Initial-Condition Free Estimation of Fixed Effects Dynamic Panel Data Models

Zhenlin Yang*

School of Economics, Singapore Management University

Email: zlyang@smu.edu.sg

May 31, 2014

Abstract

It is well known that (quasi) MLE of dynamic panel data (DPD) models with short panels depends on the assumptions on the initial values; ignoring them or a wrong treatment of them will result in inconsistency or serious bias. This paper introduces a initial-condition free method for estimating the fixed-effects DPD models, through a simple modification of the quasi-score. An outer-product-of-gradients (OPG) method is also proposed for robust inference. The MLE of Hsiao, Pesaran and Tahmiscioglu (2002, Journal of Econometrics), where the initial observations are modeled, is extended to quasi MLE and an OPG method is proposed for robust inference. Consistency and asymptotic normality for both estimation strategies are established, and the two methods are compared through Monte Carlo simulations. The proposed method performs well in general, whether the panel is short or not. The quasi MLE performs comparably, except when model does not contain time-varying regressor, or the panel is not short and the dynamic parameter is small. The proposed method is much simpler and easier to apply.

Key Words: Bias reduction; Consistency; Asymptotic normality; Dynamic panel; Fixed effects; Modified quasi-score; Robust standard error; Short panel.

JEL classifications: C10, C13, C23, C15

1 Introduction

Fixed-effects dynamic panel data (FE-DPD) models covering short time periods have played over the last three decades an important role in empirical microeconometric research. Various estimation methods have been developed for the model, including maximum likelihood (ML) type estimators (see, e.g., Nickell 1981; Hsiao et al. 2002; Hsiao 2003, Ch. 4; Binder et al. 2005; Bun and Carree 2005; Phillips and Sul 2007; Gouriéroux et al. 2010; Kruiniger 2013), the instrumental variables (IV) estimators and generalized method-of-moments (GMM) estimators (see, e.g., Holtz-Eakin et al. 1988; Arellano and Bond 1991; Hahn 1997; Ahn and Schmidt 1995; Kiviet 1995; Blundell and Bond 1998; Alvarez and Arellano 2003).

*I am grateful to Singapore Management University for financial support under Grant C244/MSS12E007.

It is widely acknowledged that the consistent ML-type estimators of the FE-DPD model can be much more efficient than the consistent GMM-type estimators. However, the main difficulty in the ML estimation of the DPD models with short panels is the modeling of the initial observations of the response vector, say y_0. This is because y_0 may be exogenous in the sense that it varies autonomously, independent of other variables in the model; or endogenous in the sense that it is generated in the same way as the other values of the response vector y in the latter time periods. In case that y_0 is endogenous, it depends on the processes starting values and the past values of time-varying regressors, both of which are not observable, leading to incidental parameters. The traditional way of handling this problem is to predict these quantities using the observed values of the regressors (Anderson and Hsiao, 1980, 1981). In case of fixed effects models, the incidental parameters problem also occurs in the model itself (the fixed effects), but this problem can be resolved by first-differencing or some kind of orthogonal transformations.

Hsiao, Pesaran and Tahmiscioglu (2002) proposed an ML estimator (MLE) where the initial differences are explicitly modeled, Bun and Caree (2005) proposed a bias-corrected least squares dummy variable (LSDV) estimator, and Gouriéroux, et al. (2010) proposed an estimator based on indirect inference. In essence, all three estimators are likelihood based, and thus are more efficient than the GMM/IV estimators. All three estimators depend on normality assumption for inference. Besides, Gouriéroux, et al.'s estimator depends on the way that initial observations are generated (from a stationary process), and Bun and Caree's estimator, though does not depend explicitly on the way the initial observations are generated, requires either an iterative numerical root-finding process or a multi-dimensional root-finding process. Kruiniger (2013), following Chamberlain's (1980) formulation, showed that the MLE of the pure FE-DPD model can still be large-N consistent when the errors display arbitrary heterogeneity. Kruiniger called this the quasi MLE (QMLE). In our formulation, the word 'quasi' means that the errors may be nonnormal and that the likelihood may have ignored the initial observations.

In this paper, I propose a unified approach to estimate the PDP model with fixed effects based on the conditional quasi-likelihood, conditional on the initial differences Δy_1. Clearly, this conditional likelihood is incorrect whether exogenous y_0 is exogenous or endogenous, and thus maximizing this conditional likelihood may produce inconsistent or asymptotically biased estimators (see, e.g., Bhargava and Sargan, 1983). An analytical expression of the expectation of the quasi-score function is derived, which turns out to be independent of the unobservables and the process starting time, leading to a modified score function. Solving the modified normal equation leads to the modified QML-type estimators that are free from the specification of the distribution of the initial observations, and also robust against the misspecification of the error distributions. To make inferences about the model parameters, a robust variance-covariance (VC) matrix estimator is desired. As the modified score function can be written a sum of N independent terms, an outer-product-of-gradients (OPG) estimator of the variance of the modified score naturally emerges. This together with the estimated Hessian matrix gives a robust (sandwich) estimator of the VC matrix of the proposed estimator. Consistency and asymptotic normality of the proposed estimators and the validity of the robust VC matrix estimator are established. The proposed methods of model estimation and inference are both very simple, avoid explicit modeling of the initial observations, and thus can be easily used by applied researchers. Compared with the traditional methods where the full likelihood is used and the initial observations are modeled, there may be only a negligible loss in efficiency, but there a significant gain in the applicability. The proposed estimation method remains valid when the time dimension (T) goes large, and more importantly, it automatically corrects the bias caused by ignoring the initial observations under the large-N and large-T set up of, e.g, Hahn and Kuersteiner (2002).

As the proposed method is closely related to that of HPT, a comparison is necessary,. For this, I first extend HPT's ML estimation framework to quasi ML estimation framework, and provide formal results on consistency, asymptotic normality, and robust standard errors. Extensive Monte Carlo experiments are run and the results evidence an excellent finite sample performance of the proposed method, whether the panel is short or not. The quasi MLE performs comparably, except when model does not contain time-varying regressor, or the panel is not short and the dynamic parameter is small. The proposed method is much simpler and easier to apply.

Section 2 describes the model and some basic assumptions. Section 3 introduces the initial-condition free method of estimating the FE-DPD model, and presents formal theories on consistency, asymptotic normality and robust standard errors. Section 4 extends HPT's ML estimation framework, and presents a similar set of theoretical results. Section 5 presents partial Monte Carlo results, and Section 6 concludes the paper.

2 Model and Basic Assumptions

Consider the dynamic panel data (DPD) model of the form:

$$y_{it} = \rho y_{i,t-1} + x_{it}'\beta + z_i'\gamma + \mu_i + v_{it}, \quad i = 1, \ldots, N, \quad t = \ldots, -1, 0, 1, \ldots, T, \tag{2.1}$$

where the scalar parameter ρ characterizes the dynamic effect, x_{it} is a $p \times 1$ vector of time-varying exogenous variables, z_i is a $q \times 1$ vector of time-invariant exogenous variables that may include the constant term, dummy variables representing individuals' gender, race, etc., and β and γ are the usual regression coefficients. The $\{\mu_i\}$ are the non-observable individual effects and $\{v_{it}\}$ are the idiosyncratic errors, assumed to be independent and identically distributed (iid) across both i and t with mean zero and variance σ_v^2.[1]

[1] The iid assumption along the time dimension can be relaxed, e.g., for each i, $\{v_{it}\}$ follows an AR process.

Denoting $y_t = (y_{1t}, \ldots, y_{Nt})'$, $x_t = (x_{1t}, \ldots, x_{Nt})'$, and $z = (z_1, \ldots, z_N)'$, the model can be represented conveniently in vector form,

$$y_t = \rho y_{t-1} + x_t'\beta + z\gamma + \mu + v_t, \;\; t = \ldots, -1, 0, 1, \ldots, T. \tag{2.2}$$

In this paper, I focus on *short panels* where $N \to \infty$ but T is fixed and small.[2] The individual-specific effects μ can be either *random* (uncorrelated with the time-varying regressors), or *fixed* (may be correlated with some of the time-varying regressors), giving the so-called *random effects* (RE) or *fixed effects* (FE) DPD models. I will focus on the latter. I adopt a similar framework as Hsiao, Pesaran and Tahmiscioglu (2002), or HPT hereafter:

Assumption A: (i) *Data collection starts from the* 0*th period; the processes start from the* $-m$*th period, i.e.,* m *periods before the start of data collection,* $m = 0, 1, \ldots$*, and then evolve according to the model specified by (2.2);* (ii) *Starting positions of the process* y_{-m} *are treated as exogenous; hence the exogenous variables* (x_t, z) *and the errors* u_t *start to have impact on the response from period* $-m+1$ *onwards; and* (iii) *All exogenous quantities* (y_{-m}, x_t, z) *are considered as random and inferences proceed by conditioning on them.*

Thus, when $m = 0$, the *initial observations*, y_0, are exogenous, when $m \geq 1$, y_0 becomes endogenous, and when $m = \infty$, the processes have reached stationarity by the time the data collection starts. Note that in reality, the exact value of m is unknown.

Clearly, the Model (2.2) can be consistently estimated by the maximum likelihood (ML) or quasi-ML (QML) method when the vector of unobserved individual-specific effects μ is uncorrelated with the time-varying regressors x_t and is treated as a random vector of iid elements. When μ in model (2.2) is correlated with (some of) the time-varying regressors x_t in an unknown manner as in many economic applications, μ acts as if they are N free parameters, and with T fixed the model cannot be consistently estimated due to the incidental parameters problem. A standard practice is to transform the data to eliminate the fixed effects μ, such as first-differencing, demeaning, and orthogonal transformation.[3] In this paper, I follow HPT and eliminate the fixed effects μ by first-differencing (2.2),

$$\Delta y_t = \rho \Delta y_{t-1} + \Delta x_t \beta + \Delta v_t, \quad t = \ldots, -1, 0, 1, 2, \ldots, T. \tag{2.3}$$

The parameters left in Model (2.3) are $\psi = \{\rho, \beta', \sigma_v^2\}'$. Note that Model (2.3) is defined only for $t \geq 2$ as, e.g., Δy_1 depends on its lag $\Delta y_0 = y_0 - y_{-1}$ where y_{-1} is not observed.

[2]While the inference methodologies will be developed for short panels, they remain valid when T becomes large, and in fact they even work better than the traditional ones developed for large N and large T panels.

[3]HPT show that their ML estimators are invariant to the choice of the transformation matrix. This approach is attractive as it does not require any knowledge on μ, but the side effect is that all the time-invariant regressors are wiped out by such a transformation. Being unable to estimate the effects of time-invariant regressors may be one of the practical weakness of the fixed effects model although it has the strong attraction of allowing one to use the panel data to establish causation under much weaker assumptions (Cameron and Trivedi, 2005, Ch. 21).

To formulate the likelihood function for ML or QML estimation of Model (2.3) when T is fixed and small, it is essential to have the marginal distribution of Δy_1. This is so because, under the fundamental set-up in Assumption A, Δy_1 is always endogenous even when y_0 is exogenous and thus contains useful and non-negligible information about the structural parameters in the model. While the marginal distribution of Δy_1 can be obtained under Assumption A, it clearly depends on y_{-m} and $\{\Delta x_s, s = -m+2, \ldots, 0\}$, which are not observed and hence induces another incidental parameters problem. HPT propose a 'predictive' model for Δy_1 based on the observables $\{\Delta x_1', \ldots, \Delta x_T'\}'$, see Section 4 for details. In this paper, I proceed with a different route and propose simple estimation and inference methods that are free from the specifications of the initial conditions and are robust against misspecifications of the error distributions.

As usual, I use $|\cdot|$, $\|\cdot\|$ and $\mathrm{tr}(\cdot)$ to denote, respectively, the determinant, the Frobenius norm, and the trace of a matrix. I use $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ to denote the largest and smallest eigenvalues of a real symmetric matrix A, and $\otimes$ to denote the Kronecker product.

3 Initial-Condition Free Estimation of FE-DPD Models

3.1 The modified QML estimator

The proposed method starts from the conditional Gaussian likelihood function, given Δy_1 and $\Delta x_1, \ldots, \Delta x_T$. Let $\Delta Y = \{\Delta y_2', \Delta y_3', \ldots, \Delta y_T'\}'$, $\Delta Y_{-1} = \{\Delta y_1', \Delta y_2', \ldots, \Delta y_{T-1}'\}'$, $\Delta X = \{\Delta x_2', \Delta x_3', \ldots, \Delta x_T'\}'$, and $\Delta v = \{\Delta v_2', \Delta v_3', \ldots, \Delta v_T'\}'$. The model (2.3) is further compacted in matrix form,

$$\Delta Y = \rho \Delta Y_{-1} + \Delta X \beta + \Delta v. \tag{3.1}$$

Under normality of $\{v_t\}$, the joint distribution of Δv is easily seen as

$$\Delta v \sim N(0, \sigma_v^2 \Omega), \quad \text{where} \quad \Omega = \sigma_v^{-2} \mathrm{Var}(\Delta v) = (C \otimes I_N), \quad \text{and}$$

$$C = \begin{pmatrix} 2 & -1 & 0 & \ldots & 0 & 0 & 0 \\ -1 & 2 & -1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & 2 & -1 \\ 0 & 0 & 0 & \ldots & 0 & -1 & 2 \end{pmatrix}_{(T-1)\times(T-1)}.$$

The joint distribution of Δv translates directly to the conditional joint distribution of ΔY since the Jacobian of the transformation is one, giving the conditional quasi Gaussian log-likelihood of ψ in terms of $\Delta y_2, \ldots, \Delta y_T$, conditional on Δy_1 and $\Delta x_1, \ldots, \Delta x_T$:

$$\ell(\psi) = -\frac{n(T-1)}{2}\log(2\pi) - \frac{n(T-1)}{2}\log(\sigma_v^2) - \frac{1}{2}\log|\Omega| - \frac{1}{2\sigma_v^2}\Delta v(\theta)'\Omega^{-1}\Delta v(\theta), \tag{3.2}$$

where $\theta = (\rho, \beta')'$ and $v(\theta) = \Delta Y - \rho \Delta Y_{-1} - \Delta X \beta$. Clearly, maximizing (3.2) is equivalent to running an ordinary least square (OLS) of ΔY on $\Delta \mathbb{X} \equiv (\Delta Y_{-1}, \Delta X)$, resulting in the conditional QML estimators (QMLEs) of θ and σ_v^2 as,

$$\hat{\theta} = (\Delta \mathbb{X}' \Omega^{-1} \Delta \mathbb{X})^{-1} \Delta \mathbb{X}' \Omega^{-1} \Delta Y \quad \text{and} \quad \hat{\sigma}_v^2 = \tfrac{1}{N(T-1)} \Delta \hat{v}' \Omega^{-1} \Delta \hat{v}, \tag{3.3}$$

where $\Delta \hat{v} = \Delta Y - \Delta \mathbb{X}\, \hat{\theta}$, and the dependence of a quantity on N, e.g., $\hat{\theta}$, is suppressed. Note that $\hat{\theta}$ and $\hat{\sigma}_v^2$ can never be the exact ML estimators (MLEs) as the loglikelihood function (3.2) can never be the exact loglikelihood function even if the errors are truly normal, simply because Δy_1 can never be strictly exogenous. This is not a problem when T is large and in this case the effect of Δy_1 becomes negligible, although there is an issue on the order of finite sample bias as discussed latter.

We now investigate the consequence of treating Δy_1 as exogenous when T is fixed and small. Let $\psi_0 = \{\rho_0, \beta_0', \sigma_{v0}^2\}'$ be the true value of $\psi = \{\rho, \beta', \sigma_v^2\}'$. The conditional quasi score function $S(\psi) = \frac{\partial}{\partial \psi} \ell(\psi)$ has the elements corresponding to ρ, β, and σ^2:

$$S(\psi) = \begin{cases} \frac{1}{\sigma_v^2} \Delta Y_{-1}' \Omega^{-1} \Delta v(\theta), \\ \frac{1}{\sigma_v^2} \Delta X' \Omega^{-1} \Delta v(\theta), \\ \frac{1}{2\sigma_v^4} \Delta v(\theta)' \Omega^{-1} \Delta v(\theta) - \frac{N(T-1)}{2\sigma_v^2}. \end{cases} \tag{3.4}$$

It is well known that (i) for a regular QML-type estimator to be consistent, it is necessary that the averaged score at the true value ψ_0 of the parameter vector ψ converges to zero in probability as N tends to infinity, i.e., $\text{plim}_{N \to \infty} \frac{1}{N} S(\psi_0) = 0$; (ii) for it to be unbiased, it is necessary that $\text{E}[S(\psi_0)] = 0$; and (iii) for it to be asymptotically unbiased, it is necessary that $\text{E}[S(\psi_0)] = o(\sqrt{N})$. These are easily seen to be the cases for the last two elements of $S(\psi_0)$, but not for the first element as shown below. Let $\Delta v = \Delta v(\theta_0)$.

Lemma 3.1: *Assume Models* (2.2) *and* (2.3) *satisfy Assumptions A. We have,*

$$\sigma_{v0}^{-2} \text{E}(\Delta Y_{-1}' \Omega^{-1} \Delta v) = N \text{tr}[C^{-1} D(\rho_0)], \tag{3.5}$$

where $D(\rho)$ is a $(T-1) \times (T-1)$ matrix and has the following expression:

$$D(\rho) = \begin{pmatrix} -1 & 2-\rho & -(1-\rho)^2 & \dots & -\rho^{T-5}(1-\rho)^2 & -\rho^{T-4}(1-\rho)^2 \\ 0 & -1 & 2-\rho & \dots & -\rho^{T-6}(1-\rho)^2 & -\rho^{T-5}(1-\rho)^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 2-\rho \\ 0 & 0 & 0 & \dots & 0 & -1 \end{pmatrix}.$$

Note that the result (3.5) only requires the existence of the second moment of v_{it}. Clearly,

$\text{tr}[C^{-1}D(\rho_0)] \neq 0$ which is independent of N. Thus, $\text{plim}_{N\to\infty}\frac{1}{NT}\sigma_{v0}^{-2}\text{E}(\Delta Y_{-1}'\Omega^{-1}\Delta v) \neq 0$ for a fixed T, showing that the conditional QMLE $\hat{\rho}$ defined in (3.3) is inconsistent. The result in Lemma 3.1 immediately suggests that modifying the score element for ρ by subtracting $N\text{tr}[C^{-1}D(\rho_0)]$ from it would fix the inconsistency problem of $\hat{\rho}$. Now, with (3.5) and the fact that other score elements have zero expectation, the modified score function becomes

$$S^*(\psi) = \begin{cases} S_\rho^*(\psi) = \frac{1}{\sigma_v^2}\Delta Y_{-1}'\Omega^{-1}\Delta v(\theta) - N\text{tr}(C^{-1}D(\rho)), \\ S_\beta^*(\psi) = \frac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta v(\theta), \\ S_{\sigma_v^2}^*(\psi) = \frac{1}{2\sigma_v^4}\Delta v(\theta)'\Omega^{-1}\Delta v(\theta) - \frac{N(T-1)}{2\sigma_v^2}. \end{cases} \tag{3.6}$$

Very interestingly, the modifier $N\text{tr}(C^{-1}D(\rho))$ of the score function depends only on the parameter ρ.[4] Hence, the modified QML estimators based on solving $S^*(\psi) = 0$ is free from (i) the unknown processes starting time designated by m, (ii) the specification of the distribution of the initial differences Δy_1, and (iii) the time-varying regressors included in the model. These, in particular (ii), are very useful features and it stands in contrast to the modeling strategy followed by HPT. The modified QML estimator $\hat{\psi}^*$ of ψ is thus,

$$\hat{\psi}^* = \arg\left\{S^*(\psi) = 0\right\}. \tag{3.7}$$

This root-finding process can be simplified by first solving the equations for β and σ^2, given ρ, resulting in the constrained estimates of β and σ^2 as

$$\tilde{\beta}^*(\rho) = (\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}\Delta Y(\rho) \quad \text{and} \quad \tilde{\sigma}_v^{*2}(\rho) = \frac{1}{N(T-1)}\Delta\tilde{v}(\rho)'\Omega^{-1}\Delta\tilde{v}(\rho), \tag{3.8}$$

where $\Delta Y(\rho) = \Delta Y - \rho\Delta Y_{-1}$ and $\Delta\tilde{v}(\rho) = \Delta Y(\rho) - \Delta X\tilde{\beta}^*(\rho)$. Substituting $\tilde{\beta}^*(\rho)$ and $\tilde{\sigma}_v^{*2}(\rho)$ into the first component of $S^*(\psi)$ for β and σ^2, and solving the resulted concentrated estimating equation gives the modified QMLE of ρ as $\hat{\rho}^* = \arg\{S_c^*(\rho) = 0\}$, where

$$S_c^*(\rho) = \frac{\Delta Y_{-1}'\Omega^{-1}\Delta\tilde{v}(\rho)}{N(T-1)\tilde{\sigma}_v^{*2}(\rho)} - \frac{\text{tr}(C^{-1}D(\rho))}{T-1}. \tag{3.9}$$

The unconstrained modified QMLEs for β and σ^2 are thus $\hat{\beta}^* \equiv \tilde{\beta}^*(\hat{\rho}^*)$ and $\hat{\sigma}_v^{*2} \equiv \tilde{\sigma}_v^{*2}(\hat{\rho}^*)$.

