A Modified First-Difference Estimator in a Panel Data Model with Unobservable Factors both in the Errors and the Regressors when the Time Dimension is Small

Giovanni Forchini[1]

University of Surrey

Bin Peng[2]

University of Technology Sydney

20 January 2017

Abstract

Panel data models with factor structures in both the errors and the regressors have received considerable attention recently. In these models the errors and the regressors are correlated and the standard estimators are inconsistent. This paper shows that, for such models, a modified first-difference estimators (in which the time and the cross-sectional dimensions are interchanged) is consistent as the cross-sectional dimension grows but the time dimension is small. Although the estimator has a non-standard asymptotic distribution, *t* and *F* tests have standard asymptotic distribution under the null hypothesis.

---

[1] Address for correspondence: Giovanni Forchini, Department of Economics, Umeå University, Sweden. E-mail: Giovanni.Forchini@umu.se. This research was partially supported by Australian Research Council grant DP0985432.

[2] Address for correspondence: Bin Peng, Economics Discipline Group, University of Technology Sydney, Ultimo NSW, Australia, 2007. E-mail: Bin.Peng@uts.edu.au.

Introduction

A considerable literature on the effects of common shocks on linear models has been developed in the last decade (e.g. Andrews (2005) and Pesaran (2006) among several others). It is standard in the literature to model the common shocks using factor structures in both the error and the regressors. This induces cross-sectional dependence as well as correlation between the errors and the regressors (i.e. endogeneity).

Models with these properties often occur in finance. For example, in models of executive compensations one assumes that the log of the compensation (defined either as cash compensation or total compensation) of executive i at time t is linearly related to assets, returns on assets, stock returns, dummies indicating the level of responsibility and gender. Since companies within the same industry face similar demand and supply conditions in each period, industry shocks in each period may affect the dependent variable directly through the error term and indirectly through the explanatory variables. In another important financial example, the debt/equity choice (leverage) is explained by firm size, returns, asset structure, the risk (and possibly other variables). Once again, industry specific shocks may affect both the error terms and some of the explanatory variables.

Most of the current literature focuses on the case where both the time series and the cross-section dimensions are large. The case where the time dimension is fixed has only recently received attention by Andrews (2005), Kuersteiner and Prucha (2013), and Forchini and Peng (2015). A characteristic of these contributions is the imposition of assumptions conditional on the unobserved factors, and the use of the notion of stable convergence in the sense of Rényi (1963) to obtain the asymptotic distributions of the statistics studied.

Forchini and Peng (2015) consider a linear panel data model in which the shocks are captured by common factors in the errors and some of the regressors. They study consistency and asymptotic distribution of the ordinary least squares (OLS) estimator of the slope parameters when the cross sectional dimension is fixed under assumptions similar to those of Andrews (2005) and Kuersteiner and Prucha (2013). Forchini and Peng (2015) show that the OLS estimator has a non-standard asymptotic distribution, however t and F tests have standard asymptotic distributions under the null hypothesis. Since, in the model considered, the OLS estimator of the slope parameters can be interpreted the standard fixed effects estimator with the time-series and cross-section dimensions interchanged, it is natural to investigate whether the first difference estimator with time-series and cross-section dimension interchanged - referred to as the modified first difference (MFD) estimator - is also consistent and whether it has a non-standard asymptotic distribution. This is what we do in this paper. The analysis of the MFD estimator is more challenging than that for the OLS estimator studied by Forchini and Peng (2015) because it involves the use of heterogeneous m-dependent processes (which are not required for the later).

A panel data model

A typical linear panel data model contains both time-invariant z_i and time-variant x_i explanatory variables and can be written in vector form as

(1) $$\underset{(T\times 1)}{y_i} = \underset{(T\times 1)}{\tau_0} + \underset{(T\times k)}{z_i}\underset{(k\times 1)}{\alpha_0} + \underset{(T\times p)}{x_i}\underset{(p\times 1)}{\beta_0} + e_i = \tau_0 + \underset{T\times(k+p)}{\omega_i}\begin{pmatrix}\alpha_0 \\ \beta_0\end{pmatrix} + e_i,$$

where $\omega_i = (z_i, x_i)$, $i = 1,2,...,N$. In the above equation α_0 and β_0 denote unknown parameters and e_i is a random vector. The term τ_0 can be considered as an unknown parameter or an unobserved effect possibly correlated with the regressors. The dimension of the various vectors and matrices is reported in brackets underneath each quantity the first time it is used. It is assumed, following Andrews (2005) and Pesaran (2006), that the errors and the time variant regressors may be affected by common shocks which can be captured using factors structures:

