbb{E} \left( U'Z_{t}^{weak} \right) \right) \]

\[ = -\kappa \lim_{N,T \to \infty} \frac{1}{N} \text{tr} \left[ \mathbb{P}_0 \left( \mathbb{E} \left( \epsilon' \tilde{\epsilon}_{t-1} \right) + \mathbb{E} \left( \left( \eta - \alpha_0 \eta_{t-1} \right)' \eta_{t-1} \right) \right) \right]. \]
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

\[ \text{plim}_{N,T \to \infty} \left( \frac{1}{N} \text{tr} \left[ E(U'U) M f_0 f_0' - 1 f_0' f_0 \right] \right)_{l=1,\ldots,L} = -\kappa. \]

- 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

\[ Y = \alpha_0 Y_{-1} + \lambda_0 f_0 + U. \]

By recursively applying the model we find

\[ Y = \lambda_0 F_0 + E + Y_{\text{init}}, \]

where \( F \) is the \( T \times 1 \) vector with entries \( F_t = \sum_{\tau=0}^{t-1} \alpha_0^\tau f_{0,t-\tau} \), and \( E \) and \( Y_{\text{init}} \) are the \( T \times N \) matrices with entries \( E_{it} = \eta_{it} + \sum_{\tau=0}^{t-1} \alpha_0^\tau \epsilon_{t-\tau} \), and \( Y^\text{init}_{it} = \alpha_0^t Y_{i0} \). We denote lagged versions of \( F_0 \) and \( E \) by \( F_{0,-1} \) and \( E_{-1} \), etc.

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

\[ \frac{\left( \text{plim}_{N,T \to \infty} \frac{1}{NT} y_{-1}' z \right)^2}{\text{plim}_{N,T \to \infty} \frac{1}{NT} z' z} - \text{plim}_{N,T \to \infty} \left( \max_{\lambda} \frac{1}{NT} y_{-1}' \mathbb{P} I_T \otimes \lambda y_{-1} \right) > 0, \]  

(10)
where $\tilde{\lambda} = (\lambda^0, \lambda)$. If $f_t$ is mean zero and independent across $t$ we have

$$
\frac{1}{NT}y_{-1}'z = \frac{1}{NT} \text{tr}(Y_{-1}Y_{-2}) \\
= \frac{1}{NT} \text{tr}(E_{-1}E_{-2}) + \frac{1}{NT} \|\lambda\|^2(F_{-1}'F_{-2}) + o_p(1) \\
= \frac{\alpha_0}{1 - \alpha_0^2} (\sigma^2 + \Sigma_{\lambda}\Sigma_f) + o_p(1),
$$

$$
\frac{1}{NT}z'z = \frac{1}{NT} \text{tr}(Y_{-1}Y_{-1}) \\
= \frac{1}{1 - \alpha_0^2} (\sigma_f^2 + \Sigma_{\lambda}\Sigma_f) + \sigma^2 + o_p(1),
$$

and

$$
\max_\lambda \frac{1}{NT} y_{-1}'\mathbb{P}_{\lambda} Y_{-1} = \max_\lambda \frac{1}{NT} \text{tr}(Y_{-1}'\mathbb{P}_\lambda Y_{-1}) \\
= \max_\lambda \frac{1}{NT} \text{tr}(F_{-1}\lambda'\mathbb{P}_\lambda \lambda F_{-1}') + o_p(1) \\
= \frac{1}{NT} \|\lambda\|^2\|F_{-1}\|^2 + o_p(1) \\
= \frac{1}{1 - \alpha_0^2} \Sigma_{\lambda}\Sigma_f + o_p(1).
$$

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