Furthermore, $\text{tr}[C^{-1}D(\rho_0)] \neq 0$ even when T is not fixed as seen from the result below:

Lemma 3.2: *For* $|\rho| < 1$, *we have* $\text{tr}[C^{-1}D(\rho)] = -\frac{1}{1-\rho} + \frac{1-\rho^T}{T(1-\rho)^2}$.

The implication of this result is as follows. If T also tends to infinity, the conditional QMLE $\hat{\rho}$, ignoring the initial conditions, can indeed have the desired asymptotic behavior as being consistent and asymptotically normally distributed, but its finite sample behavior

[4] It can be intuitively explained as a factor that recovers the neglected information contained in the initial differences Δy_1.

can be poor as $\hat{\rho}$ encounters a finite sample bias of order $O(T^{-1})$ instead of the desired order $O((NT)^{-1})$. This bias can be significant if T is not large relative to N. Clearly, the proposed modification automatically eliminates such as a bias.

3.2 Asymptotic properties of the modified QMLE

To study formally the asymptotic properties (consistency and asymptotic normality) of the modified QMLE $\hat{\psi}^* = (\hat{\rho}^*, \hat{\beta}^{*\prime}, \hat{\sigma}_v^{*2})'$, additional regularity conditions and definitions are needed. In general, let $\mathcal{P}$ be the parameter space for ρ assumed to be compact. Consistency of $\hat{\rho}^*$ follows if $S_c^*(\rho)$ and its non-stochastic counter part, say $\bar{S}_c^*(\rho)$, are such that $S_c^*(\rho) - \bar{S}_c^*(\rho) = o_p(1)$ uniformly in $\rho \in \mathcal{P}$, $|\bar{S}_c^*(\rho)| > 0$ for $\rho \neq \rho_0$, and $\bar{S}_c^*(\rho_0) = 0$, see, e.g., van der Vaart (1998, ch. 5). This result is intuitive quite clear: as $\hat{\rho}^*$ is a zero of $S_c^*(\rho)$ and ρ_0 is a zero of $\bar{S}_c^*(\rho_0)$, the difference between $\hat{\rho}^*$ and ρ_0 would vanish if the difference between the objective functions vanishes as N tends to infinity. Obviously, $\bar{S}_c^*(\rho_0)$ can be defined as follows. Let $\{\bar{\beta}^*(\rho), \bar{\sigma}_v^{*2}(\rho)\} = \arg\{\mathrm{E}[S_\beta^*(\psi)] = 0,\ \mathrm{E}[S_{\sigma_v^2}^*(\psi)] = 0\}$. Then, $\bar{S}_c^*(\rho_0) = \frac{1}{N(T-1)}\mathrm{E}[S_\rho^*(\psi)]|_{\beta=\bar{\beta}^*(\rho), \sigma_v^2=\bar{\sigma}_v^{*2}(\rho)}$, i.e.,

$$\bar{S}_c^*(\rho_0) = \frac{\mathrm{E}[\Delta Y_{-1}\Omega^{-1}\Delta v(\rho, \bar{\beta}^*(\rho))]}{N(T-1)\bar{\sigma}_v^{*2}(\rho)} - \frac{\mathrm{tr}(C^{-1}D(\rho))}{T-1}. \tag{3.10}$$

See Appendix (Proof of Theorem 3.1) for details. Once the consistency of $\hat{\rho}^*$ is established, the consistency of $\hat{\beta}^*$ and $\hat{\sigma}_v^{*2}$ follows almost immediately.

Although the exact distribution of the initial differences Δy_{-1} is not required in the modified QML estimation, Δy_{-1} needs to meet some minimum requirements in order to establish consistency and asymptotic normality of $\hat{\psi}^*$. Under the general specifications given at the end of Section 2 (Assumption A and Model (2.3)), by continuous substitutions from the tst period back to the $-m$th period, we obtain

$$\Delta y_t = \rho_0^{m+t-1}\Delta y_{-m+1} + \Delta x_t(m, \rho_0)\beta_0 + \Delta v_t(m, \rho_0), \tag{3.11}$$

where $\Delta x_t(m, \rho) = \sum_{s=-m+2}^{t} \rho^{t-s}\Delta x_s$ and $\Delta v_t(m, \rho) = \sum_{s=-m+2}^{t} \rho^{t-s}\Delta v_s$. We have,

$$\begin{aligned} \Delta y_1 &= \rho_0^m \phi(y_{-m}) + \Delta x_1(m, \rho_0)\beta_0 + \rho_0^m[\Delta y_{-m+1} - \phi(y_{-m})] + \Delta v_1(m, \rho_0) \\ &= \Delta\eta_1 + \rho_0^m v_{-m+1} + \Delta v_1(m, \rho_0), \end{aligned} \tag{3.12}$$

where $\phi(y_{-m}) = \mathrm{E}(\Delta y_{-m+1}|y_{-m})$ and $\Delta\eta_1 = \rho_0^m\phi(y_{-m}) + \Delta x_1(m, \rho_0)\beta_0$.

Assumption B: *The errors* $\{v_{it}\}, i = 1, \ldots, N,\ t = -m \ldots, -1, 0, 1, \ldots, T$, *are iid with mean zero, variance* σ_v^2, *and finite moments* $\mathrm{E}(v_{it}^{4+\delta})$ *for some* $\delta > 0$.

Assumption C: *The parameter space* $\mathcal{P}$ *is compact, and* ρ_0 *is in its interior.*

Assumption D: *(i) The time-varying regressors are exogenous or weakly exogenous,* *and their values* $x_{it}, i = 1, \ldots, N,\ t = -m \ldots, -1, 0, 1, \ldots, T$, *are uniformly bounded, (ii)* $\lim_{N\to\infty} \frac{1}{N}\Delta\eta_1'\Delta\eta_1$ *exists, and (iii)* $\lim_{N\to\infty} \frac{1}{NT}\Delta X'\Delta X$ *exists and is nonsingular.*

Assumption B is standard for the QML-type estimation. Assumption D(ii) imposes a general condition on the mean of the initial differences. It implies that when m is finite it is not necessary that $|\rho| < 1$, but the conditional expectations of the initial endowments Δy_{-m+1} given the process starting positions y_{-m}, $\phi(y_{-m})$, must have finite second moments; when $m = \infty$ it is necessary that $|\rho| < 1$ and in this case the effect of $\phi(y_{-m})$ vanishes. Assumption D(iii) ensures the identification of β and σ_v^2 once ρ is identified. Let $g(\rho) = -\frac{1}{T-1}\mathrm{tr}[C^{-1}D(\rho)]$. The identification uniqueness condition for ρ_0 is given as follows.

Assumption E: $\frac{1}{NT}(\rho_0 - \rho)\mathrm{E}(\Delta Y_{-1}'\Omega^{-1}\Delta Y_{-1}) + \bar{\sigma}_v^{*2}(\rho)g(\rho) - \sigma_{v0}^2 g(\rho_0) \neq 0,\ \forall \rho \neq \rho_0$.

Let $H^*(\psi) = \frac{\partial}{\partial\psi'}S^*(\psi)$ be the modified Hessian matrix, $\Sigma^*(\psi_0) = -\frac{1}{N(T-1)}\mathrm{E}\big[H^*(\psi_0)\big]$, and $\Gamma^*(\psi_0) = \mathrm{Var}[\frac{1}{\sqrt{N(T-1)}}S^*(\psi_0)]$. The theorem below summarizes the asymptotic properties of the modified QMLEs defined in (3.7)-(3.9). Its proof is given in Appendix.

Theorem 3.1: *Assume Model* (2.3) *satisfies Assumptions A-E. Then, the modified QMLE* $\hat{\psi}^*$ *is a consistent estimator for* ψ_0, *i.e.,* $\hat{\psi}^* \xrightarrow{p} \psi_0$ *as* $N \to \infty$, *and*

$$\sqrt{N(T-1)}(\hat{\psi}^* - \psi_0) \xrightarrow{D} N\big[0,\ \lim_{N\to\infty} \Sigma^{*-1}(\psi_0)\Gamma^*(\psi_0)\Sigma^{*-1}(\psi_0)\big],$$

where $\Gamma^*(\psi_0), \Sigma^*(\psi_0)$ *and its inverse are assumed to exist for large enough* N.

Theorem 1 shows that $\hat{\psi}^*$ is $\sqrt{N}$-consistent for a fixed T. It is easy to see that the result of Theorem 1 remains valid when T tends to infinity as well and in this case $\hat{\psi}^*$ is $\sqrt{N(T-1)}$-consistent. It is reasonable to believe that the proposed estimators should perform better in terms of bias than the QML estimators based on large N and large T setting as the score function for ρ is bias corrected. This is confirmed by our Monte Carlo results.

3.3 Initial-condition free estimation of the VC matrix

To facilitate the practical applications of the proposed estimation method, a method of estimating the variance of $\hat{\psi}^*$ is desired. First, $\Sigma^*(\psi_0)$ can be consistently estimated by $-\frac{1}{N(T-1)}H^*(\hat{\psi}^*)$, where $H^*(\psi)$ has the typical components:

$$\begin{aligned}
H_{\rho\rho}^*(\psi) &= -\tfrac{1}{\sigma_v^2}\Delta Y_{-1}'\Omega^{-1}\Delta Y_{-1} - N\mathrm{tr}[C^{-1}\dot{D}(\rho)],\\
H_{\rho\beta}^*(\psi) &= -\tfrac{1}{\sigma_v^2}\Delta Y_{-1}'\Omega^{-1}\Delta X,\\
H_{\rho\sigma_v^2}^*(\psi) &= -\tfrac{1}{\sigma_v^4}\Delta Y_{-1}'\Omega^{-1}\Delta v(\theta),\\
H_{\beta\beta}^*(\psi) &= -\tfrac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta X,\\
H_{\beta\sigma_v^2}^*(\psi) &= -\tfrac{1}{\sigma^4}\Delta X'\Omega^{-1}\Delta v(\theta),\\
H_{\sigma_v^2\sigma_v^2}^*(\psi) &= -\tfrac{1}{\sigma_v^6}\Delta v(\theta)'\Omega^{-1}\Delta v(\theta) + \tfrac{N(T-1)}{2\sigma_v^4},
\end{aligned}$$

where $\dot{D}(\rho) = \frac{\partial}{\partial \rho} D(\rho)$. Note that the modified Hessian matrix $H^*(\psi_0)$ depends only on the observables and the model parameter vector ψ_0 of which consistent estimate is available. The estimate $-\frac{1}{N(T-1)} H^*(\hat{\psi}^*)$ of $\Sigma^*(\psi_0)$ is thus free from the specification of the initial conditions. Its consistency is given in Theorem 2 below and proved in Appendix.

The estimation of $\Gamma^*(\psi_0)$ is trickier as the nature of a dynamic panel data model renders the closed-form expression of $\text{Var}[S^*(\psi_0)]$, which can be derived using (3.4), infeasible as it contains the unobservables $(y_{-m}, x_s, s = -m, \ldots, -1)$ and hence the traditional plug-in method cannot be applied, unless the initial observations are modeled (approximated using the observables) as in HPT. Note that the likelihood function used is quasi in two senses: the exact error distribution is unknown but treated as Gaussian and the initial differences are endogenous but treated as exogenous. Thus, $\Gamma^*(\psi_0) \neq \Sigma^*(\psi_0)$ even when the errors are exactly normal, and a consistent estimator of $\Gamma^*(\psi_0)$ is always needed for inference based on the modified estimation. As our asymptotics are based on large N and fixed T and the modified score function can be decomposed into a sum of N independent terms, an outer product of gradients (OPG) estimator of $\Gamma^*(\psi_0)$ naturally emerges.

Let $\Delta y_{i-} = \{\Delta y_{i1}, \Delta y_{i2}, , \ldots, \Delta y_{i,T-1}\}'$, $\Delta v_{i-} = \{\Delta v_{i2}, \Delta v_{i3}, , \ldots, \Delta v_{i,T}\}'$, and $\Delta x_{i-} = \{\Delta x_{i2}, \Delta x_{i3}, , \ldots, \Delta x_{i,T}\}'$. The key elements of the modified score can be written as

$$\Delta Y_{-1}' \Omega^{-1} \Delta v = \textstyle\sum_{i=1}^N \Delta y_{i-}' C^{-1} \Delta v_{i-} \equiv \sum_{i=1}^N g_{\rho i}(\theta_0), \tag{3.13}$$

$$\Delta X' \Omega^{-1} \Delta v = \textstyle\sum_{i=1}^N \Delta x_{i-}' C^{-1} \Delta v_{i-} \equiv \sum_{i=1}^N g_{\beta i}(\theta_0), \tag{3.14}$$

$$\Delta v' \Omega^{-1} \Delta v = \textstyle\sum_{i=1}^N \Delta v_{i-}' C^{-1} \Delta v_{i-} \equiv \sum_{i=1}^N g_{\sigma_v^2 i}(\theta_0), \tag{3.15}$$

which are all sums of N independent items. Define the $N \times (p+2)$ gradient matrix:

$$G^*(\psi_0) = \left\{ \tfrac{1}{\sigma_{v0}^2} g_{\rho i}(\psi_0) - \text{tr}(C^{-1} D(\rho_0)),\ \tfrac{1}{\sigma_{v0}^2} g_{\beta i}(\psi_0),\ \tfrac{1}{2\sigma_{v0}^4} g_{\sigma_v^2 i}(\psi_0) - \tfrac{T-1}{2\sigma_{v0}^2} \right\}. \tag{3.16}$$

Thus, we have $S^*(\psi_0) = G^*(\psi_0)' 1_N$ where 1_N is an N-vector of ones. An OPG estimator of $\Gamma^*(\psi_0)$ is $\frac{1}{NT} G^*(\hat{\psi}^*)' G^*(\hat{\psi}^*)$. Finally, the sandwich estimator of the VC matrix of $\hat{\psi}^*$ takes the form: $H^*(\hat{\psi}^*)^{-1} G^*(\hat{\psi}^*)' G^*(\hat{\psi}^*) H^*(\hat{\psi}^*)^{-1}$. It is seen that the gradients $G^*(\psi_0)$ contains only the observables and the model parameters ψ_0, its outer product evaluated at $\hat{\psi}^*$ leads to a initial-condition free estimate of $\text{Var}[S^*(\psi_0)]$. This together with $H^*(\hat{\psi}^*)$ lead to a initial-condition free estimate of $\text{Var}(\hat{\psi}^*)$, of which consistency is given below.

Theorem 3.2: *Assume Model* (2.3) *satisfies Assumptions A-E. Then, as* $N \to \infty$,

$$\frac{1}{N(T-1)} G^*(\hat{\psi}^*)' G^*(\hat{\psi}^*) - \Gamma^*(\psi_0) \xrightarrow{p} 0,$$

and hence $N(T-1) H^*(\hat{\psi}^*)^{-1} G^*(\hat{\psi}^*)' G^*(\hat{\psi}^*) H^*(\hat{\psi}^*)^{-1} - \Sigma^*(\psi_0)^{-1} \Gamma^*(\psi_0) \Sigma^*(\psi_0)^{-1} \xrightarrow{p} 0.$

Obviously, the OPG approach for robust VC matrix estimation depends on the large N

and fixed T setting. When T also tends to infinity, alternative method of robust VC matrix estimation is desired. This is in contrast the parameter estimation problem.

4 QML Estimation Utilizing Initial Observations

The inference methods proposed in the early section capture the compound effects of the processes from past to the initial observations through a simple modification on the score function. The proposed methods are free from the specifications of the distributions of the initial differences, and thus are very easy to apply. However, the initial differences contain useful information about the structural parameters of the model, and hence it is natural to wonder if there would be any efficiency loss if the initial observations are not fully utilized. To address this issue, I in this section first provide some theoretical discussions alone the lines of HPT, and then in next section present some Monte Carlo results. I extend HPT's discussions based on ML estimation to QML estimation framework, establish formally the consistency and asymptotic normality of the full QML estimators, and introduce a simple method for robust VC matrix estimation.

Rewrite the expression for Δy_1 given in (3.12) as,

$$\Delta y_1 = \Delta\eta_0 + \Delta x_1\beta_0 + \rho_0^m v_{-m+1} + \Delta v_1(m, \rho_0), \tag{4.1}$$

where $\Delta\eta_0 = \rho_0^m \phi(y_{-m}) + \Delta x_0(m, \rho_0)\beta_0$. Note that $\Delta\eta_0$ together with $\Delta x_1\beta_0$ makes up the exogenous part of Δy_1. Depending the value of m, $\Delta\eta_0$ can be very complicated, containing the process starting positions y_{-m} and the past values of the time-varying regressors $\{x_{-m+1}, \ldots, x_{-1}\}$ of which all are not observed. It also contains the parameters β_0 and ρ_0. Thus, $\Delta\eta_0$ acts like an $N\times 1$ vector of unknown parameters, and hence cannot be consistently estimated due to the incidental parameters problem.

If $\Delta\eta_0$ can be approximated linearly by the observed values of the regressors, i.e.,

$$\Delta\eta_0 = \alpha_0 1_n + \alpha_1\Delta x_1 + \ldots, \alpha_T\Delta x_T + e \equiv \Delta\tilde{\mathbf{x}}\ \alpha + e,$$

with e being the vector of approximation errors, iid with mean zero and variance σ_e^2 and independent of v_t, then, we have a 'predictive' model for Δy_1,

$$\Delta y_1 = \Delta\tilde{\mathbf{x}}\ \alpha + \Delta x_1\beta + \Delta v_1^{\diamond}, \tag{4.2}$$

where $\Delta v_1^{\diamond} = e + \rho^m v_{-m+1} + \Delta v_1(m, \rho)$. HPT argued that this approximation is valid if the regressors are generated from either of the following two processes:

$$\begin{aligned} x_{it} &= \mu_i + gt + \textstyle\sum_{j=0}^{\infty} a_j\varepsilon_{i,t-j}, \quad \sum_{j=0}^{\infty}|a_j| < \infty, && (4.3)\\ \Delta x_{it} &= g + \textstyle\sum_{j=0}^{\infty} d_j\varepsilon_{i,t-j}, \quad \sum_{j=0}^{\infty}|d_j| < \infty, && (4.4)\end{aligned}$$

where μ_i can either be fixed or random, and ε_{it} are iid$(0, \sigma_\varepsilon^2)$. It can be shown that under (4.3) and (4.4), the Assumptions D(ii)-(iii) hold.