(2) $$\underset{(T\times 1)}{e_i} = \underset{(T\times m)}{F_T}\,\underset{(m\times 1)}{\gamma_i} + \varepsilon_i$$

and

(3) $$x_i = F_T \underset{(m\times p)}{\Gamma_i} + v_i \,.$$

In equations (2) and (3), $F_T = \left(f_1, f_2, ..., f_T\right)'$ denotes a matrix of unobservable common factors, γ_i is a vector of unobservable factor loadings and ε_i is a purely idiosyncratic unobservable random vector with zero mean and constant covariance matrix. The vector v_i is interpreted as the value of the regressors that would be observed in the absence of common shocks. Finally, Γ_i is an unobservable factor loading matrix. The presence of factors in both the errors and the x_i's makes these regressors endogenous. Notice that our notation is a bit redundant to emphasise the distinction between the fixed term τ_0 and random term $F_T\gamma_i$.

The estimator we consider is a first difference estimator of the parameters α_0 and β_0 in which the cross-section and the time dimensions have been interchanged:

(4) $$\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} = \left[\sum_{i=2}^{N}\left(\omega_i - \omega_{i-1}\right)'\left(\omega_i - \omega_{i-1}\right)\right]^{-1} \sum_{i=2}^{N}\left(\omega_i - \omega_{i-1}\right)'\left(y_i - y_{i-1}\right).$$

We will refer to this as the modified first difference (MFD) estimator. Given the conditional independence assumption C.5 between the factor loadings formulated below on has the moment conditions $E\left[x_i'e_i \mid F_T\right] = E\left[x_i'e_j \mid F_T\right] = \Gamma' F_T' F_T \gamma$. By choosing $j = i-1$, the empirical moment conditions $\frac{1}{N}\sum_{i=2}^{N} x_i'\left(e_i - e_{i-1}\right) = 0$ yield the estimator above. Averaging the moments over both i and j, the empirical moment conditions $\frac{1}{N}\sum_{j=1}^{N}\frac{1}{N}\sum_{i=1}^{N} x_i'\left(e_i - e_j\right) = 0$ give the OLS estimator of α_0 and β_0

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}_{OLS} = \left[\sum_{i=1}^{N}\left(\omega_i - \bar{\omega}\right)'\left(\omega_i - \bar{\omega}\right)\right]^{-1} \sum_{i=1}^{N}\left(\omega_i - \bar{\omega}\right)'\left(y_i - \bar{y}\right) ,$$

where $\bar{y} = \left(1/N\right)\sum_{i=1}^{N} y_i$ and $\bar{\omega} = \left(1/N\right)\sum_{i=1}^{N}\omega_i$.

It is natural to state the assumptions conditional on the factors F_T (c.f. Andrews (2005), Kuersteiner and Prucha (2013), Forchini and Peng (2015)). In order to do this we will think of all variables as defined on a probability space $\left(\Omega, \mathcal{A}, P\right)$. The sigma algebra generated by the random vector $vec\left(F_T\right)$ is denoted by $\mathcal{F} = \left\{\omega \in \mathcal{A} : vec\left(F_T\right)\left(\omega\right) \in \mathcal{B}^{Tm}\right\}$, where $\mathcal{B}^{Tm}$ is the Borel sigma algebra in R^{Tm}. Notice that $\mathcal{F}$ is a sub-algebra of $\mathcal{A}$. Notice also that expectations and probabilities conditional on $\mathcal{F}$ are unique up to a.s. equivalence, so that for example two conditional expectations which differ only on sets of probability zero are regarded as equivalent. In the rest of this paper, $\left\|\cdot\right\|_2$ denotes the Euclidean norm for a vector and the Frobenius norm for a matrix.

The following assumptions are needed for consistency of the MFD estimator.

Assumption C.1. *The sequence of random vectors* $\{\varepsilon_i, i = 1,...,n\}$ *is conditionally independent given* $\mathcal{F}$, $E[\varepsilon_i \mid \mathcal{F}] = 0$ *a.s. and* $E\left[\|\varepsilon_i\|_2^{1+\delta} \mid \mathcal{F}\right] < \Delta < \infty$ *a.s. for some* $\delta > 0$, $i = 1,....,N$.