Letting $\Delta v^\diamond = (\Delta v_1^{\diamond\prime}, \Delta v_2', \ldots, \Delta v_T')'$. It is easy to show that by construction and under the strict exogeneity of $(\Delta x_1, \ldots, \Delta x_T)$, $\mathrm{E}(\Delta v_1^\diamond) = 0$,

$$\mathrm{E}(\Delta v_1^\diamond \Delta v_1^{\diamond\prime}) = \sigma_e^2 I_n + \sigma_v^2 c_m(\rho) I_n = \sigma_v^2(\phi_e + c_m(\rho)) I_n, \text{ and} \tag{4.5}$$

$$\mathrm{E}(\Delta v_1^\diamond \Delta v_t') = -\sigma_v^2 I_N, \text{ for } t = 2, \text{ and } 0, \text{ for } t = 3, \ldots, T, \tag{4.6}$$

where $\phi_e = \sigma_e^2/\sigma_v^2$ and $c_m(\rho) = \frac{2}{1+\rho} - \frac{\rho^{2m}(1-\rho)}{1+\rho}$. Under the normality assumptions of $\{v_t\}$ and e, we have $\Delta v^\diamond \sim N(0, \sigma_{v0}^2 \Omega_0^\diamond)$, where $\Omega_0^\diamond \equiv \Omega^\diamond(\omega_0) = C^\diamond(\omega_0) \otimes I_N$, and

$$C^\diamond(\omega) = \begin{pmatrix} \omega & -1 & 0 & \ldots & 0 & 0 & 0 \\ -1 & 2 & -1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & 2 & -1 \\ 0 & 0 & 0 & \ldots & 0 & -1 & 2 \end{pmatrix}_{T \times T},$$

where $\omega = \phi_e + c_m(\rho)$, treated as a free parameter, as in HPT, to be estimated jointly with the other model parameters.[5] The joint distribution of $\Delta v^\diamond$ is directly converted to the joint distribution of $\Delta Y^\diamond = (\Delta y_1', \Delta_2', \ldots, \Delta y_T')'$ or the likelihood function, as the Jacobian of the transformation is unity. Under (3.1) and (4.2)-(4.4), the full quasi-Gaussian loglikelihood of $\psi = (\theta', \sigma_v^2, \omega)'$, $\theta = (\rho, \beta', \alpha')'$, incorporating the initial observations is written as

$$\ell^\diamond(\psi) = -\frac{NT}{2}\log(2\pi) - \frac{NT}{2}\log(\sigma_v^2) - \frac{1}{2}\log|\Omega^\diamond(\omega)| - \frac{1}{2\sigma_v^2}\Delta v^\diamond(\theta)'\Omega^\diamond(\omega)^{-1}\Delta v^\diamond(\theta), \tag{4.7}$$

where $\Delta v^\diamond(\theta) = \Delta Y^\diamond - \Delta X^\diamond\, \theta$, and, letting r be the column rank of $\Delta\tilde{\mathbf{x}}$,

$$\Delta X^\diamond = \begin{pmatrix} 0_{N\times 1} & \Delta x_1 & \Delta\tilde{\mathbf{x}} \\ \Delta y_1 & \Delta x_2 & 0_{N\times r} \\ \vdots & \vdots & \vdots \\ \Delta y_{T-1} & \Delta x_T & 0_{N\times r} \end{pmatrix},$$

Given ω, the constrained QML estimators of θ and σ_v^2 are

$$\tilde{\theta}^\diamond(\omega) = (\Delta X^{\diamond\prime}\Omega^\diamond(\omega)^{-1}\Delta X^\diamond)^{-1}\Delta X^{\diamond\prime}\Omega^\diamond(\omega)^{-1}\Delta Y^\diamond \text{ and} \tag{4.8}$$

$$\tilde{\sigma}_v^{\diamond 2}(\omega) = \tfrac{1}{NT}\Delta\tilde{v}^\diamond(\omega)'\Omega^\diamond(\omega)^{-1}\Delta\tilde{v}^\diamond(\omega), \tag{4.9}$$

[5]Note that $c_m(\rho)$ takes a slightly different form from HPT. However, the exact expression is not important as $c_m(\rho)$ and ϕ_e are integrated to form a single parameter ω. The advantage of doing so is that it avoids the issue that m is also unknown. This treatment is valid as long as the model contains at least one time-varying regressor, due to the presence of ϕ_e. When the model does not contain any time-varying regressor, $\omega = c_m(\rho)$, which cannot be treated as a free parameter when $m = \infty$ or $\rho = 0$, as $c_\infty(\rho) = \frac{1}{1+\rho}$ and $c_m(0) = 2$.

where $\Delta\tilde{v}^\diamond(\omega) = \Delta Y^\diamond - \Delta X^\diamond \tilde{\theta}^\diamond(\omega)$. Substituting $\tilde{\theta}^\diamond(\omega)$ and $\tilde{\sigma}_v^{\diamond 2}(\omega)$ back into (4.7) gives the concentrated loglikelihood for ω:

$$\ell_c^\diamond(\omega) = -\frac{NT}{2}[\log(2\pi) + 1] - \frac{1}{2}\log|\Omega^\diamond(\omega)| - \frac{NT}{2}\log\hat{\sigma}_v^{\diamond 2}(\omega). \tag{4.10}$$

Maximizing (4.10) gives the full QMLE $\hat{\omega}^\diamond$ of ω. The unconstrained full QMLEs of θ and σ_v^2 are thus $\hat{\theta}^\diamond \equiv \tilde{\theta}(\hat{\omega}^\diamond)$ and $\hat{\sigma}_v^{\diamond 2} \equiv \tilde{\sigma}_v^{\diamond 2}(\hat{\omega}^\diamond)$. The QMLE of ψ is denoted as $\hat{\psi}^\diamond$.

The expressions given in (4.2), (4.5) and (4.6) show that the initial endowments Δy_1 can indeed be useful in estimating ρ and σ_v^2, and ignoring them as in the original conditional QML estimation would result in inconsistency in model estimation. The modified QML estimation captures the compound effect of the processes on the initial observations through a simple modification on the score function, resulting in consistent estimation of the model parameters. In contrast, the full QML estimation is through a model for the initial observations as in HPT. The tradeoff, however, in utilizing Δy_1 through (4.2) for model estimation is that an additional $pT + 2$ parameters have to be estimated. It is clear that this is beneficial only if N is significantly larger than $pT + 2$. However, doing so the information about ρ contained in Δy_1 is not utilized in estimating ρ, and hence the efficiency in estimating ρ cannot be increased by incorporating the initial observations into the model estimation.

Based on this set up, it is clear that the identification of the model parameters depends crucially on the identification of the nuisance parameter ω and the full rank condition of $\Delta X^\diamond$. Let $[\bar{\theta}(\omega), \bar{\sigma}_v^2(\omega)]$ be the solution of $\max_{\theta,\sigma_v^2}[\ell^\diamond(\psi)]$. We need the following conditions.

Assumption C′: *ω_0 is in the interior of a compact parameter space $\mathcal{W}$.*

Assumption D′: (*i*) *The regressors are generated by either Model* (4.3) *or Model* (4.4)*;* (*ii*) *All the elements in Δx_{it} have finite high-than 4th moments; and* (*iii*) $\frac{1}{NT}\Delta X^{\diamond\prime}\Delta X^\diamond$ *is positive definite almost surely for large enough N.*

Assumption E′: $\lim_{N\to\infty}\frac{1}{NT}[\log|\sigma_{v0}^2\Omega_0^\diamond| - \log|\bar{\sigma}_v^2(\omega)\Omega^\diamond(\omega)|]$*, for any $\omega \neq \omega_0$.*

Let $S^\diamond(\psi)$ and $H^\diamond(\psi)$ be, respectively, the score vector and the Hessian matrix of the full QML estimation. The asymptotic properties of the QMLE of ψ_0 are give as follows.

Theorem 4.1: *Assume Model* (2.3) *satisfies Assumptions A, B, and C′-E′. Then, the full QML estimator $\hat{\psi}^\diamond$ is consistent, i.e.,* $\text{plim}_{N\to\infty}(\hat{\psi}^\diamond - \psi_0) = 0$*, and is asymptotically normally distributed, i.e.,*

$$\sqrt{NT}(\hat{\psi}^\diamond - \psi_0) \xrightarrow{D} N\Big[0,\ \lim_{N\to\infty}\Sigma^\diamond(\psi_0)^{-1}\Gamma^\diamond(\psi_0)\Sigma^\diamond(\psi_0)^{-1}\Big],$$

where $\Gamma^\diamond(\psi_0) = \text{Var}[\frac{1}{\sqrt{NT}}S^\diamond(\psi_0)]$ *assumed to exist, and* $\Sigma^\diamond(\psi_0) = \frac{1}{NT}\text{E}[H^\diamond(\psi_0)]$ *assumed to exist and to be nonsingular, for large enough N.*

For robust inferences, a method for estimating the VC matrix of $\hat{\psi}^\diamond$ is desired. As in Section 3, $\Sigma^\diamond(\psi_0)$ can be consistently estimated by $-\frac{1}{NT}H^\diamond(\hat{\psi}^\diamond)$ and $\Gamma^\diamond(\psi_0)$ by an OPG method. First, the score function of the full loglikelihood has the form

$$S^\diamond(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X^{\diamond\prime}\Omega^{\diamond-1}\Delta v^\diamond(\theta), \\ \frac{1}{2\sigma_v^4}\Delta v^\diamond(\theta)'\Omega^{\diamond-1}\Delta v^\diamond(\theta) - \frac{NT}{2\sigma_v^2}, \\ \frac{1}{2\sigma_v^2}\Delta v^\diamond(\theta)'\Omega^{\diamond-1}\dot{\Omega}\Omega^{\diamond-1}\Delta v^\diamond(\theta) - \frac{1}{2}\text{tr}(\Omega^{\diamond-1}\dot{\Omega}), \end{cases}$$

where $\dot{\Omega} = \dot{C}\otimes I_N$ and $\dot{C}$ is a $T\times T$ matrix of element 1 on its top-left corner and 0 elsewhere. Similar to Section 3, let $\Delta y_{i-} = \{0, \Delta y_{i1}, \ldots, \Delta y_{i,T-1}\}'$, $\Delta v_{i-}^\diamond = \{\Delta v_{i1}^\diamond, \Delta v_{i2}, \ldots, \Delta v_{i,T}\}'$, and $\Delta x_{i-}^\diamond = \{\Delta x_{i1}^\diamond, \Delta x_{i2}^\diamond, , \ldots, \Delta x_{i,T}^\diamond\}'$, where $\Delta x_{i1}^{\diamond\prime}$ is the ith row of $\Delta X^\diamond$, $\Delta x_{i2}^{\diamond\prime}$ is the $(N+i)$th row, etc. An OPG estimator of $\Gamma^\diamond(\psi_0)$ is thus $\frac{1}{NT}G^\diamond(\hat{\psi}^\diamond)'G^\diamond(\hat{\psi}^\diamond)$, where

$$G^\diamond(\psi_0) = \big\{\tfrac{1}{\sigma_0^2}g_{\theta i}^\diamond(\psi_0),\ \tfrac{1}{2\sigma_0^4}g_{\sigma_v^2 i}^\diamond(\psi_0) - \tfrac{T}{2\sigma_0^2},\ \tfrac{1}{2\sigma_0^2}g_{\omega i}^\diamond(\psi_0) - \tfrac{1}{2}\text{tr}(C^{\diamond-1}\dot{C})\big\}_{N\times(p+2)},$$

and $\{g_{\theta i}^\diamond(\psi_0), g_{\sigma_v^2 i}^\diamond(\psi_0), g_{\omega i}^\diamond(\psi_0)\}$ are independent and are such that

$$\Delta X^{\diamond\prime}\Omega^{\diamond-1}\Delta v^\diamond = \textstyle\sum_{i=1}^N \Delta x_{i-}^{\diamond\prime}C^{\diamond-1}\Delta v_{i-}^\diamond \equiv \sum_{i=1}^N g_{\theta i}^\diamond(\psi_0), \quad (4.11)$$

$$\Delta v^{\diamond\prime}\Omega^{\diamond-1}\Delta v^\diamond = \textstyle\sum_{i=1}^N \Delta v_{i-}^{\diamond\prime}C^{\diamond-1}\Delta v_{i-}^\diamond \equiv \sum_{i=1}^N g_{\sigma_v^2 i}^\diamond(\psi_0), \quad (4.12)$$

$$\Delta v^{\diamond\prime}\Omega^{\diamond-1}\dot{\Omega}\Omega^{\diamond-1}\Delta v^\diamond = \textstyle\sum_{i=1}^N \Delta v_{i-}^{\diamond\prime}C^{\diamond-1}\dot{C}C^{\diamond-1}\Delta v_{i-}^\diamond \equiv \sum_{i=1}^N g_{\omega i}^\diamond(\psi_0). \quad (4.13)$$

The Hessian matrix $H^\diamond(\psi)$ has the typical elements:

$$\begin{aligned}
H_{\theta\theta'}^\diamond(\psi) &= -\tfrac{1}{\sigma_v^2}\Delta X^{\diamond\prime}\Omega^{\diamond-1}\Delta X^\diamond, \\
H_{\theta\sigma_v^2}^\diamond(\psi) &= -\tfrac{1}{\sigma_v^4}\Delta X^{\diamond\prime}\Omega^{\diamond-1}\Delta v^\diamond(\theta), \\
H_{\theta\omega}^\diamond(\psi) &= -\tfrac{1}{\sigma_v^2}\Delta X^{\diamond\prime}\Omega^{\diamond-1}\dot{\Omega}\Omega^{\diamond-1}\Delta v^\diamond(\theta), \\
H_{\sigma_v^2\sigma_v^2}^\diamond(\psi) &= -\tfrac{1}{\sigma_v^6}\Delta v^\diamond(\theta)'\Omega^{\diamond-1}\Delta v^\diamond(\theta) + \tfrac{NT}{2\sigma_v^4}, \\
H_{\sigma_v^2\omega}^\diamond(\psi) &= -\tfrac{1}{2\sigma_v^4}\Delta v^\diamond(\theta)'\Omega^{\diamond-1}\dot{\Omega}\Omega^{\diamond-1}\Delta v^\diamond(\theta), \\
H_{\omega\omega}^\diamond(\psi) &= -\tfrac{1}{\sigma_v^2}\Delta v^\diamond(\theta)'\Omega^{\diamond-1}\dot{\Omega}\Omega^{\diamond-1}\dot{\Omega}\Omega^{\diamond-1}\Delta v^\diamond(\theta) + \tfrac{1}{2}\text{tr}(\Omega^{\diamond-1}\dot{\Omega}\Omega^{\diamond-1}\dot{\Omega}).
\end{aligned}$$

Plugging $\hat{\psi}^\diamond$ for ψ in $H^\diamond(\psi)$ leads to a consistent estimator of $\Sigma^\diamond(\psi_0)$. Under (4.2), one can readily verify that $E[S^\diamond(\psi_0)] = 0$. Consistency of $\frac{1}{NT}H^\diamond(\hat{\psi}^\diamond)$ for $\Sigma^\diamond(\psi_0)$ and the consistency of $\frac{1}{NT}G^{\diamond\prime}(\hat{\psi}^\diamond)G^\diamond(\hat{\psi}^\diamond)$ for $\Gamma^\diamond(\psi_0)$ can be proved in the same manner as in Section 3.

Theorem 4.2: *Assume Assumptions A, C, and D′-E′ hold. Then, as $N\to\infty$,*

$$\tfrac{1}{NT}G^\diamond(\hat{\psi}^\diamond)'G^\diamond(\hat{\psi}^\diamond) - \Gamma^\diamond(\psi_0) \xrightarrow{p} 0,$$

and hence $NT\ H^\diamond(\hat{\psi}^\diamond)^{-1}G^\diamond(\hat{\psi}^\diamond)'G^\diamond(\hat{\psi}^\diamond)H^\diamond(\hat{\psi}^\diamond)^{-1} - \Sigma^\diamond(\psi_0)^{-1}\Gamma^\diamond(\psi_0)\Sigma^\diamond(\psi_0)^{-1} \xrightarrow{p} 0.$

Similar to the result of Theorem 3.2, the validity of the result of Theorem 4.2 depends on the large N and fixed T set up. When T is also large, an alternative method for robust VC matrix estimation is desired. However, in this case the results of Theorem 4.1 remain valid as the effect of Model (4.2) for the initial differences Δy_1 becomes small.

5 Monte Carlo Results

Monte Carlo experiments are carried out to (*i*) investigate finite sample performance of the proposed estimator, the modified QMLE (MQMLE), and the OPG estimator of the standard error of the MQMLE and (*ii*) compare with the closely related estimators, the QMLE based on the full likelihood (FQMLE) related to Hsiao et al. (2002), and the QMLE based on the conditional likelihood (CQMLE), under different initial conditions, and different error distributions. A particularly interesting question to be answered is how much more efficient can FQMLE be than MQMLE, as the former takes use of the initial observations. Also interested is how bad can CQMLE be when T is small and how does it improve as T grows. Is there still a comparative advantage of MQMLE over CQMLE when T grows with N? The following data generating process (dgp) is used in the Monte Carlo experiments:

$$y_t = \rho y_{t-1} + \beta_0 \iota_n + x_t \beta_1 + z\gamma + \mu + v_t,$$

where y_t, y_{t-1}, x_t, and z are all $N \times 1$ vectors. The elements of x_t are generated in a similar fashion as in Hsiao et al. (2002),[6] and the elements of z are randomly generated from $Bernoulli(0.5)$. The error (v_t) distributions can be (*i*) normal, (*ii*) normal mixture (10% $N(0,4)$ and 90% $N(0,1)$), or (*iii*) chi-square. The fixed effects μ are generated according to $\frac{1}{T}\sum_{t=1}^{T} x_t + e$ where e is a vector of iid random numbers. A simpler model, the pure FE-DPD model without regressors, is given a separate consideration.

The parameters' values are chosen to be $\beta_0 = 5, \beta_1 = \gamma = \sigma_\mu = \sigma_v = 1$, $\rho \in [0.9, 0.9]$, $T \in \{3, 6, \ldots\}$, $N = \{10, 20, 50, 100, 200\}$, and $\{m = 0, 5, 50\}$. Each set of Monte Carlo results is based on 1000 samples. Due to the space constraints, only a subset of results, corresponding to $m = 50$, are reported. The reported results are the Monte Carlo mean and standard deviation (sd), the averaged standard error ($\widehat{\text{se}}$) calculated based on Hessian only, and the averaged robust standard error ($\widehat{\text{rse}}$) calculated based on the sandwich estimator.

The case of small T. Tables 1-6 summarize the results for $T = (3, 6)$, $N = (50, 100, 200)$, and $\rho = (.8, .4, 0, -.4, -.8)$. The results reveal the followings: (*i*) when T is much smaller than N, MQMLE and FQMLE give almost identical results, (*ii*) the proposed the robust variance estimator works well for both MQMLE and FQMLE, (*iii*) CQMLE can be very biased and the bias does not reduce as N goes large, (*iv*) the sds of MQMLE and FQMLE can be slightly larger than that of CQMLE when ρ is positive and large, but otherwise are about

[6] The detail is: $x_t = \mu_x + gt1_n + \zeta_t$, $(1-\phi_1 L)\zeta_t = v_t + \phi_2 v_{t-1}$, $v_t \sim N(0, \sigma_1^2 I_N)$, $\mu_x = e + \frac{1}{T+m+1}\sum_{t=-m}^{T} v_t$, and $e \sim N(0, \sigma_2^2)$. The reported results correspond to $(g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (0.01, 0.5, 0.5, 2, 1)$ similar to HPT. Alternatively, x_t can be randomly generated from $N(0, 4I_N)$.

the same, (v) when T increases (even from 3 to 6), the bias of CQMLE drops significantly, and (vi) the sds or ses become smaller when N or T or both increase.

Changing the value of m does not affect much the overall performance of MQMLE and FQMLE, but affects significantly the performance of CQMLE. As m decreases, CQMLE improves and when $m = 0$ CQMLE performs quite reasonably even when T is as small as 3, though there still exists some noticeable bias. This is well expected as in this case y_0 is exogenous, and there is no cumulative impact on Δy_1 from the past, which makes Δy_1 much less influential on the subsequent estimation.

The case T is not small. When T is not so small compared with N, the model for Δy_1 can contain many extra parameters, which is exactly $(Tp+2)$ of them where p is the number of the time-varying regressors. In this case, there should be a reasonable concern on the finite sample performance of FQMLE when $|\rho|$ is very small as in this case the 'predicting' power of the model for Δy_1 might not be strong enough, and the extra variability caused by estimating these extra parameters may be too big. To address these issues, Monte Carlo experiments are run with T not so small relative to N and $|\rho|$ is small. The reported results in Table 7 correspond to $(N, T) = (10, 6), (20, 16)$, and $(30, 26)$, and $\rho = 0.1, 0.0$, and -0.1. The results show that MQMLE still performs very well in general, but FQMLE does not perform satisfactorily. MQMLE also outperforms CQMLE, but the latter performs reasonably well. These results confirm our theories that the proposed estimation method works for both small and large T. The ses and rses are not reported as they are unstable for the FQMLEs.