Assumption C.2. *The sequence of random matrices* $\{(z_i, v_i), i = 1,...,n\}$ *is conditionally independent given* $\mathcal{F}$ *with* $E\left[\|(z_i, v_i)\|_2^{2+\delta} \mid \mathcal{F}\right] < \Delta < \infty$ *a.s. for some* $\delta > 0$, $i = 1,....,N$. *Moreover,*

$$\min_{i\geq 1} \lambda_{\min}\left\{E\left[(z_i, v_i)'(z_i, v_i) \mid \mathcal{F}\right] - E\left[(z_i, v_i) \mid \mathcal{F}\right]' E\left[(z_i, v_i) \mid \mathcal{F}\right]\right\} \geq \tau > 0.$$

Assumption C.3. *The sequence of random vectors* $\{\gamma_i, i = 1,...,n\}$ *is conditionally independent given* $\mathcal{F}$, $E[\gamma_i \mid \mathcal{F}] = \gamma$ *a.s. and* $E\left[\|\gamma_i\|_2^{1+\delta} \mid \mathcal{F}\right] < \Delta < \infty$ *a.s. for some* $\delta > 0$, $i = 1,....,N$.

Assumption C.4. *The sequence of random vectors* $\{vec(\Gamma_i), i = 1,...,n\}$ *is conditionally independent given* $\mathcal{F}$, $E\left[vec(\Gamma_i) \mid \mathcal{F}\right] = vec(\Gamma)$ *a.s. and* $E\left[\|\Gamma_i\|_2^{2+\delta} \mid \mathcal{F}\right] \leq \Delta < \infty$ *a.s. for some* $\delta > 0$, $i = 1,....,N$.

Assumption C.5. *The random vectors* ε_i, $vec(z_i, v_i)$, γ_i *and* $vec(\Gamma_i)$ *are conditionally independent given* $\mathcal{F}$ *for all* $i = 1,2,...,N$. *For* $\forall F \in \mathcal{F}$, $\|F\| < \infty$.

Following Andrews (2005) the assumptions are formulated conditional on $\mathcal{F}$ (see also Kuersteiner and Prucha (2013) and). The assumptions are conditional versions of what is normally required for consistency of estimators. Bounding the minimum eigenvalue away from zero in Assumption C.2 ensures the invertibility of some of the matrices needed for establishing consistency and conditional asymptotic normality. Notice that conditional independence does not imply that the quantities ε_i, $vec(z_i, v_i)$, γ_i and $vec(\Gamma_i)$ are independent unconditionally (e.g.

Andrews (2005) and Prakasa Rao (2009)). Assumption C.5 allows ε_i, $vec(z_i, v_i)$, γ_i and $vec(\Gamma_i)$ to be unconditionally correlated, and to depend on the factors. Notice that if F_T is a scalar and it is equal to one, our model becomes a random effect model, if however, F_T is scalar but random we have a fixed effect model because the term $F_T \gamma_i$ is unconditionally correlated with the regressors. Notice also that we do not impose any restrictions on the factors because in the rest of the paper we derive asymptotic results for the case where the cross sectional dependence goes to infinity. Although F_T does not change with i, but it is a random matrix.

Our first result is given below.

Theorem 1. *Given Assumptions C.1-C.5,*

(1) $E\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix}$.

(2) $\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \to \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix}$ *a.s. as* N *tends to infinity.*

Thus, the MFD estimator is unbiased and a.s. consistent for a fixed time. To obtain the asymptotic distribution of $\hat{\alpha}$ and $\hat{\beta}$, we need slightly stronger versions of Assumptions C.1 and C.3 requiring the existence of higher order moments.

Assumption N.1. *The sequence of random vectors* $\{\varepsilon_i, i = 1, \ldots, n\}$ *is conditionally independent given* $\mathcal{F}$, $E[\varepsilon_i \mid \mathcal{F}] = 0$ *a.s.,* $E[\varepsilon_i \varepsilon_i' \mid \mathcal{F}] = \Sigma_{\varepsilon i}$ *a.s. and* $E\left[\|\varepsilon_i\|^{2+\delta} \mid \mathcal{F}\right] < \Delta < \infty$ *for some* $\delta > 0$ *a.s.,* $i = 1, \ldots., N$. *Moreover,* $\min_{i \geq 1} \lambda_{\min}\left(\Sigma_{\varepsilon_i}\right) \geq \tau > 0$.

Assumption N.3. *The sequence of random vectors* $\{\gamma_i, i=1,...,n\}$ *is conditionally independent given* $\mathcal{F}$ *,* $E[\gamma_i \mid \mathcal{F}] = \gamma$ *a.s.,* $\operatorname{cov}[\gamma_i \mid \mathcal{F}] = \Sigma_{\gamma i}$ *a.s. and* $E\left[\left\|\gamma_i\right\|^{2+\delta} \mid \mathcal{F}\right] < \Delta < \infty$ *for some* $\delta > 0$ *a.s.,* $i = 1,....,N$.

The asymptotic distribution of the MFD estimator is given in the following theorem.

Theorem 2. *If Assumptions N.1, C.2, N.3, C.4 and C.5 hold, then*

$$\text{(5)}\qquad \sqrt{N}\left(\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix}-\begin{pmatrix}\alpha_0\\ \beta_0\end{pmatrix}\right) \to^{D} \left(B(F_T)\right)^{-1} A(F_T)^{1/2}\, N\left(0, I_{p+k}\right) \textit{(stably),}$$

where

$$A(F_T) = \lim_{N\to\infty}\frac{1}{N}\sum_{i=2}^{N-1} E\left[\left(2\omega_i-\omega_{i-1}-\omega_{i+1}\right)'\left(F_T\Sigma_{\gamma_i}F_T+\Sigma_{\varepsilon_i}\right)\left(2\omega_i-\omega_{i-1}-\omega_{i+1}\right)\mid \mathcal{F}\right],$$
$$B(F_T) = \lim_{N\to\infty}\frac{1}{N}\sum_{i=2}^{N}\left(2E\left[\omega_i'\omega_i \mid \mathcal{F}\right]-E\left[\omega_i'\omega_{i-1}+\omega_{i-1}'\omega_i \mid \mathcal{F}\right]\right).$$

Since the convergence in (5) is stable in the sense of Rényi (1963), one can interpret the asymptotic distribution of $\sqrt{N}\left(\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix}-\begin{pmatrix}\alpha_0\\ \beta_0\end{pmatrix}\right)$ as being normal conditional on the factors F_T with mean vector zero and covariance matrix $\left(B(F_T)\right)^{-1} A(F_T)\left(B(F_T)\right)^{-1}$. Since the factors are not observable, the conditional distribution is not useful. When conditioning is removed, the MFD estimator has asymptotic covariance-matrix-mixed normal distribution with mixing density given by the density function of the factors which is unknown (cf. Andrews (2005), Kuersteiner and Prucha (2013), and Forchini and Peng (2015)).

Although the asymptotic distribution is non-standard, tests of hypothesis on the coefficients can rely on the standard *t* and *F* tests. In fact, it follows from Theorem 2 that

$$\left(A(F_T)\right)^{-1/2} B(F_T)\sqrt{N}\left(\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix}-\begin{pmatrix}\alpha_0\\ \beta_0\end{pmatrix}\right)\rightarrow^{D} N\left(0, I_{p+k}\right).$$

Thus, t and F tests for restrictions on α_0 and β_0 will have standard distributions under the null hypothesis if we can find $\hat{A}\rightarrow A(F_T)$ and $\hat{B}\rightarrow B(F_T)$ in probability as $N\rightarrow\infty$. This is the case if we choose:

(6) $$\hat{B}=\frac{1}{N}\sum_{i=2}^{N}(\omega_i-\omega_{i-1})'(\omega_i-\omega_{i-1}),$$

(see the proof of Theorem 1 in the Appendix), and

(7) $$\hat{A}=\frac{1}{N}\sum_{i=1}^{N-1}\hat{g}_i\hat{g}_i{}'+\frac{1}{N}\sum_{i=2}^{N-1}\hat{g}_i\hat{g}_{i-1}{}'+\frac{1}{N}\sum_{i=2}^{N-1}\hat{g}_{i-1}\hat{g}_i{}'$$

where $\hat{g}_{i-1}=(\omega_i-\omega_{i-1})'\left(y_i-\omega_i\hat{\theta}-y_{i-1}+\omega_{i-1}\hat{\theta}\right)$ and $\hat{\theta}=\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix}$.

Theorem 3 *Suppose that the conditions of Theorem 2 hold and, in addition,* $E\left[\left\|(z_i,v_i)\right\|_2^{4+\delta}\mid\mathcal{F}\right]<\infty$ *and* $E\left[\left\|\Gamma_i\right\|_2^{4+\delta}\mid\mathcal{F}\right]<\infty$. *Let* $\hat{A}$ *be defined as in (7). Then* $\hat{A}\rightarrow^{P} A(F_T)$.

Notice that given the assumptions needed in Theorem 3, convergence of the t and F test statistics to their asymptotic null distributions requires slightly stronger assumptions than those needed for the derivation of the asymptotic distribution of the MFD estimator. Notice also that the distributions of the t and F test statics under the alternative hypothesis are non-standard.