Pure Dynamic Model. Dropping the regressor from the data generating process gives a pure dynamic model. As discussed in Hsiao et al. (2002), and in Section 4 (Footnote 5, in particular), treating ω as an independent parameter is generally valid when model contains at least one time-varying regressor, but may not be so with a pure dynamic model. When $m = 50$, $\omega = c_m(\rho) \approx c_\infty(\rho) = \frac{1}{1+\rho}$, showing that treating ω as an independent parameter and estimating it jointly with the common parameters may generate some undesirable impacts, dependent upon the magnitude of ρ. Table 8 presents some results that typically show that when ρ is large, the performance of FQMLE may not be satisfactorily. In contrast, the proposed MQMLE performs well or satisfactorily in general.

6 Conclusion and Discussion

A simple modification on the score element of dynamic parameter of the conditional quasi-likelihood, conditional on the initial differences Δy_1, of the fixed effects dynamic panel data model is derived. This modification leads to a modified QML estimator of the fixed effects DPD model that is free from the specification of the initial conditions. Formal theories on consistency, asymptotic normality and robust VC matrix are presented. While the method is developed based on short panel set up, it remains valid when the time dimension also grows and more importantly it automatically corrects the bias caused by ignoring the initial observations in the usual large-N and large-T set up, i.e., using the conditional likelihood given the initial differences. An OPG estimator of the robust VC matrix is also given.

To compare the proposed methods with those given in Hsiao et al. (2002), their ML estimation and inference framework is first extended to the QML estimation and inference framework, formal theories on consistency, asymptotic normality and robust VC matrix provided to suite for the situations where the errors are not exactly normally distributed. Monte Carlo results show an excellent finite sample performance of the proposed methods. Compared with the full QMLE where the initial differences are modeled, there is generally no loss in efficiency, but there are gains in terms of reduced bias and increased versatility/stability with respect to the changes in parameter values and structures of the model. In particular, the proposed methods outperforms the full QMLE under the pure DPD model, and under a general model where $|\rho|$ is small and T is not so small relative to N. Considering the simplicity and stability of the proposed methods, it is recommended to the applied researchers.

Some immediate extensions of the methods can be made by relaxing the independence assumption across the time dimension. In particular, the constant C matrix given below (3.1) can be replaced by an parametric matrix capturing the serial correlations. The method may also be extended to allow the errors to be dependent in the cross-sectional direction, e.g., the errors follow a spatial autoregressive process, or the model contains a spatial lag of the response variable. However, doing so clearly invalidates the OPG method for the robust VC matrix estimation, besides the complications in deriving the modification factor. Nevertheless, the results presented in this paper shows that in tackling the initial condition problem for dynamic panel data models with small and fixed T, modifications on the score element for the dynamic parameter seem a promising way to go.

References

[1] Ahn, S., Schmidt, P., 1995. Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68, 5-27.

[2] Alvarez, J., Arellano, M., 2003. The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71, 1121-1159.

[3] Arellano, M., Bover, O., 1995. Another look at the instrumental-variable estimation of error-components models. *Journal of Econometrics* 68, 29-51.

[4] Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58*, 277-297.

[5] Binder, M., Hsiao, C., Pesaran, M. H., 2005. Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory* 21, 795-837.

[6] Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115-143.

[7] Bun, M.J., Carree, M.A., 2005. Bias-corrected estimation in dynamic panel data models. *Journal of Business and Economic Statistics* 23, 200-210.

[8] Bun, M., Kiviet, J., 2006. The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data model. *Journal of Econometrics 132*, 409-444.

[9] Cameron, A. C., Trivedi, P. K. (2005). *Microeconometrics, Methods and Applications.* Cambridge University Press.

[10] Gouriéroux, C. Phillips, P. C. B, Yu, J., 2010. Indirect inference for dynamic panel models. *Journal of Econometrics* 157, 68-77.

[11] Hahn, J., 1997. Efficient estimation of panel data models with sequential moment restrictions. *Journal of Econometrics* 79, 1-21.

[12] Hahn, J., Kuersteiner, G., 2002. Asymptotically unbiased inference for a dynamic model with fixed effects when both n and T are large. *Econometrica* 70, 1639-1657.

[13] Hahn, J., Newey, W., 2004. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72, 1295-1319.

[14] Han, C., Phillips, P.C.B., 2010. GMM estimation for dynamic panels with fixed effects 17 and strong instruments at unity. *Econometric Theory* 26, 119-151.

[15] Holtz-Eakin, D., Newey, W., Rosen, H.S., 1988. Estimating vector autoregressions with panel data. *Econometrica* 56, 1371-1395.

[16] Hsiao, C., Pesaran, M. H., Tahmiscioglu, A. K., 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics 109*, 107-150.

[17] Hsiao, C., 2003. *Analysis of Panel Data, 2nd edition.* Cambridge University Press.

[18] Kiviet, I., 1995. On bias, inconsistency and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68, 53-78.

[19] Kruiniger, H., 2013. Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions. *Journal of Econometrics* 173, 175-188.

[20] Phillips, P.C.B., Sul, D., 2007. Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics* 137, 162-188.

[21] van der Vaart, A. W., 1998. *Asymptotic Statistics.* Cambridge University Press.

[22] Ziliak, J.P., 1997. Efficient estimation with panel data when instruments are predetermined: An empirical comparison of moment-condition estimators. *Journal of Business and Economic Statistics* 15, 419-431.

Appendix: Proofs of Results

Proof of Lemma 3.1: $\mathrm{E}(\Delta Y_{-1}'\Omega^{-1}\Delta v) = \mathrm{tr}[\Omega^{-1}\mathrm{E}(\Delta v \Delta Y_{-1}')]$, and $\mathrm{E}(\Delta v \Delta Y_{-1}') = \{\mathrm{E}(\Delta v_t \Delta y_s')\}$ for $t = 2, 3, \ldots, T$ and $s = 1, 2, \ldots, T-1$. For $t = 2$, $\mathrm{E}(\Delta v_2 \Delta y_1') = \mathrm{E}(\Delta v_2 y_1') = \mathrm{E}(\Delta v_2 v_1') = -\sigma_{v0}^2 I_N$, $\mathrm{E}(\Delta v_2 \Delta y_2') = \rho_0 \mathrm{E}(\Delta v_2 \Delta y_1') + \mathrm{E}(\Delta v_2 \Delta v_2') = (2-\rho_0)\sigma_{v0}^2 I_N$, and for $s = 3, \ldots, T-1$, $\mathrm{E}(\Delta v_2 \Delta y_s') = -\rho_0^{s-3}(1-\rho_0)^2\sigma_{v0}^2 I_N$. Similarly, for $t = 3$, $\mathrm{E}(\Delta v_3 \Delta y_1') = 0$, $\mathrm{E}(\Delta v_3 \Delta y_2') = -\sigma_{v0}^2 I_N$, $\mathrm{E}(\Delta v_3 \Delta y_3') = (2-\rho_0)\sigma_{v0}^2 I_N$, and for $s = 4, \ldots, T-1$, $\mathrm{E}(\Delta v_3 \Delta y_s') = -\rho_0^{s-4}(1-\rho_0)^2\sigma_{v0}^2 I_N$. The other elements of $\{\mathrm{E}(\Delta v_t \Delta y_s')\}$ corresponding to $t = 4, \ldots, T-1$ can be found in the same way, and the result of Lemma 1 thus follows.

Proof of Lemma 3.2: In connection to Hsiao, et al. (2002, p. 143), we first derive a simplified expression for the inverse of C as: $C^{-1} = \frac{1}{T}\{a_{ts},\ \ t, s = 0, 1, \ldots, T-2\}$, where

$$a_{ts} = \begin{cases} (T-t-1)(s+1), & s \le t, \\ (T-s-1)(t+1), & s > t. \end{cases}$$

The rest is just straightforward but tedious algebraic manipulation.

For proving Theorems 3.1 and 3.2, we need the simplified expressions for $\tilde{\beta}^*(\rho)$, $\tilde{\sigma}_v^{*2}(\rho)$, and $S_m^\circ(\rho)$, defined by (3.8) and (3.9), and their population counter parts $\bar{\beta}^*(\rho)$,

$\bar{\sigma}_v^{*2}(\rho)$, and $\bar{S}_c^*(\rho)$, defined above (3.10) and in (3.10). First,

$$\Delta v(\theta) = \Delta Y - \rho \Delta Y_{-1} + \Delta X\beta = \Delta Y_{-1}(\rho_0 - \rho) + \Delta X(\beta_0 - \beta) + \Delta v. \qquad \text{(A.1)}$$

Using (A.1), the random components of the modified score function at ψ are written as

$$\begin{aligned}
\Delta Y_{-1}'\Omega^{-1}\Delta v(\theta) &= \Delta Y_{-1}'\Omega^{-1}\Delta Y_{-1}(\rho_0-\rho) + \Delta Y_{-1}'\Omega^{-1}\Delta X(\beta_0-\beta) + \Delta Y_{-1}'\Omega^{-1}\Delta v,\\
\Delta X'\Omega^{-1}\Delta v(\theta) &= \Delta X'\Omega^{-1}\Delta Y_{-1}(\rho_0-\rho) + \Delta X'\Omega^{-1}\Delta X(\beta_0-\beta) + \Delta X'\Omega^{-1}\Delta v,\\
\Delta v(\theta)'\Omega^{-1}\Delta v(\theta) &= \Delta Y_{-1}'\Omega^{-1}\Delta Y_{-1}(\rho_0-\rho)^2 + (\beta_0-\beta)'\Delta X'\Omega^{-1}\Delta X(\beta_0-\beta)\\
&\quad + \Delta v'\Omega^{-1}\Delta v + 2\Delta Y_{-1}'\Omega^{-1}\Delta X(\rho_0-\rho)(\beta_0-\beta)\\
&\quad + 2\Delta Y_{-1}'\Omega^{-1}\Delta v(\rho_0-\rho) + 2(\beta_0-\beta)'\Delta X'\Omega^{-1}\Delta v,
\end{aligned}$$

which are all linear or quadratic in ρ and β. Let $M = I - \Omega^{-\frac{1}{2}}\Delta X(\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-\frac{1}{2}}$, where I is a $N(T-1)$-dimensional identity matrix, and $\Omega^{-\frac{1}{2}}$ is the symmetric square root of Ω. The modified QMLEs of β_0 and σ_{v0}^2, for a given ρ, defined in (3.8) can be written as,

$$\begin{aligned}
\tilde{\beta}^*(\rho) &= \tilde{\beta}^*(\rho_0) - (\rho-\rho_0)(\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Delta Y_{-1} && \text{(A.2)}\\
\tilde{\sigma}_v^{*2}(\rho) &= \tilde{\sigma}_v^{*2}(\rho_0) + (\rho-\rho_0)^2 Q_1 - 2(\rho-\rho_0)Q_2, && \text{(A.3)}
\end{aligned}$$

where $Q_1 = \frac{1}{N(T-1)}\Delta Y_{-1}'\Omega^{-\frac{1}{2}}M\Omega^{-\frac{1}{2}}\Delta Y_{-1}$ and $Q_2 = \frac{1}{N(T-1)}\Delta Y_{-1}'\Omega^{-\frac{1}{2}}M\Omega^{-\frac{1}{2}}\Delta v$. The modified concentrated estimating function defined in (3.9) can be written as,

$$S_c^*(\rho) = \frac{(\rho_0-\rho)Q_1 + Q_2}{\tilde{\sigma}_v^{*2}(\rho)} - \frac{\mathrm{tr}(C^{-1}D(\rho))}{T-1}. \qquad \text{(A.4)}$$

Now, solving $\mathrm{E}[S_\beta^*(\psi)] = 0$ and $\mathrm{E}[S_{\sigma_v^2}^*(\psi)] = 0$, we obtain, in relation to (3.10),

$$\begin{aligned}
\bar{\beta}^*(\rho) &= \beta_0 - (\rho-\rho_0)(\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}\mathrm{E}(\Delta Y_{-1}), \text{ and} && \text{(A.5)}\\
\bar{\sigma}_v^{*2}(\rho) &= \sigma_{v0}^2 + (\rho-\rho_0)^2[\mathrm{E}(Q_1) + \mathrm{E}(Q_3)] - \tfrac{2\sigma_{v0}^2}{T-1}(\rho-\rho_0)\mathrm{tr}(C^{-1}D), && \text{(A.6)}
\end{aligned}$$

where $Q_3 = \frac{1}{N(T-1)}[(\Delta Y_{-1}' - \mathrm{E}(\Delta Y_{-1}'))\Omega^{-\frac{1}{2}}(I-M)\Omega^{-\frac{1}{2}}(\Delta Y_{-1}' - \mathrm{E}(\Delta Y_{-1}'))]$, and $D = D(\rho_0)$. Substituting $\bar{\beta}^*(\rho)$ and $\bar{\sigma}_v^{*2}(\rho)$ into $\mathrm{E}[S_\rho^*(\psi)]$ for β and σ_v^2 leads to $\bar{S}_c^*(\rho)$ defined by (3.10) as,

$$\bar{S}_c^*(\rho) = \frac{(\rho_0-\rho)[\mathrm{E}(Q_1)+\mathrm{E}(Q_3)]}{\bar{\sigma}_v^{*2}(\rho)} + \frac{\sigma_{v0}^2\mathrm{tr}(C^{-1}D)}{(T-1)\bar{\sigma}_v^{*2}(\rho)} - \frac{\mathrm{tr}(C^{-1}D(\rho))}{(T-1)}. \qquad \text{(A.7)}$$

The following lemma is very useful in simplifying the proofs of the main results.

Lemma A.1: *If the assumptions of Theorem 3.1 hold, we have when N is large,*
(i) $\frac{1}{N}\mathrm{E}(\Delta y_1'\Delta y_1) = O(1)$ *and* $\frac{1}{N}[\Delta y_1'\Delta y_1 - \mathrm{E}(\Delta y_1'\Delta y_1)] = o_p(1)$;
(ii) $\frac{1}{NT}\mathrm{E}(\Delta Y_{-1}'\Omega^{-1}\Delta Y_{-1}) = O(1)$ *and* $\frac{1}{NT}[\Delta Y_{-1}'\Omega^{-1}\Delta Y_{-1} - \mathrm{E}(\Delta Y_{-1}'\Omega^{-1}\Delta Y_{-1})] = o_p(1)$;
(iii) $\frac{1}{NT}\mathrm{E}(\Delta Y_{-1}'\Omega^{-1}\Delta v) = O(1)$ *and* $\frac{1}{NT}[\Delta Y_{-1}'\Omega^{-1}\Delta v - \mathrm{E}(\Delta Y_{-1}'\Omega^{-1}\Delta v)] = o_p(1)$;
(iv) $(\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}\Delta Y_{-1} = O_p(1)$.

Proof: For the first part of (i), we have by (3.12) and Assumption D(ii), $\frac{1}{N}\mathrm{E}(\Delta y_1'\Delta y_1) = \frac{1}{N}\mathrm{tr}[\mathrm{Var}(\Delta y_1) + \mathrm{E}(\Delta y_1)\mathrm{E}(\Delta y_1')] = \sigma_{v0}^2 c_m(\rho_0) + \frac{1}{N}\Delta\eta_1'\Delta\eta_1 = O(1)$. For the second part of (i), we have $\frac{1}{N}(\Delta y_1'\Delta y_1 - \mathrm{E}(\Delta y_1'\Delta y_1)] = \frac{2}{N}\Delta\eta_1'\Delta v_1^\circ + \frac{1}{N}[\Delta v_1^{\circ\prime}\Delta v_1^\circ - \mathrm{E}(\Delta v_1^{\circ\prime}\Delta v_1^\circ)] = o_p(1)$, by Chebyshev inequality, where $\Delta v_1^\circ = \Delta v_1(m, \rho_0)$ to differentiate from $\Delta v_1 = v_1 - v_0$.

For the first part of (ii), similar to (3.12), under Assumption A and Model (2.3), continuous substitutions from the tth period, $t = 2, 3, \ldots, T$, back to the 2th period lead to

$$\Delta y_t = \rho^{t-1}\Delta y_1 + \sum_{s=2}^{t}\rho^{t-s}\Delta x_s\beta + \sum_{s=2}^{t}\rho^{t-s}\Delta v_s. \tag{A.8}$$

Define $a_{-1}(\rho) = (1, \rho, \ldots, \rho^{T-2})'$, and the $(T-1)\times(T-1)$ matrices:

$$A(\rho) = \begin{pmatrix} 0 & 1 & \rho & \ldots & \rho^{T-3} \\ 1 & 0 & 0 & \ldots & 0 \\ \rho & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-3} & 0 & 0 & \ldots & 0 \end{pmatrix} \quad \text{and} \quad B_{-1}(\rho) = \begin{pmatrix} 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 \\ \rho & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{T-3} & \rho^{T-4} & \ldots & 1 & 0 \end{pmatrix},$$

and denote $a_{-1} \equiv a_{-1}(\rho_0)$, $A \equiv A(\rho_0)$, $B_{-1} \equiv B_{-1}(\rho_0)$, and $\Delta\eta = \Delta X\beta_0$. We have

$$\begin{aligned} \Delta Y_{-1} &= a_{-1}\otimes\Delta y_1 + (B_{-1}\otimes I_N)(\Delta\eta + \Delta v), \quad \text{and} & \text{(A.9)} \\ \mathrm{Var}(\Delta Y_{-1}) &= \sigma_{v0}^2[c_m(\rho_0)a_{-1}a_{-1}' + B_{-1}CB_{-1}' + A]\otimes I_N. & \text{(A.10)} \end{aligned}$$

It is obvious that $\lambda_{\min}(\Omega) = \lambda_{\min}(C) = c_{\min} > 0$ and $\lambda_{\max}(\Omega) = \lambda_{\max}(C) = c_{\max} < \infty$. Thus, $\frac{1}{NT}\mathrm{E}(\Delta Y_{-1}'\Omega^{-1}\Delta Y_{-1}) = \frac{1}{NT}\mathrm{tr}[\Omega^{-1}\mathrm{E}(\Delta Y_{-1}\Delta Y_{-1}')] \le \frac{1}{NT}c_{\min}^{-1}\mathrm{tr}[\mathrm{E}(\Delta Y_{-1}\Delta Y_{-1}')] = O(1)$, by (i), (A.9) and (A.10). For the second part of (ii), we have by (A.9) and $\Omega = C\otimes I_N$,

$$\begin{aligned} &\tfrac{1}{NT}[\Delta Y_{-1}'\Omega^{-1}\Delta Y_{-1} - \mathrm{E}(\Delta Y_{-1}'\Omega^{-1}\Delta Y_{-1})] = \textstyle\sum_{k-1}^{5} R_k, \text{ where} \\ &\quad R_1 = \tfrac{1}{NT}(a_{-1}'a_{-1})[\Delta y_1'\Delta y_1 - \mathrm{E}(\Delta y_1'\Delta y_1)], \\ &\quad R_2 = \tfrac{1}{NT}\{\Delta v'((B_{-1}'C^{-1}B_{-1})\otimes I_N)\Delta v - \mathrm{E}[\Delta v'((B_{-1}'C^{-1}B_{-1})\otimes I_N)\Delta v]\}, \\ &\quad R_3 = \tfrac{2}{NT}[(a_{-1}'C^{-1}B_{-1})\otimes(\Delta y_1' - \mathrm{E}(\Delta y_1'))]\Delta\eta, \\ &\quad R_4 = \tfrac{2}{NT}\{((a_{-1}'C^{-1}B_{-1})\otimes\Delta y_1')\Delta v - \mathrm{E}[((a_{-1}'C^{-1}B_{-1})\otimes\Delta y_1')\Delta v]\}, \text{ and} \\ &\quad R_5 = \tfrac{2}{NT}\Delta\eta'[(B_{-1}'C^{-1}B_{-1})\otimes I_N)]\Delta v. \end{aligned}$$

Using (3.12), (i) and Chebyshev inequality, it is straightforward to show that $R_k = o_p(1), k = 1, \ldots, 5$, leading to second part of (ii). The proof of (iii) follows closely that of (ii).

For (iv), note that $c_{\max}^{-1}\Delta X'\Delta X \le \Delta X'\Omega^{-1}\Delta X \le c_{\min}^{-1}\Delta X'\Delta X$. It follows by Assumption D(iii) that $NT(\Delta X'\Omega^{-1}\Delta X)^{-1} = O(1)$. It is easy to show that $\frac{1}{NT}(\Delta X'\Omega^{-1}\Delta Y_{-1}) = O_p(1)$ by (A.9), the result (iv) thus follows. ■

Proof of Theorem 3.1: We first prove the consistency of $\hat{\rho}^*$. By Theorem 5.9 of van der Vaart (1998, p. 46), it amounts to show (i) $\sup_{\rho\in\mathcal{P}} \left|S_c^*(\rho) - \bar{S}_c^*(\rho)\right| \xrightarrow{p} 0$, and (ii) for every $\epsilon > 0$, $\inf_{\rho:\ |\rho-\rho_0|\geq\epsilon} \left|\bar{S}_c^*(\rho)\right| > 0 = \left|\bar{S}_c^*(\rho_0)\right|$. To show (i), we have by (A.4) and (A.7),

$$\begin{aligned} S_c^*(\rho) - \bar{S}_c^*(\rho) &= -(\rho-\rho_0)\frac{Q_1 - \mathrm{E}(Q_1) - \mathrm{E}(Q_3)}{\bar{\sigma}_v^{*2}(\rho)} + \frac{Q_2 - \frac{\sigma_{v0}^2}{T-1}\mathrm{tr}(C^{-1}D)}{\bar{\sigma}_v^{*2}(\rho)} \\ &\quad +[(\rho-\rho_0)Q_1 + Q_2]\frac{\tilde{\sigma}_v^{*2}(\rho) - \bar{\sigma}_v^{*2}(\rho)}{\tilde{\sigma}_v^{*2}(\rho)\bar{\sigma}_v^{*2}(\rho)}. \end{aligned}$$

Thus, it suffices to show that $Q_1 - \mathrm{E}(Q_1) = o_p(1)$, $Q_2 - \frac{\sigma_{v0}^2}{T-1}\mathrm{tr}(C^{-1}D) = o_p(1)$, $\mathrm{E}(Q_3) = o(1)$, $\tilde{\sigma}_v^{*2}(\rho) - \bar{\sigma}_v^{*2}(\rho) = o_p(1)$ uniformly in $\rho \in \mathcal{P}$, and $\tilde{\sigma}_v^{*2}(\rho)$ is bounded away from zero with probability one uniformly in $\rho \in \mathcal{P}$ for large enough N. It is easy to show that $Q_1 = \frac{1}{N(T-1)}\Delta Y_{-1}'\Omega^{-1}\Delta Y_{-1} + o_p(1)$ and $Q_2 = \frac{1}{N(T-1)}\Delta Y_{-1}'\Omega^{-1}\Delta v + o_p(1)$. Thus, the first two results follow from Lemma A.1(i)-(ii). For the third result we have by (A.10),

$$\begin{aligned} \mathrm{E}(Q_3) &= \tfrac{1}{N(T-1)}\mathrm{E}[(\Delta Y_{-1}' - \mathrm{E}(\Delta Y_{-1}'))\Omega^{-\frac{1}{2}}(I-M)\Omega^{-\frac{1}{2}}(\Delta Y_{-1}' - \mathrm{E}(\Delta Y_{-1}'))] \\ &= \tfrac{1}{N(T-1)}\mathrm{tr}\{\Omega^{-\frac{1}{2}}(I-M)\Omega^{-\frac{1}{2}}\mathrm{Var}(\Delta Y_{-1})\} \\ &\leq \tfrac{1}{N(T-1)}\lambda_{\max}[\mathrm{Var}(\Delta Y_{-1})][\lambda_{\min}(\Omega)]^{-1}\mathrm{tr}(I-M) = o(1), \end{aligned}$$

where note that $\mathrm{tr}(I-M) = p$. For the fourth result, we have by (A.3) and (A.6),

$$\begin{aligned} \tilde{\sigma}_v^{*2}(\rho) - \bar{\sigma}_v^{*2}(\rho) &= \tilde{\sigma}_v^{*2}(\rho_0) - \sigma_{vo}^2 + (\rho-\rho_0)^2[Q_1 - \mathrm{E}(Q_1) - \mathrm{E}(Q_3)] \\ &\quad +2(\rho-\rho_0)[Q_2 - \tfrac{\sigma_{v0}^2}{T-1}\mathrm{tr}(D^{-1}D)]. \end{aligned}$$

It is easy to show $\tilde{\sigma}_v^{*2}(\rho_0) = \frac{1}{N(T-1)}\Delta v'\Omega^{-\frac{1}{2}}M\Omega^{-\frac{1}{2}}\Delta v \xrightarrow{p} \sigma_{v0}^2$ and $Q_2 - \frac{\sigma_{v0}^2}{T-1}\mathrm{tr}(D^{-1}D) \xrightarrow{p} 0$. Further, $Q_1 - \mathrm{E}(Q_1) = o_p(1)$ and $\mathrm{E}(Q_3) = o(1)$ as shown above. Thus, $\tilde{\sigma}_v^{*2}(\rho) - \bar{\sigma}_v^{*2}(\rho) = o_p(1)$ uniformly in $\rho \in \mathcal{P}$. To show the last result, note from (3.8),

$$\begin{aligned} \tilde{\sigma}_v^{*2}(\rho) &= \tfrac{1}{N(T-1)}\Delta Y(\rho)'\Omega^{-\frac{1}{2}}M\Omega^{-\frac{1}{2}}\Delta Y(\rho) \\ &= \tfrac{1}{N(T-1)}\Delta Y(\rho)'\Omega^{-1}\Delta Y(\rho) + o_p(1) \\ &\geq \tfrac{1}{N(T-1)}c_{\max}^{-1}\Delta Y(\rho)'\Delta Y(\rho) + o_p(1). \end{aligned}$$

As $c_{\max}^{-1}$ is strictly positive and $\frac{1}{N(T-1)}\Delta Y(\rho)'\Delta Y(\rho)$ is positive with probability one uniformly in $\rho \in \mathcal{P}$, $\tilde{\sigma}_v^{*2}(\rho)$ is bounded away from zero with probability one uniformly in $\rho \in \mathcal{P}$. As $\tilde{\sigma}_v^{*2}(\rho) - \bar{\sigma}_v^{*2}(\rho) = o_p(1)$ uniformly in $\rho \in \mathcal{P}$, $\bar{\sigma}_v^{*2}(\rho)$ is bounded away from zero uniformly in $\rho \in \mathcal{P}$, and this finishes the proof of (i).

To show (ii), first note from (A.5)-(A.7) that $\bar{\beta}^*(\rho_0) = \beta_0$, $\bar{\sigma}_v^{*2}(\rho_0) = \sigma_{v0}^2$ and $\bar{S}_c^*(\rho_0) = 0$. To show that ρ_0 is the unique zero of $\bar{S}_c^*(\rho)$, it is equivalent to show that ρ_0 is the unique zero of $(\rho_0 - \rho)\mathrm{E}(Q_1) + \bar{\sigma}_v^{*2}(\rho)g(\rho) - \sigma_{v0}^2 g(\rho_0)$, which is given by Assumption D.

Now, for the consistency of $\hat{\beta}^*$, note $\hat{\beta}^* = \tilde{\beta}(\rho_0) - (\hat{\rho}^* - \rho_0)(\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}\Delta Y_{-1}$. It is easy to show that $(\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}\Delta v = o_p(1)$ by Assumption D, and thus as $N \to \infty$, $\tilde{\beta}(\rho_0) \xrightarrow{p} \beta_0$. By Lemma A.1(*iii*), $(\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}\Delta Y_{-1} = O_p(1)$. Since $\hat{\rho}^* \xrightarrow{p} \rho_0$, it follows that $\hat{\beta}^* \xrightarrow{p} \beta_0$. Finally, $\hat{\sigma}_v^{*2} = \tilde{\sigma}_v^{*2}(\rho_0) + (\hat{\rho}^* - \rho_0)^2 Q_1 + 2(\hat{\rho}^* - \rho_0)Q_2 \xrightarrow{p} \sigma_{v0}^2$, as $\tilde{\sigma}_v^{*2}(\rho_0) \xrightarrow{p} \sigma_{v0}^2$, $\hat{\rho}^* \xrightarrow{p} \rho_0$, and both Q_1 and Q_2 are $O_p(1)$ by Lemma A.1(*i*).

To establish the asymptotic normality for $\hat{\psi}^*$, we have by the mean value theorem,

$$0 = \frac{1}{\sqrt{N(T-1)}}S^*(\hat{\psi}^*) = \frac{1}{\sqrt{N(T-1)}}S^*(\psi_0) + \frac{1}{N(T-1)}H^*(\tilde{\psi})\sqrt{N(T-1}(\hat{\psi}^* - \psi_0)$$

where the elements of $\tilde{\psi} = (\tilde{\rho}, \tilde{\beta}, \tilde{\sigma}_v^2)'$ lie in the segment joining the corresponding elements of $\hat{\psi}^*$ and ψ_0. As $\hat{\psi}^* \xrightarrow{p} \psi_0$, it must be that $\tilde{\psi} \xrightarrow{p} \psi_0$. Thus,

$$\sqrt{N(T-1}(\hat{\psi}^* - \psi_0) = \left[\tfrac{1}{N(T-1)}H^*(\tilde{\psi})\right]^{-1}\frac{1}{\sqrt{N(T-1)}}S^*(\psi_0).$$

We need to show: (*i*) $\frac{1}{N(T-1)}[H^*(\tilde{\psi}) - H^*(\psi_0)] = o_p(1)$, (*ii*) $\frac{1}{N(T-1)}[H^*(\psi_0) - \mathrm{E}(H^*(\psi_0))] = o_p(1)$, and (*iii*) $\frac{1}{\sqrt{N(T-1)}}S^*(\psi_0) \xrightarrow{D} N(0, \lim_{N\to\infty}\Gamma^*(\psi_0))$. The proofs of (*i*) and (*ii*) are straightforward applications of Lemma A.1. Finally, the proof of (*iii*) amounts to check the conditions of Lindeberg-Feller central limit theorem for a sum of independent random vectors (see, e.g., van der Vaart, 1998, p. 20), applied to $S^*(\psi_0) = G^*(\psi_0)1_N$ defined in (3.16). These conditions are implied by the existence of $\Gamma^*(\psi_0)$ for large enough N. ■

Proof of Theorem 3.2: First, the convergence result, $\frac{1}{N(T-1)}H^*(\hat{\psi}^*) - \Sigma^*(\psi_0) \xrightarrow{p} 0$, is proved in the proof of Theorem 3.1. To show $\frac{1}{N(T-1)}G^*(\hat{\psi}^*)'G^*(\hat{\psi}^*) - \Gamma^*(\psi_0) \xrightarrow{p} 0$, we have from (3.13)-(3.16), $\mathrm{Var}[S^*(\psi_0)] = \mathrm{Var}[G^*(\psi_0)'1_N] = \sum_{i=1}^N \mathrm{Var}[g_i^*(\psi_0)]$, where

$$g_i^*(\psi_0) = \left(\tfrac{1}{\sigma_{v0}^2}g_{\rho,i}^* - \mathrm{tr}(C^{-1}D),\ \ \tfrac{1}{\sigma_{v0}^2}g_{\beta,i}^*,\ \ \tfrac{1}{2\sigma_{v0}^4}g_{\sigma_v^2,i}^* - \tfrac{T-1}{2\sigma_{v0}^2}\right)',$$

and $g_{\rho,i}^* \equiv g_{\rho,i}^*(\theta_0)$, $g_{\beta,i}^* \equiv g_{\beta,i}^*(\theta_0)$, and $g_{\sigma_v^2,i}^* \equiv g_{\sigma_v^2,i}^*(\theta_0)$, defined in (3.13)-(3.15). Thus,

$$\mathrm{Var}[S^*(\psi_0)] = \begin{pmatrix} \frac{1}{\sigma_{v0}^4}\sum_{i=1}^N \mathrm{Var}(g_{\rho,i}^*) & \frac{2}{\sigma_{v0}^4}\sum_{i=1}^N \mathrm{Cov}(g_{\rho,i}^*, g_{\beta,i}^*) & \frac{1}{\sigma_{v0}^6}\sum_{i=1}^N \mathrm{Cov}(g_{\rho,i}^*, g_{\sigma_v^2,i}^*) \\ \sim & \frac{1}{\sigma_{v0}^4}\sum_{i=1}^N \mathrm{Var}(g_{\beta,i}^*) & \frac{1}{\sigma_{v0}^6}\sum_{i=1}^N \mathrm{Cov}(g_{\beta,i}^*, g_{\sigma_v^2,i}^*) \\ \sim & \sim & \frac{1}{\sigma_{v0}^8}\sum_{i=1}^N \mathrm{Var}(g_{\sigma_v^2,i}^*) \end{pmatrix}.$$

Now, $G^*(\hat{\psi}^*)'G^*(\hat{\psi}^*) = \sum_{i=1}^N g_i^*(\hat{\psi}^*)g_i^*(\hat{\psi}^*)$, which takes a similar form as $\mathrm{Var}[S^*(\psi_0)]$, but with 'Var' and 'Cov' replaced by their sample versions, and ψ_0 by $\hat{\psi}^*$. As $\hat{\sigma}_v^{*,2}$ is consistent for σ_{v0}^2, and so is its power for the corresponding power of σ_{v0}^2, it suffices to show that

(*i*) $\frac{1}{N}\sum_{i=1}^N [g_{\rho,i}^{*2}(\hat{\theta}^*) - \mathrm{E}(g_{\rho,i}^{*2}(\theta_0))] \xrightarrow{p} 0$,

(ii) $\frac{1}{N}\sum_{i=1}^{N}[g^*_{\beta,i}(\hat\theta^*)g^*_{\beta,i}(\hat\theta^*)' - \mathrm{E}(g^*_{\beta,i}(\theta_0))g^*_{\beta,i}(\theta_0))'] \xrightarrow{p} 0$,
(iii) $\frac{1}{N}\sum_{i=1}^{N}[g^{*2}_{\sigma_v^2,i}(\hat\theta^*) - \mathrm{E}(g^{*2}_{\sigma_v^2,i}(\theta_0))] \xrightarrow{p} 0$,
(iv) $\frac{1}{N}\sum_{i=1}^{N}[g^*_{\rho,i}(\hat\theta^*)g^*_{\beta,i}(\hat\theta^*)' - \mathrm{E}(g^*_{\rho,i}(\theta_0))g^*_{\beta,i}(\theta_0))'] \xrightarrow{p} 0$,
(v) $\frac{1}{N}\sum_{i=1}^{N}[g^*_{\rho,i}(\hat\theta^*)g^*_{\sigma_v^2,i}(\hat\theta^*) - \mathrm{E}(g^*_{\rho,i}(\theta_0))g^*_{\sigma_v^2,i}(\theta_0))] \xrightarrow{p} 0$, and
(vi) $\frac{1}{N}\sum_{i=1}^{N}[g^*_{\beta,i}(\hat\theta^*)g^*_{\sigma_v^2,i}(\hat\theta^*) - \mathrm{E}(g^*_{\beta,i}(\theta_0))g^*_{\sigma_v^2,i}(\theta_0))] \xrightarrow{p} 0$,

where in the elements involving $g^{*2}_{\rho,i}(\hat\theta^*)$ and $g^{*2}_{\sigma_v^2,i}(\hat\theta^*)$, we have used the easily proved results $\frac{1}{N}\sum_{i=1}^{N} g^*_{\rho,i}(\hat\theta^*) \xrightarrow{p} \sigma^2_{v0}\mathrm{tr}(C^{-1}D)$ and $\frac{1}{N}\sum_{i=1}^{N} g^*_{\sigma_v^2,i}(\hat\theta^*) \xrightarrow{p} (T-1)\sigma^2_{v0}$.

We give a detailed proof of (i) as it is one of the most complicated cases due to the involvement of $\Delta y_1 = \{\Delta y_{i1}, i = 1,\ldots,N\}'$. The proofs of the rest are similar or easier. From (3.13), $g^*_{\rho,i}(\hat\theta^*) = \Delta y'_{i-}C^{-1}\Delta v_{i-}(\hat\theta^*)$, and similar to (A.1),

$$\Delta v_{i-}(\hat\theta^*) = -\Delta y_{i-}(\hat\rho^* - \rho_0) - \Delta x_{i-}(\hat\beta^* - \beta_0) + \Delta v_{i-},$$

noting that $\Delta v_{i-} = \Delta v_{i-}(\theta_0)$. These lead to,

$$\begin{aligned}
& \tfrac{1}{N}\textstyle\sum_{i=1}^{N}[g^{*2}_{\rho,i}(\hat\theta^*) - \mathrm{E}(g^{*2}_{\rho,i}(\theta_0))] \\
= \ & \tfrac{1}{N}\textstyle\sum_{i=1}^{N}[(\Delta y'_{i-}C^{-1}\Delta v_{i-})^2 - \mathrm{E}((\Delta y'_{i-}C^{-1}\Delta v_{i-})^2)] \\
& +(\hat\rho^* - \rho_0)^2\tfrac{1}{N}\textstyle\sum_{i=1}^{N}(\Delta y'_{i-}C^{-1}\Delta y_{i-})^2 \\
& +(\hat\beta^* - \beta_0)'\tfrac{1}{N}\textstyle\sum_{i=1}^{N}(\Delta x'_{i-}C^{-1}\Delta y_{i-}\Delta y'_{i-}C^{-1}\Delta x_{i-})(\hat\beta^* - \beta_0) \\
& +(\hat\rho^* - \rho_0)\tfrac{2}{N}\textstyle\sum_{i=1}^{N}(\Delta y'_{i-}C^{-1}\Delta y_{i-}\Delta y'_{i-}C^{-1}\Delta x_{i-})(\hat\beta^* - \beta_0) \\
& -(\hat\rho^* - \rho_0)\tfrac{2}{N}\textstyle\sum_{i=1}^{N}(\Delta y'_{i-}C^{-1}\Delta y_{i-})(\Delta y'_{i-}C^{-1}\Delta v_{i-}) \\
& -(\hat\beta^* - \beta_0)'\tfrac{1}{N}\textstyle\sum_{i=1}^{N}(\Delta x'_{i-}C^{-1}\Delta y_{i-}\Delta y'_{i-}C^{-1}\Delta v_{i-}).
\end{aligned}$$

To show that the first term is $o_p(1)$, we have by (A.9),

$$\Delta y_{i-} = \Delta y_{i1}a_{-1} + B_{-1}\Delta\eta_{i-} + B_{-1}\Delta v_{i-},$$

where $\eta_{i-} = (\eta_{i2}, \eta_{i3}, \ldots, \eta_{i,T})'$. By (4.1), $\Delta y_{i1} = \Delta\eta_{i1} + \Delta v^{\circ}_{i1}$, which are, respectively, the elements of $\Delta\eta_1 = \rho_0^m\phi(y_{-m}) + \Delta x_1(m,\rho_0)$ and $\Delta v^{\circ}_1 = \rho_0^m v_{-m+1} + \Delta v_1(m,\rho_0)$. Thus,

$$\Delta y_{i-} = (\Delta\eta_{i1}a_{-1} + B_{-1}\Delta\eta_{i-}) + (\Delta v^{\circ}_{i1}a_{-1} + B_{-1}\Delta v_{i-}) \equiv \Delta\mu_i + \Delta\varepsilon_i, \text{ and}$$

$$\begin{aligned}
\tfrac{1}{N}\textstyle\sum_{i=1}^{N}(\Delta y'_{i-}C^{-1}\Delta v_{i-})^2 \ = \ & \tfrac{1}{N}\textstyle\sum_{i=1}^{N}(\Delta\mu'_{i-}C^{-1}\Delta v_{i-})^2 + \tfrac{1}{N}\sum_{i=1}^{N}(\Delta\varepsilon'_{i-}C^{-1}\Delta v_{i-})^2 \\
& +\tfrac{2}{N}\textstyle\sum_{i=1}^{N}(\Delta\mu'_{i-}C^{-1}\Delta v_{i-})(\Delta\varepsilon'_{i-}C^{-1}\Delta v_{i-}).
\end{aligned}$$

It can shown that $\mathrm{Var}[\frac{1}{N}\sum_{i=1}^{N}(\Delta\mu'_{i-}C^{-1}\Delta v_{i-})^2] = o(1)$, and thus by Chebyshev inequality, $\frac{1}{N}\sum_{i=1}^{N}[(\Delta\mu'_{i-}C^{-1}\Delta v_{i-})^2 - \mathrm{E}((\Delta\mu'_{i-}C^{-1}\Delta v_{i-})^2)] = o_p(1)$. By Kolmogorov law of large number, $\frac{1}{N}\sum_{i=1}^{N}(\Delta\varepsilon'_{i-}C^{-1}\Delta v_{i-})^2 \xrightarrow{p} c$, as it is a sum of iid random variables with a finite mean, say c. Finally, by Cauchy-Schwartz inequality, $\frac{1}{N}\sum_{i=1}^{N}[(\Delta\mu_{i-}'C^{-1}\Delta v_{i-})(\Delta\varepsilon_{i-}'C^{-1}\Delta v_{i-}) - \mathrm{E}(\Delta\mu_{i-}'C^{-1}\Delta v_{i-})(\Delta\varepsilon_{i-}'C^{-1}\Delta v_{i-})] = o_p(1)$. These lead to

$$\frac{1}{N}\sum_{i=1}^{N}[(\Delta y_{i-}'C^{-1}\Delta v_{i-})^2 - \mathrm{E}((\Delta y_{i-}'C^{-1}\Delta v_{i-})^2)] = o_p(1).$$

Now, by the representation for Δy_{i-} given above and by Assumption C, the summations in all the other five terms are easily seen to be $O_p(1)$, and thus these five terms are all $o_p(1)$ as $\hat{\rho}^* - \rho_0 = o_p(1)$ and $\hat{\beta}^* - \beta_0 = o_p(1)$. ■

Proof of Theorem 4.1: The consistency of $(\hat{\rho}^\diamond, \hat{\beta}^\diamond, \hat{\sigma}_v^\diamond)$ of the structural parameters $(\rho, \beta, \sigma_v^2)$ depends crucially on the consistency of $\hat{\omega}^\diamond$, and the latter is proved by applying Theorem 5.7 of van der Vaart (1998). Let $\bar{\ell}_c^\diamond(\omega) = \max_{\theta,\sigma_v^2}\mathrm{E}[\ell(\psi)]$. It suffices to show that (i) $\sup_{\omega\in\mathcal{W}}|\ell_c^\diamond(\omega) - \bar{\ell}_c^\diamond(\omega)| = o_p(1)$, and (ii) for every $\epsilon > 0$, $\sup_{|\omega-\omega_0|\geq\epsilon}\bar{\ell}_c^\diamond(\omega) < \bar{\ell}_c^\diamond(\omega_0)$.