Small sample results

This section provides some Monte Carlo evidence on the properties of the MFD estimator studied in this paper. To simplify the simulation we assume that all the regressors are affected by common shocks and compare MFD with the CCEMG and CCEP estimators of Pesaran (2006). These have been designed for a model similar to ours but for a panel in which both the cross-section and the time dimensions are large. We will also include in the comparison the OLS estimator which is consistent but has a non-standard asymptotic distribution as shown by Forchini and Peng (2015).

The data generating process is:

$$y_{it} = \tau_t + \beta_1 x_{i1t} + \beta_2 x_{i2t} + e_{it}$$

$$e_{it} = \gamma_{i3} + \gamma_{i1} f_{1t} + \gamma_{i2} f_{2t} + \varepsilon_{it}$$

$$x_{ijt} = \Gamma_{ij3} + \Gamma_{ij1} f_{1t} + \Gamma_{ij2} f_{2t} + v_{ijt},$$

where $j = 1,2$, $i = 1,2,...,N$, $t = 1,...,T$, $\beta_1 = \beta_2 = 1$, $\tau_t \sim U(0,1)$. The common factors are generated as: $f_{j,-50} = 0,\ \ f_{jt} = \rho_{fj} f_{jt-1} + i.i.N(0,1)$ for $j = 1,2,\ t = -49,...,0,...,T,$ where $\rho_{f1} = 1$, $\rho_{f2} = 0.95$. The ε_{it} are generated as AR(1) process

$$\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{i,t-1} + \sigma_i \left(1-\rho_{i\varepsilon}^2\right)^{1/2} \zeta_{it}$$

for $i = 1,...,[N/2]$ and as MA(1)

$$\varepsilon_{it} = \sigma_i \left(1+\theta_{i\varepsilon}{}^2\right)^{-1/2} \left(\zeta_{it} + \theta_{i\varepsilon}\zeta_{it-1}\right)$$

for $i = [N/2]+1,...,N$, where $\zeta_{it} \sim N(0, S_T)$, $S_T = T + |f_{11}| + ... + |f_{1T}| + ... + |f_{2T}|$, $\sigma_i^2 \sim U(0.5, 1.5)$, $\rho_{i\varepsilon} \sim U[0.05, 0.95]$ and $\theta_{i\varepsilon} \sim U[0,1]$. The v_{it}'s follow

$$v_{ijt} = \rho_{vij} v_{ijt-1} + N\left(0, |f_{1t}|^2\right) \text{ for } j = 1,2,\ t = 0,...,T$$
$$v_{ij,-50} = 0, \quad \rho_{vij} \sim U[0.05, 0.95].$$

The factor loadings are generated as $\Gamma_{ijp} \sim N\left(\mu_{jp}, \sigma_{\Gamma_i}^2\right)$ and $\mu_{jp} \sim U(-0.5, 1.5)$ for $p = 1,2,3$ and as $\gamma_{i1} \sim N\left(1, \sigma_{\gamma_1}^2\right)$, $\gamma_{i2} \sim N\left(1, \sigma_{\gamma_2}^2\right)$, $\gamma_{i3} \sim N(0,1)$, $\sigma_{\Gamma_i}^2 \sim U(0, S_T)$, $\sigma_{\gamma_1}^2 = 0.5\left(|f_{21}| + |f_{22}| + ... + |f_{2T}|\right)$ and $\sigma_{\gamma_2}^2 = 0.5$. Notice that in this set-up the factor loadings Γ_{ijk} and γ_{ik} and the error terms ε_{it} and v_{ijt} are independent conditional on $\mathcal{F}$ but they are *not independent* unconditionally.

Table 1 reports bias and RMSE for the estimators of β_1 considered. Notice that we consider two MFD estimators corresponding to two different random permutations of the units i (MFD1 and MFD2), the OLS, the CCEP and CCEMG estimators. Table 1 shows that, for the small T case, the OLS and MFD estimators have similar bias and RMSE although the OLS estimator seems to perform marginally better. The CCEP and CCEMG estimators of Pesaran (2006) have been designed for large N and T and this can be noticed in the larger bias and RMSE when T is very small.