To show (i), let $A(\omega) = \Delta X^{\diamond\prime}\Omega^\diamond(\omega)^{-1}\Delta X^\diamond$ and $b(\omega) = \Delta X^{\diamond\prime}\Omega^\diamond(\omega)^{-1}\Delta v^\diamond$. It is easy to show that the optimal solution to $\bar{\ell}_c^\diamond(\omega) = \max_{\beta,\sigma_v^2}\mathrm{E}[\ell^\diamond(\psi)]$ is: $\bar{\beta}^\diamond(\omega) = \beta_0 + A(\omega)^{-1}b(\omega)$ and $\bar{\sigma}_v^{\diamond 2}(\omega) = \frac{\sigma_{v0}^2}{NT}\mathrm{tr}[\Omega^\diamond(\omega)^{-1}\Omega^\diamond] - \frac{1}{NT}\mathrm{E}[b(\omega)']\mathrm{E}[A(\omega)]^{-1}\mathrm{E}[b(\omega)]$, leading to

$$\bar{\ell}_c^\diamond(\omega) = -\frac{NT}{2}[\log(2\pi) + 1] - \frac{1}{2}\left|\bar{\sigma}_v^{\diamond 2}(\omega)\Omega^\diamond(\omega)\right|.$$

Thus, $\ell_c^\diamond(\omega) - \bar{\ell}_c^\diamond(\omega) = -\frac{NT}{2}[\log\hat{\sigma}_v^{\diamond 2}(\omega) - \log\bar{\sigma}_v^{\diamond 2}(\omega)]$. It suffices to show that $\hat{\sigma}_v^{\diamond 2}(\omega) - \bar{\sigma}_v^{\diamond 2}(\omega)$ converges to zero uniformly in $\omega \in \mathcal{W}$ and that $\bar{\sigma}_v^{\diamond 2}(\omega)$ is bounded away from zero uniformly in $\omega \in \mathcal{W}$. Note $\hat{\sigma}_v^{\diamond 2}(\omega) = \frac{1}{NT}\Delta v^{\diamond\prime}\Omega^\diamond(\omega)^{-1}\Delta v^\diamond - \frac{1}{NT}b(\omega)'A(\omega)^{-1}b(\omega)$. These lead to,

$$\begin{aligned}\hat{\sigma}_v^{\diamond 2}(\omega) - \bar{\sigma}_v^{\diamond 2}(\omega) &= \frac{1}{NT}\left\{\Delta v^{\diamond\prime}\Omega^\diamond(\omega)^{-1}\Delta v^\diamond - \sigma_{v0}^2\mathrm{tr}[\Omega^\diamond(\omega)\Omega^\diamond]\right\} \\ &\quad -\frac{1}{NT}\left\{b(\omega)'A(\omega)^{-1}b(\omega) - \mathrm{E}[b(\omega)']\mathrm{E}[A(\omega)]^{-1}\mathrm{E}[b(\omega)]\right\}.\end{aligned}$$

We have, $\frac{1}{NT}\left|\Delta v^{\diamond\prime}\Omega^\diamond(\omega)^{-1}\Delta v^\diamond - \sigma_{v0}^2\mathrm{tr}[\Omega^\diamond(\omega)^{-1}\Omega^\diamond]\right| \leq \lambda_{\min}^{-1}[\Omega^\diamond(\omega)]\frac{1}{NT}\left|\Delta v^{\diamond\prime}\Delta v^\diamond - \sigma_{v0}^2\mathrm{tr}(\Omega^\diamond)\right|$. As $\Omega^\diamond(\omega) = C^\diamond(\omega) \otimes I_N, \Omega^\diamond(\omega)^{-1}$ and $C^\diamond(\omega)$ is positive definite uniformly in $\omega \in \mathcal{W}$, $\lambda_{\min}[\Omega^\diamond(\omega)$ is strictly positive, and hence $\frac{1}{NT}\left|\Delta v^{\diamond\prime}\Omega^\diamond(\omega)^{-1}\Delta v^\diamond - \sigma_{v0}^2\mathrm{tr}[\Omega^\diamond(\omega)^{-1}\Omega^\diamond]\right| = o_p(1)$ uniformly in $\omega \in \mathcal{W}$. Now, similar to the proof of Lemma 3.2, we express the inverse of $C^\diamond(\omega)$ simply as: $C^\diamond(\omega)^{-1} = \frac{1}{1+T(\omega-1)}\{\ a_{ts}(\omega)\ \}_{T\times T}$, where

$$a_{ts}(\omega)\begin{cases}(T-t)(s\omega-(s-1)),\ s\leq t,\\ (T-s)(t\omega-(t-1)),\ s>t,\end{cases} \qquad t,s = 0,1,2,\ldots,T-1.$$

The representation greatly facilitates the proof of uniform convergence of the second term, i.e., $\frac{1}{NT}\left\{b(\omega)'A(\omega)^{-1}b(\omega) - \mathrm{E}[b(\omega)']\mathrm{E}[A(\omega)]^{-1}\mathrm{E}[b(\omega)]\right\} = o_p(1)$ uniformly in $\omega \in \mathcal{W}$. The detail is tedious but straightforward, and is made available from the author upon request.

To show (ii), if $\Delta v^\diamond$ were normal, i.e., $\Delta v^\diamond \sim N(0, \sigma_{v0}\Omega^\diamond)$, then $\mathrm{Exp}[\ell^\diamond(\psi_0)]$ would be the true pdf of $\Delta v^\diamond$. By Jensen inequality, we have $\mathrm{E}[\ell^\diamond(\psi)] \leq \mathrm{E}[\ell^\diamond(\psi_0)]$. It can be easily seen that $\mathrm{E}[\ell^{\diamond}(\psi_0)] = \bar{\ell}_c^{\diamond}(\omega_0)$. Now, if the true distribution of $\Delta v^{\diamond}$ is not normal but has the same first two moments as normal, i.e., $\Delta v^{\diamond} \sim (0, \sigma_{v0}\Omega^{\diamond})$, we see that the inequality, $\mathrm{E}[\ell^{\diamond}(\psi)] \leq \mathrm{E}[\ell^{\diamond}(\psi_0)]$, still holds as $\mathrm{E}[\ell^{\diamond}(\psi_0)]$ depends on only the first two moments of $\Delta v^{\diamond}$. Thus, $\mathrm{E}[\ell^{\diamond}(\psi)] \leq \mathrm{E}[\ell^{\diamond}(\psi_0)] = \bar{\ell}_c^{\diamond}(\omega_0)$, and hence $\bar{\ell}^{\diamond}(\omega) \leq \bar{\ell}_c^{\diamond}(\omega_0)$, as long as $\Delta v^{\diamond} \sim (0, \sigma_{v0}\Omega^{\diamond})$. This and Assumption E$'$ lead to the result *(ii)*.

The proof of asymptotic normality is analogous to that of Theorem 3.1. ■

Proof of Theorem 4.2: The proof is analogous to that of Theorem 3.2. ■

Table 1. Empirical Mean(sd)[$\widehat{se}$]{$\widehat{rse}$} of CMLE, MMLE and FMLE, $N = 50, T = 3$

dgp	par	CMLE	MMLE	FMLE
1	**0.8**	0.6562 (.0535)	0.8049 (.0684)[.0626]{.0644}	0.8047 (.0681)[.0657]{.0649}
	1.0	0.9415 (.0571)	1.0015 (.0619)[.0585]{.0573}	1.0014 (.0619)[.0591]{.0575}
	1.0	0.9196 (.1308)	0.9907 (.1550)[.1523]{.1495}	0.9907 (.1548)[.1535]{.1497}
2	**0.8**	0.6550 (.0652)	0.8016 (.0751)[.0620]{.0699}	0.8010 (.0731)[.0649]{.0684}
	1.0	0.9404 (.0576)	0.9999 (.0620)[.0581]{.0576}	0.9997 (.0617)[.0586]{.0573}
	1.0	0.9149 (.2471)	0.9883 (.2887)[.1526]{.2684}	0.9877 (.2874)[.1538]{.2669}
3	**0.8**	0.6571 (.0590)	0.8036 (.0735)[.0620]{.0688}	0.8029 (.0717)[.0649]{.0681}
	1.0	0.9442 (.0564)	1.0032 (.0620)[.0581]{.0592}	1.0030 (.0617)[.0586]{.0590}
	1.0	0.9111 (.1877)	0.9820 (.2200)[.1511]{.2060}	0.9813 (.2188)[.1521]{.2051}
1	**0.4**	0.2833 (.0516)	0.4004 (.0606)[.0582]{.0587}	0.4009 (.0608)[.0595]{.0589}
	1.0	0.9725 (.0533)	1.0013 (.0553)[.0543]{.0530}	1.0015 (.0554)[.0544]{.0530}
	1.0	0.9282 (.1377)	0.9752 (.1528)[.1455]{.1409}	0.9758 (.1529)[.1461]{.1413}
2	**0.4**	0.2846 (.0597)	0.4024 (.0673)[.0582]{.0619}	0.4018 (.0642)[.0594]{.0600}
	1.0	0.9723 (.0537)	1.0015 (.0558)[.0542]{.0527}	1.0014 (.0557)[.0543]{.0526}
	1.0	0.9343 (.2646)	0.9847 (.2931)[.1476]{.2628}	0.9844 (.2929)[.1479]{.2622}
3	**0.4**	0.2832 (.0556)	0.4017 (.0645)[.0585]{.0620}	0.4017 (.0631)[.0597]{.0609}
	1.0	0.9718 (.0550)	1.0007 (.0564)[.0545]{.0538}	1.0007 (.0561)[.0545]{.0536}
	1.0	0.9393 (.2083)	0.9888 (.2313)[.1479]{.2031}	0.9888 (.2308)[.1483]{.2030}
1	**0.0**	-0.0848 (.0510)	0.0023 (.0564)[.0537]{.0531}	0.0025 (.0560)[.0540]{.0530}
	1.0	0.9924 (.0531)	0.9992 (.0535)[.0526]{.0514}	0.9992 (.0535)[.0526]{.0514}
	1.0	0.9490 (.1393)	0.9781 (.1480)[.1429]{.1379}	0.9783 (.1482)[.1430]{.1380}
2	**0.0**	-0.0839 (.0547)	0.0034 (.0592)[.0535]{.0540}	0.0027 (.0572)[.0537]{.0526}
	1.0	0.9923 (.0534)	0.9992 (.0544)[.0522]{.0505}	0.9991 (.0544)[.0522]{.0505}
	1.0	0.9467 (.2637)	0.9774 (.2808)[.1431]{.2569}	0.9769 (.2793)[.1429]{.2563}
3	**0.0**	-0.0860 (.0504)	0.0021 (.0569)[.0540]{.0547}	0.0020 (.0556)[.0542]{.0537}
	1.0	0.9924 (.0521)	0.9993 (.0526)[.0527]{.0510}	0.9993 (.0526)[.0527]{.0510}
	1.0	0.9562 (.2034)	0.9866 (.2164)[.1443]{.1995}	0.9865 (.2159)[.1443]{.1992}
1	**-0.4**	-0.4558 (.0433)	-0.3980 (.0469)[.0480]{.0467}	-0.3980 (.0470)[.0478]{.0466}
	1.0	1.0031 (.0505)	1.0001 (.0509)[.0525]{.0512}	1.0000 (.0509)[.0525]{.0512}
	1.0	0.9621 (.1364)	0.9775 (.1407)[.1406]{.1367}	0.9776 (.1408)[.1406]{.1367}
2	**-0.4**	-0.4541 (.0449)	-0.3973 (.0479)[.0473]{.0467}	-0.3975 (.0476)[.0471]{.0463}
	1.0	1.0010 (.0514)	0.9978 (.0518)[.0518]{.0502}	0.9978 (.0518)[.0518]{.0502}
	1.0	0.9502 (.2806)	0.9660 (.2899)[.1390]{.2503}	0.9660 (.2898)[.1390]{.2502}
3	**-0.4**	-0.4576 (.0457)	-0.4007 (.0492)[.0476]{.0468}	-0.4007 (.0491)[.0474]{.0466}
	1.0	1.0055 (.0495)	1.0024 (.0497)[.0523]{.0513}	1.0024 (.0497)[.0523]{.0513}
	1.0	0.9602 (.2012)	0.9757 (.2074)[.1403]{.1929}	0.9757 (.2075)[.1403]{.1929}
1	**-0.8**	-0.8238 (.0328)	-0.7993 (.0342)[.0353]{.0341}	-0.7993 (.0342)[.0351]{.0341}
	1.0	1.0072 (.0533)	1.0018 (.0534)[.0529]{.0519}	1.0018 (.0534)[.0529]{.0519}
	1.0	0.9679 (.1396)	0.9728 (.1410)[.1383]{.1356}	0.9728 (.1410)[.1383]{.1356}
2	**-0.8**	-0.8250 (.0361)	-0.8005 (.0372)[.0352]{.0342}	-0.8004 (.0372)[.0350]{.0341}
	1.0	1.0046 (.0544)	0.9991 (.0543)[.0529]{.0510}	0.9991 (.0543)[.0529]{.0510}
	1.0	0.9820 (.2868)	0.9872 (.2897)[.1404]{.2579}	0.9872 (.2897)[.1404]{.2579}
3	**-0.8**	-0.8238 (.0335)	-0.7992 (.0348)[.0354]{.0338}	-0.7992 (.0348)[.0352]{.0338}
	1.0	1.0060 (.0549)	1.0006 (.0549)[.0529]{.0515}	1.0006 (.0549)[.0529]{.0515}
	1.0	0.9717 (.2089)	0.9767 (.2110)[.1389]{.1925}	0.9768 (.2110)[.1389]{.1925}

Table 2. Empirical Mean(sd)[$\widehat{se}$]{$\widehat{rse}$} of CMLE, MMLE and FMLE, $N = 100, T = 3$

dgp	par	CMLE	MMLE	FMLE
1	**0.8**	0.7005 (.0320)	0.8015 (.0371)[.0352]{.0358}	0.8015 (.0370)[.0362]{.0357}
	1.0	0.9643 (.0346)	0.9999 (.0365)[.0359]{.0356}	0.9999 (.0365)[.0360]{.0356}
	1.0	0.9466 (.0937)	0.9940 (.1040)[.1047]{.1032}	0.9941 (.1041)[.1051]{.1032}
2	**0.8**	0.6980 (.0363)	0.7981 (.0390)[.0350]{.0385}	0.7979 (.0381)[.0360]{.0377}
	1.0	0.9629 (.0360)	0.9980 (.0373)[.0356]{.0353}	0.9979 (.0372)[.0358]{.0352}
	1.0	0.9434 (.1915)	0.9918 (.2121)[.1047]{.1954}	0.9917 (.2118)[.1050]{.1951}
3	**0.8**	0.7010 (.0339)	0.8013 (.0386)[.0350]{.0372}	0.8014 (.0383)[.0361]{.0369}
	1.0	0.9655 (.0352)	1.0010 (.0375)[.0357]{.0355}	1.0011 (.0374)[.0359]{.0355}
	1.0	0.9425 (.1407)	0.9901 (.1555)[.1044]{.1487}	0.9902 (.1555)[.1047]{.1488}
1	**0.4**	0.3184 (.0314)	0.4012 (.0352)[.0339]{.0339}	0.4012 (.0351)[.0343]{.0338}
	1.0	0.9892 (.0336)	1.0022 (.0344)[.0341]{.0336}	1.0022 (.0344)[.0341]{.0336}
	1.0	0.9610 (.0981)	0.9944 (.1053)[.1031]{.1021}	0.9944 (.1054)[.1032]{.1021}
2	**0.4**	0.3176 (.0336)	0.4003 (.0355)[.0338]{.0351}	0.4003 (.0346)[.0342]{.0342}
	1.0	0.9860 (.0340)	0.9988 (.0348)[.0339]{.0333}	0.9988 (.0348)[.0339]{.0333}
	1.0	0.9566 (.1930)	0.9907 (.2068)[.1028]{.1949}	0.9907 (.2067)[.1029]{.1948}
3	**0.4**	0.3166 (.0323)	0.3996 (.0347)[.0339]{.0346}	0.3995 (.0342)[.0343]{.0341}
	1.0	0.9874 (.0352)	1.0003 (.0361)[.0340]{.0337}	1.0003 (.0361)[.0340]{.0337}
	1.0	0.9613 (.1446)	0.9951 (.1549)[.1032]{.1494}	0.9951 (.1549)[.1033]{.1493}
1	**0.0**	-0.0706 (.0316)	-0.0015 (.0348)[.0335]{.0332}	-0.0014 (.0348)[.0336]{.0331}
	1.0	1.0000 (.0349)	0.9991 (.0354)[.0336]{.0332}	0.9991 (.0354)[.0336]{.0332}
	1.0	0.9652 (.0964)	0.9882 (.1012)[.1013]{.0999}	0.9883 (.1011)[.1013]{.0999}
2	**0.0**	-0.0686 (.0331)	0.0012 (.0347)[.0336]{.0343}	0.0011 (.0340)[.0337]{.0336}
	1.0	1.0039 (.0329)	1.0029 (.0331)[.0336]{.0331}	1.0030 (.0331)[.0336]{.0331}
	1.0	0.9734 (.2017)	0.9976 (.2115)[.1024]{.1948}	0.9975 (.2114)[.1024]{.1946}
3	**0.0**	-0.0687 (.0322)	0.0007 (.0344)[.0336]{.0341}	0.0006 (.0341)[.0337]{.0337}
	1.0	1.0020 (.0333)	1.0010 (.0336)[.0335]{.0329}	1.0010 (.0336)[.0335]{.0329}
	1.0	0.9669 (.1478)	0.9903 (.1551)[.1016]{.1455}	0.9903 (.1551)[.1016]{.1454}
1	**-0.4**	-0.4522 (.0294)	-0.3998 (.0315)[.0323]{.0318}	-0.3998 (.0314)[.0322]{.0317}
	1.0	1.0097 (.0345)	1.0011 (.0349)[.0340]{.0337}	1.0011 (.0349)[.0340]{.0337}
	1.0	0.9775 (.1015)	0.9915 (.1043)[.1006]{.0995}	0.9914 (.1043)[.1006]{.0994}
2	**-0.4**	-0.4520 (.0315)	-0.4005 (.0327)[.0319]{.0319}	-0.4005 (.0324)[.0318]{.0316}
	1.0	1.0089 (.0333)	1.0002 (.0334)[.0337]{.0330}	1.0002 (.0334)[.0336]{.0330}
	1.0	0.9620 (.1904)	0.9759 (.1955)[.0991]{.1898}	0.9759 (.1954)[.0990]{.1898}
3	**-0.4**	-0.4503 (.0312)	-0.3986 (.0331)[.0320]{.0319}	-0.3986 (.0330)[.0319]{.0318}
	1.0	1.0097 (.0324)	1.0011 (.0329)[.0338]{.0332}	1.0011 (.0328)[.0338]{.0331}
	1.0	0.9645 (.1413)	0.9782 (.1452)[.0993]{.1422}	0.9783 (.1452)[.0993]{.1422}
1	**-0.8**	-0.8231 (.0230)	-0.7999 (.0240)[.0244]{.0240}	-0.7999 (.0240)[.0242]{.0240}
	1.0	1.0082 (.0354)	1.0013 (.0355)[.0344]{.0341}	1.0013 (.0355)[.0344]{.0341}
	1.0	0.9878 (.1010)	0.9925 (.1019)[.0997]{.0981}	0.9925 (.1019)[.0997]{.0981}
2	**-0.8**	-0.8213 (.0239)	-0.7983 (.0249)[.0242]{.0238}	-0.7983 (.0248)[.0241]{.0237}
	1.0	1.0073 (.0349)	1.0005 (.0348)[.0342]{.0336}	1.0005 (.0348)[.0342]{.0336}
	1.0	0.9841 (.2094)	0.9888 (.2114)[.0994]{.1944}	0.9888 (.2114)[.0994]{.1944}
3	**-0.8**	-0.8221 (.0230)	-0.7991 (.0241)[.0243]{.0241}	-0.7991 (.0240)[.0241]{.0241}
	1.0	1.0076 (.0357)	1.0007 (.0357)[.0343]{.0338}	1.0007 (.0357)[.0343]{.0338}
	1.0	0.9829 (.1519)	0.9875 (.1533)[.0992]{.1432}	0.9875 (.1533)[.0992]{.1432}

Table 3. Empirical Mean(sd)[$\widehat{se}$]{$\widehat{rse}$} of CMLE, MMLE and FMLE, $N = 200, T = 3$

dgp	par	CMLE	MMLE	FMLE
1	**0.8**	0.6820 (.0232)	0.8008 (.0282)[.0274]{.0283}	0.8009 (.0282)[.0284]{.0282}
	1.0	0.9495 (.0268)	0.9990 (.0285)[.0288]{.0289}	0.9991 (.0285)[.0290]{.0289}
	1.0	0.9378 (.0664)	0.9933 (.0754)[.0747]{.0747}	0.9935 (.0754)[.0751]{.0748}
2	**0.8**	0.6817 (.0291)	0.8011 (.0322)[.0275]{.0311}	0.8009 (.0314)[.0284]{.0304}
	1.0	0.9510 (.0279)	1.0009 (.0297)[.0289]{.0292}	1.0008 (.0296)[.0291]{.0291}
	1.0	0.9446 (.1338)	1.0017 (.1503)[.0755]{.1446}	1.0015 (.1499)[.0758]{.1443}
3	**0.8**	0.6822 (.0257)	0.8013 (.0301)[.0274]{.0301}	0.8011 (.0297)[.0284]{.0298}
	1.0	0.9510 (.0280)	1.0009 (.0296)[.0289]{.0290}	1.0008 (.0295)[.0291]{.0290}
	1.0	0.9453 (.0971)	1.0017 (.1092)[.0754]{.1095}	1.0015 (.1090)[.0757]{.1094}
1	**0.4**	0.3073 (.0229)	0.3995 (.0259)[.0255]{.0259}	0.3995 (.0258)[.0258]{.0258}
	1.0	0.9810 (.0253)	0.9993 (.0261)[.0270]{.0268}	0.9993 (.0261)[.0270]{.0268}
	1.0	0.9592 (.0672)	0.9962 (.0728)[.0733]{.0731}	0.9962 (.0727)[.0734]{.0731}
2	**0.4**	0.3068 (.0262)	0.3985 (.0279)[.0254]{.0271}	0.3985 (.0271)[.0257]{.0264}
	1.0	0.9820 (.0263)	1.0002 (.0270)[.0269]{.0267}	1.0002 (.0270)[.0269]{.0267}
	1.0	0.9545 (.1342)	0.9916 (.1445)[.0730]{.1416}	0.9915 (.1444)[.0731]{.1415}
3	**0.4**	0.3081 (.0249)	0.4002 (.0267)[.0254]{.0264}	0.4000 (.0262)[.0258]{.0261}
	1.0	0.9818 (.0260)	1.0002 (.0266)[.0269]{.0267}	1.0002 (.0266)[.0269]{.0267}
	1.0	0.9541 (.0980)	0.9911 (.1057)[.0730]{.1064}	0.9909 (.1055)[.0730]{.1063}
1	**0.0**	-0.0738 (.0224)	-0.0010 (.0243)[.0244]{.0245}	-0.0010 (.0242)[.0245]{.0244}
	1.0	1.0005 (.0269)	1.0009 (.0273)[.0265]{.0263}	1.0009 (.0273)[.0265]{.0263}
	1.0	0.9704 (.0681)	0.9946 (.0716)[.0722]{.0714}	0.9947 (.0716)[.0722]{.0714}
2	**0.0**	-0.0716 (.0241)	0.0018 (.0256)[.0245]{.0254}	0.0017 (.0249)[.0246]{.0248}
	1.0	0.9995 (.0257)	1.0001 (.0259)[.0265]{.0262}	1.0001 (.0259)[.0265]{.0262}
	1.0	0.9768 (.1392)	1.0018 (.1461)[.0728]{.1436}	1.0017 (.1458)[.0728]{.1435}
3	**0.0**	-0.0746 (.0241)	-0.0013 (.0255)[.0245]{.0248}	-0.0014 (.0253)[.0246]{.0245}
	1.0	1.0000 (.0272)	1.0003 (.0275)[.0265]{.0264}	1.0003 (.0275)[.0265]{.0264}
	1.0	0.9759 (.1039)	1.0006 (.1090)[.0726]{.1066}	1.0005 (.1090)[.0726]{.1065}
1	**-0.4**	-0.4502 (.0206)	-0.3994 (.0219)[.0225]{.0222}	-0.3994 (.0219)[.0224]{.0222}
	1.0	1.0083 (.0260)	1.0007 (.0263)[.0267]{.0266}	1.0007 (.0263)[.0267]{.0266}
	1.0	0.9815 (.0694)	0.9951 (.0713)[.0714]{.0708}	0.9951 (.0713)[.0714]{.0708}
2	**-0.4**	-0.4513 (.0214)	-0.4007 (.0225)[.0224]{.0226}	-0.4006 (.0223)[.0224]{.0225}
	1.0	1.0087 (.0267)	1.0008 (.0270)[.0267]{.0265}	1.0008 (.0270)[.0267]{.0265}
	1.0	0.9819 (.1351)	0.9956 (.1387)[.0714]{.1417}	0.9956 (.1387)[.0714]{.1417}
3	**-0.4**	-0.4500 (.0217)	-0.3991 (.0230)[.0225]{.0224}	-0.3991 (.0230)[.0224]{.0223}
	1.0	1.0067 (.0274)	0.9990 (.0276)[.0267]{.0266}	0.9990 (.0276)[.0267]{.0266}
	1.0	0.9821 (.1053)	0.9958 (.1081)[.0714]{.1051}	0.9958 (.1081)[.0714]{.1051}
1	**-0.8**	-0.8222 (.0162)	-0.8007 (.0169)[.0166]{.0164}	-0.8007 (.0169)[.0165]{.0164}
	1.0	1.0062 (.0262)	1.0000 (.0264)[.0269]{.0268}	1.0000 (.0264)[.0269]{.0268}
	1.0	0.9908 (.0683)	0.9951 (.0689)[.0707]{.0702}	0.9951 (.0689)[.0707]{.0702}
2	**-0.8**	-0.8219 (.0165)	-0.8006 (.0169)[.0165]{.0165}	-0.8006 (.0169)[.0164]{.0165}
	1.0	1.0072 (.0263)	1.0010 (.0263)[.0268]{.0264}	1.0010 (.0263)[.0268]{.0264}
	1.0	0.9880 (.1381)	0.9923 (.1393)[.0705]{.1396}	0.9923 (.1393)[.0705]{.1396}
3	**-0.8**	-0.8207 (.0166)	-0.7994 (.0172)[.0165]{.0163}	-0.7994 (.0172)[.0164]{.0163}
	1.0	1.0074 (.0269)	1.0012 (.0269)[.0268]{.0266}	1.0012 (.0269)[.0268]{.0266}
	1.0	0.9815 (.1086)	0.9857 (.1095)[.0700]{.1040}	0.9857 (.1095)[.0700]{.1040}

Table 4. Empirical Mean(sd)[$\widehat{se}$]{$\widehat{rse}$} of CMLE, MMLE and FMLE, $N = 50, T = 6$

dgp	par	CMLE	MMLE	FMLE
1	**0.8**	0.7432 (.0233)	0.7992 (.0258)[.0256]{.0254}	0.7991 (.0257)[.0261]{.0255}
	1.0	0.9928 (.0324)	1.0011 (.0329)[.0324]{.0317}	1.0011 (.0329)[.0324]{.0317}
	1.0	0.9733 (.0925)	0.9950 (.0969)[.0912]{.0891}	0.9949 (.0969)[.0912]{.0891}
2	**0.8**	0.7431 (.0259)	0.7983 (.0271)[.0254]{.0253}	0.7984 (.0268)[.0258]{.0254}
	1.0	0.9919 (.0313)	1.0001 (.0317)[.0322]{.0313}	1.0001 (.0317)[.0322]{.0314}
	1.0	0.9653 (.1859)	0.9871 (.1941)[.0905]{.1869}	0.9871 (.1941)[.0906]{.1870}
3	**0.8**	0.7448 (.0244)	0.8008 (.0268)[.0256]{.0258}	0.8009 (.0268)[.0261]{.0259}
	1.0	0.9927 (.0323)	1.0012 (.0328)[.0324]{.0318}	1.0013 (.0328)[.0324]{.0318}
	1.0	0.9738 (.1421)	0.9959 (.1485)[.0913]{.1379}	0.9960 (.1486)[.0914]{.1380}
1	**0.4**	0.3567 (.0259)	0.4007 (.0273)[.0279]{.0274}	0.4007 (.0272)[.0280]{.0274}
	1.0	1.0035 (.0335)	1.0002 (.0336)[.0322]{.0317}	1.0002 (.0336)[.0322]{.0317}
	1.0	0.9839 (.0874)	0.9946 (.0893)[.0900]{.0876}	0.9946 (.0893)[.0900]{.0876}
2	**0.4**	0.3573 (.0272)	0.4015 (.0284)[.0279]{.0274}	0.4015 (.0282)[.0280]{.0274}
	1.0	1.0045 (.0338)	1.0012 (.0339)[.0323]{.0316}	1.0012 (.0339)[.0323]{.0316}
	1.0	0.9933 (.2014)	1.0045 (.2059)[.0909]{.1908}	1.0045 (.2059)[.0909]{.1908}
3	**0.4**	0.3551 (.0277)	0.3985 (.0287)[.0277]{.0270}	0.3984 (.0287)[.0278]{.0270}
	1.0	1.0045 (.0315)	1.0012 (.0316)[.0321]{.0313}	1.0012 (.0316)[.0321]{.0313}
	1.0	0.9754 (.1433)	0.9859 (.1463)[.0892]{.1356}	0.9859 (.1463)[.0892]{.1356}
1	**0.0**	-0.0358 (.0290)	-0.0012 (.0301)[.0294]{.0290}	-0.0012 (.0301)[.0294]{.0290}
	1.0	1.0068 (.0325)	1.0000 (.0326)[.0326]{.0320}	1.0000 (.0326)[.0326]{.0320}
	1.0	0.9827 (.0862)	0.9885 (.0872)[.0889]{.0869}	0.9885 (.0872)[.0889]{.0869}
2	**0.0**	-0.0350 (.0294)	-0.0001 (.0302)[.0294]{.0285}	-0.0002 (.0301)[.0294]{.0284}
	1.0	1.0088 (.0336)	1.0020 (.0338)[.0326]{.0318}	1.0020 (.0338)[.0326]{.0318}
	1.0	0.9940 (.1984)	1.0000 (.2006)[.0900]{.1891}	1.0000 (.2006)[.0900]{.1891}
3	**0.0**	-0.0362 (.0288)	-0.0018 (.0298)[.0293]{.0285}	-0.0018 (.0298)[.0293]{.0285}
	1.0	1.0075 (.0331)	1.0008 (.0331)[.0325]{.0317}	1.0008 (.0331)[.0325]{.0317}
	1.0	0.9800 (.1409)	0.9858 (.1425)[.0887]{.1358}	0.9858 (.1425)[.0887]{.1358}
1	**-0.4**	-0.4247 (.0281)	-0.3994 (.0292)[.0286]{.0281}	-0.3994 (.0292)[.0286]{.0281}
	1.0	1.0057 (.0339)	0.9991 (.0340)[.0330]{.0322}	0.9991 (.0340)[.0330]{.0322}
	1.0	0.9889 (.0837)	0.9921 (.0842)[.0890]{.0868}	0.9921 (.0842)[.0890]{.0868}
2	**-0.4**	-0.4248 (.0282)	-0.3994 (.0290)[.0286]{.0277}	-0.3994 (.0290)[.0286]{.0277}
	1.0	1.0063 (.0330)	0.9996 (.0332)[.0330]{.0321}	0.9996 (.0332)[.0330]{.0321}
	1.0	0.9962 (.2051)	0.9995 (.2063)[.0897]{.1877}	0.9995 (.2063)[.0897]{.1877}
3	**-0.4**	-0.4263 (.0282)	-0.4009 (.0293)[.0286]{.0278}	-0.4009 (.0294)[.0286]{.0278}
	1.0	1.0087 (.0329)	1.0021 (.0331)[.0331]{.0325}	1.0021 (.0331)[.0331]{.0325}
	1.0	0.9973 (.1465)	1.0005 (.1473)[.0898]{.1370}	1.0005 (.1474)[.0898]{.1370}
1	**-0.8**	-0.8111 (.0214)	-0.7993 (.0219)[.0214]{.0209}	-0.7993 (.0219)[.0214]{.0209}
	1.0	1.0028 (.0331)	0.9995 (.0331)[.0326]{.0320}	0.9995 (.0331)[.0326]{.0320}
	1.0	0.9875 (.0877)	0.9888 (.0879)[.0885]{.0868}	0.9888 (.0879)[.0885]{.0868}
2	**-0.8**	-0.8111 (.0212)	-0.7992 (.0218)[.0214]{.0206}	-0.7992 (.0218)[.0214]{.0206}
	1.0	1.0014 (.0338)	0.9981 (.0339)[.0326]{.0318}	0.9981 (.0339)[.0326]{.0318}
	1.0	0.9951 (.2060)	0.9964 (.2065)[.0892]{.1874}	0.9964 (.2065)[.0892]{.1874}
3	**-0.8**	-0.8116 (.0204)	-0.8000 (.0211)[.0212]{.0206}	-0.8000 (.0211)[.0212]{.0206}
	1.0	1.0035 (.0332)	1.0002 (.0332)[.0325]{.0319}	1.0002 (.0332)[.0325]{.0319}
	1.0	0.9801 (.1467)	0.9813 (.1471)[.0879]{.1341}	0.9813 (.1471)[.0879]{.1341}

Table 5. Empirical Mean(sd)[$\widehat{se}$]{$\widehat{rse}$} of CMLE, MMLE and FMLE, $N = 100, T = 6$

dgp	par	CMLE	MMLE	FMLE
1	**0.8**	0.7466 (.0154)	0.8007 (.0170)[.0178]{.0178}	0.8007 (.0170)[.0181]{.0178}
	1.0	0.9820 (.0216)	1.0000 (.0217)[.0218]{.0215}	1.0000 (.0217)[.0218]{.0215}
	1.0	0.9757 (.0623)	0.9966 (.0651)[.0645]{.0638}	0.9966 (.0651)[.0646]{.0638}
2	**0.8**	0.7462 (.0180)	0.8004 (.0187)[.0177]{.0184}	0.8004 (.0186)[.0181]{.0183}
	1.0	0.9827 (.0218)	1.0008 (.0221)[.0218]{.0215}	1.0008 (.0220)[.0218]{.0215}
	1.0	0.9752 (.1387)	0.9964 (.1446)[.0645]{.1385}	0.9964 (.1446)[.0646]{.1385}
3	**0.8**	0.7458 (.0170)	0.8000 (.0182)[.0178]{.0183}	0.8000 (.0182)[.0181]{.0183}
	1.0	0.9819 (.0217)	0.9998 (.0219)[.0218]{.0217}	0.9998 (.0219)[.0218]{.0217}
	1.0	0.9779 (.0976)	0.9989 (.1018)[.0647]{.1015}	0.9989 (.1017)[.0647]{.1015}
1	**0.4**	0.3609 (.0179)	0.3995 (.0187)[.0185]{.0183}	0.3996 (.0187)[.0185]{.0183}
	1.0	1.0025 (.0207)	1.0007 (.0207)[.0210]{.0208}	1.0007 (.0207)[.0210]{.0208}
	1.0	0.9863 (.0595)	0.9956 (.0606)[.0636]{.0629}	0.9956 (.0606)[.0636]{.0629}
2	**0.4**	0.3598 (.0183)	0.3984 (.0184)[.0184]{.0183}	0.3984 (.0184)[.0185]{.0183}
	1.0	1.0016 (.0204)	0.9998 (.0204)[.0209]{.0207}	0.9998 (.0204)[.0209]{.0207}
	1.0	0.9836 (.1406)	0.9930 (.1432)[.0634]{.1373}	0.9930 (.1432)[.0634]{.1373}
3	**0.4**	0.3610 (.0180)	0.3999 (.0183)[.0185]{.0183}	0.3999 (.0183)[.0186]{.0183}
	1.0	1.0011 (.0212)	0.9993 (.0212)[.0210]{.0208}	0.9993 (.0212)[.0210]{.0208}
	1.0	0.9915 (.1009)	1.0010 (.1027)[.0639]{.1003}	1.0010 (.1027)[.0639]{.1003}
1	**0.0**	-0.0313 (.0197)	0.0007 (.0203)[.0200]{.0199}	0.0007 (.0203)[.0200]{.0198}
	1.0	1.0075 (.0224)	1.0010 (.0225)[.0214]{.0212}	1.0010 (.0225)[.0214]{.0212}
	1.0	0.9943 (.0646)	0.9997 (.0653)[.0636]{.0630}	0.9997 (.0653)[.0636]{.0630}
2	**0.0**	-0.0324 (.0193)	-0.0003 (.0197)[.0199]{.0196}	-0.0004 (.0197)[.0199]{.0196}
	1.0	1.0070 (.0215)	1.0005 (.0216)[.0214]{.0211}	1.0005 (.0216)[.0214]{.0211}
	1.0	0.9971 (.1378)	1.0026 (.1392)[.0638]{.1386}	1.0026 (.1392)[.0638]{.1386}
3	**0.0**	-0.0319 (.0197)	0.0001 (.0202)[.0200]{.0197}	0.0001 (.0202)[.0200]{.0197}
	1.0	1.0066 (.0206)	1.0001 (.0207)[.0214]{.0212}	1.0001 (.0207)[.0214]{.0212}
	1.0	0.9952 (.1043)	1.0006 (.1054)[.0636]{.1004}	1.0006 (.1054)[.0636]{.1004}
1	**-0.4**	-0.4240 (.0192)	-0.3992 (.0199)[.0200]{.0199}	-0.3992 (.0198)[.0200]{.0199}
	1.0	1.0055 (.0215)	0.9990 (.0216)[.0216]{.0215}	0.9990 (.0216)[.0216]{.0215}
	1.0	0.9958 (.0589)	0.9989 (.0593)[.0634]{.0625}	0.9989 (.0593)[.0634]{.0625}
2	**-0.4**	-0.4247 (.0195)	-0.4001 (.0201)[.0199]{.0196}	-0.4001 (.0201)[.0199]{.0196}
	1.0	1.0067 (.0217)	1.0003 (.0218)[.0216]{.0213}	1.0003 (.0218)[.0216]{.0213}
	1.0	0.9925 (.1418)	0.9956 (.1426)[.0632]{.1368}	0.9956 (.1426)[.0632]{.1368}
3	**-0.4**	-0.4252 (.0197)	-0.4005 (.0203)[.0200]{.0196}	-0.4005 (.0203)[.0200]{.0196}
	1.0	1.0064 (.0211)	0.9999 (.0212)[.0216]{.0213}	0.9999 (.0212)[.0216]{.0213}
	1.0	0.9923 (.1086)	0.9954 (.1092)[.0632]{.1006}	0.9954 (.1092)[.0632]{.1006}
1	**-0.8**	-0.8118 (.0156)	-0.7997 (.0162)[.0154]{.0152}	-0.7997 (.0162)[.0154]{.0152}
	1.0	1.0022 (.0212)	0.9986 (.0212)[.0215]{.0213}	0.9986 (.0212)[.0215]{.0213}
	1.0	0.9976 (.0660)	0.9989 (.0661)[.0633]{.0628}	0.9989 (.0661)[.0633]{.0628}
2	**-0.8**	-0.8116 (.0153)	-0.7996 (.0157)[.0153]{.0150}	-0.7996 (.0157)[.0153]{.0150}
	1.0	1.0035 (.0221)	0.9998 (.0222)[.0214]{.0212}	0.9998 (.0222)[.0214]{.0212}
	1.0	0.9914 (.1423)	0.9927 (.1427)[.0629]{.1353}	0.9927 (.1427)[.0629]{.1353}
3	**-0.8**	-0.8119 (.0153)	-0.7999 (.0158)[.0153]{.0150}	-0.7999 (.0158)[.0153]{.0150}
	1.0	1.0038 (.0220)	1.0002 (.0220)[.0214]{.0213}	1.0002 (.0220)[.0214]{.0213}
	1.0	0.9942 (.1051)	0.9955 (.1054)[.0630]{.0993}	0.9955 (.1054)[.0630]{.0993}