Table 1

		Bias					MSE				
Beta_1		T=3	7	10	15	20	T=3	7	10	15	20
OLS	N=20	0.000	0.000	-0.002	0.000	0.004	0.365	0.131	0.460	0.245	0.270
	50	0.002	-0.001	0.000	0.002	-0.002	0.285	0.338	0.362	0.152	0.167

	100	0.002	-0.001	0.000	0.000	0.001	0.237	0.282	0.301	0.106	0.117
MFD1	N=20	0.001	0.000	-0.001	0.000	0.003	0.401	0.472	0.507	0.298	0.328
	50	0.001	-0.002	0.000	0.003	-0.001	0.314	0.372	0.399	0.183	0.206
	100	0.002	0.000	0.001	0.001	0.000	0.263	0.313	0.335	0.130	0.142
MFD2	N=20	0.000	0.035	-0.003	-0.001	0.005	0.397	0.464	0.489	0.278	0.303
	50	0.002	0.000	0.001	0.002	-0.002	0.309	0.362	0.388	0.172	0.188
	100	0.002	-0.001	0.000	0.001	0.001	0.258	0.302	0.322	0.120	0.132
CCEP	N=20	0.044	-0.002	-0.002	0.000	0.005	3.975	0.515	0.520	0.296	0.319
	50	-5.964	0.000	0.000	0.001	0.000	24.856	0.404	0.411	0.184	0.199
	100	-0.870	-0.001	0.000	0.000	-0.001	7.366	0.341	0.344	0.130	0.142
CCEMG	N=20	-1.723	0.002	0.000	-0.001	0.007	11.840	0.751	0.623	0.372	0.383
	50	0.306	-0.005	0.003	0.000	0.000	8.179	0.607	0.499	0.238	0.242
	100	1.565	-0.003	0.000	0.001	0.001	8.576	0.516	0.420	0.168	0.173

Conclusion

The paper has investigated the MFD estimator for the parameters of panel data with factor structure both in the errors and some of the regressors when the time dimension is small. We have shown that the estimator is consistent but has a non-standard distribution as the cross-section dimension tends to infinity and the time dimension is small. A small Monte Carlo simulation shows that this estimator compares favourably with other estimators.

Appendix

In this appendix we make use the conditional weak law of large number and conditional central limit theorems. These are discussed by Forchini, Jiang and Peng (2015) and are reported below for convenience. Alterative versions of these results are given by Dedecker and Merlevede (2002), Majerek, Nowak and Zięba (2005), Prakasa Rao (2009), Cabrera, Rosalsky and Volodin (2012), and Yuan, Wei and Lei (2014).

Theorem A1. (Conditional Markov strong law of large numbers) Let $\{Z_i : i \geq 1\}$ *be a sequence of* $\mathcal{F}$ *-independent random variables such that* $E\left(\left|Z_i\right|^{1+\delta} \mid \mathcal{F}\right) < \Delta$ *for some* $\delta > 0$ *and some* Δ *which is* $\mathcal{F}$ *-measurable and* $\Delta < \infty$ *a.s.. Then conditionally on* $\mathcal{F}$ *,* $\frac{1}{n}\sum_{i=1}^{n}\left(Z_i - E\left(Z_i \mid \mathcal{F}\right)\right) \to 0$ *a.s..*

Theorem A2. (Conditional Liapounov central limit theorem) *Let* $\{Z_i : i \geq 1\}$ *be a sequence of* $\mathcal{F}$ *-independent random variables with conditional means* $E\left(Z_i \mid \mathcal{F}\right)$ *, conditional variances* $\sigma_i^2 = E\left(\left(Z_i - E\left(Z_i \mid \mathcal{F}\right)\right)^2 \mid \mathcal{F}\right)$ *, and* $E\left(\left(Z_i - E\left(Z_i \mid \mathcal{F}\right)\right)^{2+\delta} \mid \mathcal{F}\right) < \Delta$ *a.s. for* Δ *arbitrarily* $\mathcal{F}$ *-measurable,* $\Delta < \infty$ *a.s. and some* $\delta > 0$ *,* $i = 1, 2, \ldots$ *. If there is* η $\mathcal{F}$ *-measurable such that* $\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 > \eta > 0$ *holds a.s., then conditionally on* $\mathcal{F}$ *,* $\frac{1}{\bar{\sigma}_n\sqrt{n}}\sum_{i=1}^{n}\left(Z_i - E\left(Z_i \mid \mathcal{F}\right)\right) \to^{D} N(0,1)$*. a.s..*

Also, $\frac{1}{\bar{\sigma}_n\sqrt{n}}\sum_{i=1}^{n}\left(Z_i - E\left(Z_i \mid \mathcal{F}\right)\right) \to^{D} N(0,1)$ *(*$\mathcal{F}$ *-stably).*

The notion of $\mathcal{F}$ -stable convergence used in Theorem A2 is due to Daley and Vere-Jones (1988) (see also Kuersteiner and Prucha (2013)) by modifying the notion of stable convergence due to Rényi (1963): he sequence of random variables ξ_i, $i = 1, 2, \ldots.$ defined on a probability space $\left(\Omega, \mathcal{A}, P\right)$ is $\mathcal{F}$ -stable if for any event $B \in \mathcal{F} \subseteq \mathcal{A}$ with $P\left(B\right) > 0$ the conditional distribution of ξ_i given B tends to a limiting distribution, $\lim_{i\to\infty} P\left(\xi_i < x \mid B\right) = F_B\left(x\right)$ for every x which is a continuity point of the distribution function $F_B\left(x\right)$, written as $\xi_i \to X$ ($\mathcal{F}$ -stably) (c.f. Rényi (1963), p. 294).