Table 6. Empirical Mean(sd)[$\widehat{se}$]{$\widehat{rse}$} of CMLE, MMLE and FMLE, $N = 200, T = 6$

dgp	par	CMLE	MMLE	FMLE
1	**0.8**	0.7523 (.0111)	0.8001 (.0121)[.0117]{.0119}	0.8000 (.0121)[.0119]{.0119}
	1.0	0.9870 (.0163)	1.0006 (.0164)[.0157]{.0156}	1.0006 (.0164)[.0157]{.0156}
	1.0	0.9783 (.0432)	0.9967 (.0449)[.0455]{.0454}	0.9967 (.0449)[.0455]{.0454}
2	**0.8**	0.7523 (.0120)	0.7997 (.0122)[.0117]{.0121}	0.7997 (.0122)[.0119]{.0120}
	1.0	0.9866 (.0150)	1.0000 (.0151)[.0156]{.0155}	1.0000 (.0151)[.0157]{.0155}
	1.0	0.9749 (.0964)	0.9933 (.1000)[.0453]{.0984}	0.9933 (.1000)[.0454]{.0985}
3	**0.8**	0.7522 (.0115)	0.8001 (.0122)[.0117]{.0121}	0.8001 (.0122)[.0119]{.0120}
	1.0	0.9865 (.0158)	1.0001 (.0159)[.0157]{.0157}	1.0001 (.0159)[.0157]{.0157}
	1.0	0.9800 (.0727)	0.9985 (.0755)[.0456]{.0727}	0.9985 (.0755)[.0456]{.0727}
1	**0.4**	0.3617 (.0123)	0.4000 (.0128)[.0130]{.0129}	0.4000 (.0128)[.0130]{.0129}
	1.0	1.0020 (.0148)	1.0013 (.0148)[.0154]{.0153}	1.0013 (.0148)[.0154]{.0153}
	1.0	0.9899 (.0455)	0.9992 (.0463)[.0451]{.0449}	0.9992 (.0463)[.0451]{.0449}
2	**0.4**	0.3614 (.0128)	0.3998 (.0130)[.0130]{.0130}	0.3998 (.0130)[.0130]{.0129}
	1.0	1.0014 (.0154)	1.0007 (.0154)[.0153]{.0152}	1.0007 (.0154)[.0153]{.0152}
	1.0	0.9904 (.0987)	0.9998 (.1005)[.0452]{.0990}	0.9998 (.1005)[.0452]{.0990}
3	**0.4**	0.3620 (.0131)	0.4002 (.0136)[.0130]{.0130}	0.4002 (.0136)[.0130]{.0130}
	1.0	1.0008 (.0154)	1.0001 (.0154)[.0153]{.0153}	1.0001 (.0154)[.0153]{.0153}
	1.0	0.9863 (.0714)	0.9955 (.0726)[.0450]{.0722}	0.9955 (.0726)[.0450]{.0722}
1	**0.0**	-0.0326 (.0141)	0.0002 (.0146)[.0143]{.0142}	0.0002 (.0146)[.0143]{.0142}
	1.0	1.0053 (.0155)	0.9994 (.0155)[.0155]{.0154}	0.9994 (.0155)[.0155]{.0154}
	1.0	0.9917 (.0452)	0.9972 (.0456)[.0448]{.0446}	0.9972 (.0456)[.0448]{.0446}
2	**0.0**	-0.0325 (.0144)	0.0003 (.0147)[.0143]{.0142}	0.0003 (.0147)[.0143]{.0142}
	1.0	1.0057 (.0150)	0.9999 (.0151)[.0155]{.0154}	0.9999 (.0151)[.0155]{.0154}
	1.0	0.9902 (.1003)	0.9957 (.1014)[.0448]{.0976}	0.9957 (.1014)[.0448]{.0976}
3	**0.0**	-0.0322 (.0140)	0.0007 (.0144)[.0143]{.0143}	0.0007 (.0143)[.0143]{.0143}
	1.0	1.0058 (.0157)	0.9999 (.0158)[.0156]{.0155}	0.9999 (.0158)[.0156]{.0155}
	1.0	0.9971 (.0748)	1.0026 (.0755)[.0451]{.0728}	1.0026 (.0755)[.0451]{.0728}
1	**-0.4**	-0.4272 (.0144)	-0.4008 (.0149)[.0147]{.0145}	-0.4008 (.0149)[.0147]{.0145}
	1.0	1.0073 (.0160)	1.0000 (.0161)[.0158]{.0158}	1.0000 (.0161)[.0158]{.0158}
	1.0	0.9920 (.0450)	0.9954 (.0452)[.0447]{.0443}	0.9953 (.0452)[.0447]{.0443}
2	**-0.4**	-0.4266 (.0147)	-0.4002 (.0151)[.0146]{.0145}	-0.4002 (.0151)[.0146]{.0145}
	1.0	1.0062 (.0158)	0.9989 (.0158)[.0158]{.0157}	0.9989 (.0158)[.0158]{.0157}
	1.0	0.9923 (.1009)	0.9957 (.1015)[.0447]{.0983}	0.9957 (.1015)[.0447]{.0983}
3	**-0.4**	-0.4254 (.0150)	-0.3988 (.0155)[.0147]{.0146}	-0.3988 (.0155)[.0147]{.0146}
	1.0	1.0069 (.0159)	0.9996 (.0159)[.0159]{.0158}	0.9996 (.0159)[.0159]{.0158}
	1.0	0.9981 (.0729)	1.0015 (.0733)[.0449]{.0723}	1.0015 (.0733)[.0449]{.0723}
1	**-0.8**	-0.8130 (.0102)	-0.7998 (.0106)[.0113]{.0113}	-0.7998 (.0106)[.0114]{.0113}
	1.0	1.0033 (.0161)	0.9989 (.0161)[.0158]{.0157}	0.9989 (.0161)[.0158]{.0157}
	1.0	0.9962 (.0441)	0.9975 (.0442)[.0447]{.0445}	0.9975 (.0442)[.0447]{.0445}
2	**-0.8**	-0.8132 (.0109)	-0.8000 (.0112)[.0113]{.0112}	-0.8000 (.0112)[.0113]{.0112}
	1.0	1.0048 (.0153)	1.0005 (.0153)[.0158]{.0156}	1.0005 (.0153)[.0158]{.0156}
	1.0	0.9947 (.1012)	0.9960 (.1014)[.0446]{.0985}	0.9960 (.1014)[.0446]{.0985}
3	**-0.8**	-0.8129 (.0113)	-0.7996 (.0117)[.0114]{.0113}	-0.7996 (.0117)[.0114]{.0113}
	1.0	1.0040 (.0161)	0.9996 (.0161)[.0158]{.0157}	0.9996 (.0161)[.0158]{.0157}
	1.0	0.9959 (.0735)	0.9973 (.0736)[.0447]{.0717}	0.9973 (.0736)[.0447]{.0717}

Table 7. Empirical Mean(sd). Panels top-down-left-right: $\rho = .1, 0, -.1, \ldots$; $(N, T) = (10, 6)$ for first 3 panels, $(20, 16)$ next 3, and then $(30, 26)$; Every three rows, $(\rho, \beta, \sigma_{v0}^2)' = (\rho, 1, 1)'$

dgp	CMLE	MMLE	FMLE	CMLE	MMLE	FMLE
1	.0653(.059)	.0941(.060)	.0655(.092)	-.1124(.027)	-.1002(.027)	-.0978(.086)
	1.0052(.066)	1.0017(.066)	1.0052(.066)	1.0052(.030)	1.0015(.030)	1.0007(.039)
	.9484(.188)	.9536(.190)	.7994(.376)	.9946(.081)	.9953(.081)	.9771(.142)
2	.0713(.057)	.0993(.057)	.0602(.099)	-.1116(.028)	-.0995(.028)	-.1010(.073)
	1.0034(.066)	.9999(.066)	1.0049(.066)	1.0032(.030)	.9995(.030)	.9998(.035)
	.9382(.452)	.9440(.458)	.7212(.516)	.9897(.191)	.9904(.191)	.9627(.235)
3	.0656(.056)	.0946(.057)	.0589(.097)	-.1130(.027)	-.1008(.028)	-.1012(.063)
	1.0029(.069)	.9994(.069)	1.0038(.069)	1.0038(.029)	1.0001(.029)	1.0002(.033)
	.9592(.308)	.9647(.311)	.7683(.449)	.9889(.139)	.9896(.139)	.9730(.178)
1	-.0310(.062)	-.0020(.063)	-.0230(.084)	.0925(.015)	.0995(.015)	.0278(.089)
	1.0058(.068)	1.0012(.069)	1.0044(.069)	1.0026(.017)	1.0004(.017)	1.0226(.032)
	.9668(.204)	.9717(.206)	.8520(.364)	1.0010(.052)	1.0013(.052)	.6027(.490)
2	-.0309(.060)	-.0034(.060)	-.0402(.098)	.0930(.014)	.1000(.014)	.0181(.092)
	1.0057(.067)	1.0014(.068)	1.0067(.067)	1.0019(.016)	.9997(.016)	1.0251(.032)
	.9320(.419)	.9371(.424)	.7320(.506)	.9955(.120)	.9958(.120)	.5428(.499)
3	-.0309(.059)	-.0026(.060)	-.0302(.092)	.0931(.015)	.1001(.015)	.0258(.090)
	1.0097(.068)	1.0054(.068)	1.0094(.068)	1.0021(.016)	.9999(.016)	1.0231(.033)
	.9532(.301)	.9581(.304)	.7999(.434)	.9973(.088)	.9976(.088)	.5846(.494)
1	-.1294(.061)	-.1013(.063)	-.1168(.084)	-.0074(.015)	-.0007(.015)	-.0638(.084)
	1.0054(.070)	1.0002(.071)	1.0031(.071)	1.0024(.017)	1.0002(.017)	1.0209(.032)
	.9519(.197)	.9562(.199)	.8622(.335)	.9965(.051)	.9967(.051)	.6386(.479)
2	-.1297(.059)	-.1029(.060)	-.1313(.104)	-.0064(.016)	.0003(.016)	-.0742(.090)
	1.0081(.067)	1.0032(.067)	1.0086(.070)	1.0023(.017)	1.0001(.017)	1.0246(.034)
	.9297(.417)	.9342(.421)	.7691(.479)	.9950(.119)	.9952(.119)	.5697(.494)
3	-.1294(.063)	-.1015(.064)	-.1214(.086)	-.0075(.015)	-.0008(.015)	-.0667(.086)
	1.0080(.069)	1.0030(.069)	1.0064(.069)	1.0019(.017)	.9996(.017)	1.0212(.032)
	.9566(.300)	.9611(.303)	.8513(.399)	.9948(.088)	.9951(.088)	.6182(.484)
1	.0872(.025)	.1003(.025)	.0992(.061)	-.1069(.016)	-.1004(.016)	-.1471(.078)
	1.0077(.029)	1.0039(.029)	1.0041(.033)	1.0026(.017)	1.0003(.017)	1.0163(.031)
	.9927(.082)	.9936(.082)	.9694(.150)	.9977(.051)	.9980(.051)	.7236(.446)
2	.0864(.025)	.0994(.026)	.0940(.061)	-.1071(.016)	-.1005(.016)	-.1557(.083)
	1.0043(.028)	1.0007(.028)	1.0019(.032)	1.0021(.017)	.9999(.017)	1.0188(.032)
	.9873(.193)	.9882(.193)	.9494(.246)	.9990(.121)	.9993(.121)	.6777(.472)
3	.0877(.027)	.1007(.027)	.1013(.086)	-.1068(.016)	-.1002(.016)	-.1556(.081)
	1.0046(.029)	1.0010(.029)	1.0008(.037)	1.0026(.018)	1.0003(.018)	1.0192(.033)
	.9892(.139)	.9901(.139)	.9561(.204)	1.0007(.088)	1.0009(.088)	.6734(.471)
1	-.0135(.026)	-.0008(.026)	-.0040(.061)			
	1.0047(.029)	1.0010(.029)	1.0018(.033)			
	.9878(.082)	.9886(.082)	.9601(.161)			
2	-.0138(.026)	-.0012(.026)	-.0048(.078)			
	1.0049(.029)	1.0012(.029)	1.0022(.036)			
	.9961(.192)	.9969(.193)	.9536(.253)			
3	-.0125(.026)	.0002(.026)	-.0024(.063)			
	1.0025(.029)	.9988(.029)	.9996(.034)			
	.9843(.134)	.9851(.134)	.9562(.194)			

Table 8. Empirical Mean(sd). $\rho = .9, .6, .3$; $\sigma^2_{v0} = 1$, every second row

dpg	par	CMLE	MMLE	FMLE	CMLE	MMLE	FMLE
			$N = 20, T = 6$			$N = 20, T = 12$	
1	.9	.4940(.097)	.8373(.142)	.8012(.174)	.6884(.056)	.8791(.078)	.8332(.113)
	1	.8153(.119)	.9507(.162)	.8436(.265)	.9213(.092)	.9889(.109)	.9168(.217)
2	.9	.4946(.108)	.8240(.166)	.7885(.180)	.6895(.060)	.8772(.084)	.8281(.161)
	1	.8177(.271)	.9441(.326)	.8481(.360)	.9157(.197)	.9813(.216)	.9098(.283)
3	.9	.4961(.101)	.8383(.152)	.7892(.166)	.6880(.058)	.8795(.080)	.8283(.173)
	1	.8127(.189)	.9460(.234)	.8585(.278)	.9238(.150)	.9920(.167)	.9092(.265)
1	.6	.2910(.096)	.6074(.159)	.5650(.177)	.4511(.063)	.6004(.079)	.5798(.076)
	1	.9055(.132)	1.0119(.172)	.9400(.232)	.9692(.091)	.9974(.097)	.9885(.105)
2	.6	.2881(.098)	.5944(.158)	.5647(.186)	.4463(.064)	.5966(.082)	.5745(.075)
	1	.8944(.300)	.9936(.349)	.9297(.370)	.9833(.221)	1.0123(.230)	1.0046(.227)
3	.6	.2894(.102)	.6033(.165)	.5602(.179)	.4486(.062)	.5982(.081)	.5741(.094)
	1	.9110(.205)	1.0168(.248)	.9531(.278)	.9659(.160)	.9943(.168)	.9856(.167)
1	.3	.0708(.095)	.3036(.135)	.2864(.131)	.1810(.064)	.2922(.072)	.2842(.072)
	1	.9355(.134)	.9893(.150)	.9792(.152)	.9813(.094)	.9940(.096)	.9924(.096)
2	.3	.0699(.100)	.3121(.149)	.2881(.142)	.1851(.065)	.2977(.074)	.2871(.072)
	1	.9290(.293)	.9873(.320)	.9677(.321)	.9819(.220)	.9950(.224)	.9928(.223)
3	.3	.0750(.103)	.3132(.148)	.2908(.137)	.1847(.065)	.2971(.073)	.2877(.073)
	1	.9404(.214)	.9975(.235)	.9862(.231)	.9792(.151)	.9922(.154)	.9903(.153)
			$N = 50, T = 6$			$N = 50, T = 12$	
1	.9	.4998(.061)	.8631(.106)	.8284(.126)	.6932(.035)	.8948(.060)	.8616(.080)
	1	.8235(.076)	.9739(.113)	.9033(.203)	.9242(.059)	.9993(.075)	.9456(.177)
2	.9	.5045(.072)	.8566(.128)	.8276(.138)	.6949(.038)	.8919(.062)	.8624(.081)
	1	.8169(.165)	.9588(.211)	.9003(.253)	.9241(.134)	.9965(.151)	.9454(.216)
3	.9	.5048(.065)	.8640(.115)	.8318(.130)	.6943(.037)	.8944(.060)	.8602(.082)
	1	.8260(.128)	.9746(.168)	.9054(.234)	.9320(.094)	1.0066(.107)	.9611(.174)
1	.6	.2942(.062)	.6077(.114)	.5877(.127)	.4508(.040)	.5985(.049)	.5877(.047)
	1	.9015(.084)	1.0031(.117)	.9612(.173)	.9746(.060)	1.0016(.063)	.9979(.062)
2	.6	.2945(.066)	.6097(.119)	.5859(.116)	.4519(.040)	.6002(.049)	.5898(.047)
	1	.9057(.186)	1.0096(.225)	.9766(.236)	.9679(.132)	.9951(.137)	.9915(.136)
3	.6	.2956(.064)	.6108(.117)	.5876(.124)	.4497(.040)	.5969(.049)	.5868(.047)
	1	.9080(.140)	1.0121(.175)	.9713(.207)	.9724(.098)	.9993(.101)	.9958(.101)
			$N = 100, T = 6$			$N = 100, T = 12$	
1	.9	.5046(.041)	.8845(.084)	.8482(.090)	.6959(.025)	.9020(.047)	.8761(.054)
	1	.8281(.053)	.9915(.089)	.9478(.134)	.9249(.041)	1.0029(.055)	.9612(.150)
2	.9	.5045(.053)	.8706(.105)	.8422(.107)	.6940(.027)	.8956(.050)	.8734(.057)
	1	.8230(.113)	.9756(.150)	.9302(.189)	.9240(.094)	.9986(.107)	.9568(.179)
3	.9	.5020(.047)	.8743(.094)	.8431(.101)	.6971(.026)	.9035(.048)	.8771(.053)
	1	.8258(.089)	.9827(.126)	.9369(.169)	.9242(.065)	1.0026(.078)	.9650(.147)
1	.6	.2978(.043)	.6095(.081)	.5936(.077)	.4532(.028)	.6007(.034)	.5951(.033)
	1	.9096(.062)	1.0088(.085)	.9946(.094)	.9760(.042)	1.0028(.044)	1.0009(.044)
2	.6	.2988(.047)	.6141(.096)	.5971(.089)	.4532(.028)	.6011(.034)	.5955(.032)
	1	.9101(.128)	1.0129(.157)	.9931(.166)	.9746(.097)	1.0016(.101)	.9996(.100)
3	.6	.2927(.044)	.6013(.085)	.5838(.073)	.4517(.028)	.5989(.034)	.5936(.033)
	1	.9044(.093)	1.0015(.117)	.9874(.118)	.9735(.068)	1.0001(.071)	.9983(.071)