Lemma 1. *If Assumptions C.1-C.5 hold, then*

1) $\frac{1}{N}\sum_{i=2}^{N}\omega_i{}'\omega_i \to \lim_{N\to\infty}\frac{1}{N}\sum_{i=2}^{N}E\left[\omega_i{}'\omega_i \mid \mathcal{F}\right]$ *a.s.;*

2) $\frac{1}{N}\sum_{i=2}^{N}\omega_i{}'e_i \rightarrow \lim_{N\to\infty}\frac{1}{N}\sum_{i=2}^{N}E\left[\left(z_i,v_i\right)'\,|\,\mathcal{F}\right]+\left(0,F_T\Gamma\right)'F_T\gamma$ *a.s.;*

3) $\frac{1}{N}\sum_{i=2}^{N}\omega_i{}'e_{i-1} \rightarrow \lim_{N\to\infty}\frac{1}{N}\sum_{i=2}^{N}E\left[\left(z_i,v_i\right)'\,|\,\mathcal{F}\right]+\left(0,F_T\Gamma\right)'F_T\gamma$ *a.s.;*

4) $\frac{1}{N}\sum_{i=2}^{N}\omega_{i-1}{}'e_i \rightarrow \lim_{N\to\infty}\frac{1}{N}\sum_{i=2}^{N}E\left[\left(z_{i-1},v_{i-1}\right)'\,|\,\mathcal{F}\right]+\left(0,F_T\Gamma\right)'F_T\gamma$ *a.s.;*

5) $\frac{1}{N}\sum_{i=2}^{N}\omega_i{}'\omega_{i-1} \rightarrow \lim_{N\to\infty}\frac{1}{N}\sum_{i=2}^{N}E\left[\omega_i{}'\omega_{i-1}\,|\,\mathcal{F}\right]$ *a.s.;*

Notice that 1) and 5) could have been expressed more precisely in terms of F_T and the conditional expectation of the components of ω_i.

Proof of Lemma 1. 1) to 4) follow from Theorem A1. 5) We assume that N is an odd number – if N is an even number, we just need to adjust the grouping in the following equation accordingly and write

$$\frac{1}{N-1}\sum_{i=2}^{N}\omega_i{}'\omega_{i-1}=\frac{1}{2}\frac{1}{\left(N-1\right)/2}\sum_{i=1}^{(N-1)/2}\omega_{2i}{}'\omega_{2i-1}+\frac{1}{2}\frac{1}{\left(N-1\right)/2}\sum_{i=1}^{(N-1)/2}\omega_{2i}{}'\omega_{2i+1}$$

The two sums only involve independent terms, their convergence follows from Theorem A1.

Proof of Theorem 1. For simplicity, let $\hat{\theta}=\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix}$ and $\theta_0=\begin{pmatrix}\alpha_0\\ \beta_0\end{pmatrix}$. Unbiasedness follows from observing that

$$
\begin{aligned}
E\left[\hat{\theta}\right]-\theta_0 &= E\left[E\left[\left(\sum_{i=2}^{N}(\omega_i-\omega_{i-1})'(\omega_i-\omega_{i-1})\right)^{-1}\sum_{i=2}^{N}(\omega_i-\omega_{i-1})'(e_i-e_{i-1})\,|\,\mathcal{F}\right]\right] \\
&= E\left[\sum_{j=2}^{N}E\left[\left(\sum_{i=2}^{N}(\omega_i-\omega_{i-1})'(\omega_i-\omega_{i-1})\right)^{-1}(\omega_j-\omega_{j-1})'\,|\,\mathcal{F}\right]E\left[(e_j-e_{j-1})\,|\,\mathcal{F}\right]\right] \\
&= 0.
\end{aligned}
$$

To verify consistency we write

$$
\hat{\theta}-\theta_0=\left(\frac{1}{N}\sum_{i=2}^{N}(\omega_i-\omega_{i-1})'(\omega_i-\omega_{i-1})\right)^{-1}\frac{1}{N}\sum_{i=2}^{N}(\omega_i-\omega_{i-1})'(e_i-e_{i-1}).
$$

As N tends to infinity, conditional on $\mathcal{F}$, $\frac{1}{N}\sum_{i=2}^{N}(\omega_i-\omega_{i-1})'(\omega_i-\omega_{i-1})\to B(F_T)$ a.s. from 1) and 5) of Lemma 1. Moreover,

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=2}^{N}E\left[(\omega_i-\omega_{i-1})'(\omega_i-\omega_{i-1})\,|\,\mathcal{F}\right] \\
&=\frac{2}{N}\sum_{i=2}^{N-1}E\left[\left(\omega_i-E\left[\omega_i\,|\,\mathcal{F}\right]\right)'\left(\omega_i-E\left[\omega_i\,|\,\mathcal{F}\right]\right)|\,\mathcal{F}\right] \\
&+\frac{1}{N}E\left[\left(\omega_1-E\left[\omega_1\,|\,\mathcal{F}\right]\right)'\left(\omega_1-E\left[\omega_1\,|\,\mathcal{F}\right]\right)|\,\mathcal{F}\right] \\
&+\frac{1}{N}E\left[\left(\omega_N-E\left[\omega_N\,|\,\mathcal{F}\right]\right)'\left(\omega_N-E\left[\omega_N\,|\,\mathcal{F}\right]\right)|\,\mathcal{F}\right] \\
&+\frac{1}{N}\sum_{i=2}^{N}\left(E\left[\omega_i\,|\,\mathcal{F}\right]-E\left[\omega_{i-1}\,|\,\mathcal{F}\right]\right)'\left(E\left[\omega_i\,|\,\mathcal{F}\right]-E\left[\omega_{i-1}\,|\,\mathcal{F}\right]\right).
\end{aligned}
$$

Since the last three terms in the expression above are positive semi-definite matrices, we can write

$$
\begin{aligned}
&\lambda_{\min}\left\{\frac{1}{N}\sum_{i=2}^{N}E\left[\left(\omega_i-\omega_{i-1}\right)'\left(\omega_i-\omega_{i-1}\right)\mid\mathcal{F}\right]\right\}\\
&\geq 2\lambda_{\min}\left\{\frac{1}{N}\sum_{i=2}^{N-1}E\left[\left(\omega_i-E\left[\omega_i\mid\mathcal{F}\right]\right)'\left(\omega_i-E\left[\omega_i\mid\mathcal{F}\right]\right)\mid\mathcal{F}\right]\right\}\\
&\geq 2\min_{i\geq 1}\lambda_{\min}\left\{E\left[\left(\omega_i-E\left[\omega_i\mid\mathcal{F}\right]\right)'\left(\omega_i-E\left[\omega_i\mid\mathcal{F}\right]\right)\mid\mathcal{F}\right]\right\}
\end{aligned}
$$

by Exercise 1 on page 204 in Magnus and Neudecker (1999). Notice that

$$
\begin{aligned}
&E\left[\left(\omega_i-E\left[\omega_i\mid\mathcal{F}\right]\right)'\left(\omega_i-E\left[\omega_i\mid\mathcal{F}\right]\right)\mid\mathcal{F}\right]\\
&=E\left[\left(\left(z_i,v_i\right)-E\left[\left(z_i,v_i\right)\mid\mathcal{F}\right]\right)'\left(\left(z_i,v_i\right)-E\left[\left(z_i,v_i\right)\mid\mathcal{F}\right]\right)\mid\mathcal{F}\right]\\
&+\begin{pmatrix}0 & 0\\ 0 & E\left[\left(\Gamma_i-\Gamma\right)'F_T{}'F_T\left(\Gamma_i-\Gamma\right)\mid\mathcal{F}\right]\end{pmatrix}
\end{aligned}
$$

Again, by Exercise 1 on page 204 in Magnus and Neudecker (1999) and Assumption C.2

$$
\begin{aligned}
&\min_{i\geq 1}\lambda_{\min}\left\{E\left[\left(\omega_i-E\left[\omega_i\mid\mathcal{F}\right]\right)'\left(\omega_i-E\left[\omega_i\mid\mathcal{F}\right]\right)\mid\mathcal{F}\right]\right\}\\
&\geq\min_{i\geq 1}\lambda_{\min}\left\{E\left[\left(\left(z_i,v_i\right)-E\left[\left(z_i,v_i\right)\mid\mathcal{F}\right]\right)'\left(\left(z_i,v_i\right)-E\left[\left(z_i,v_i\right)\mid\mathcal{F}\right]\right)\mid\mathcal{F}\right]\right\}\geq\tau>0.
\end{aligned}
$$

Therefore, $\frac{1}{N}\sum_{i=2}^{N}E\left[\left(\omega_i-\omega_{i-1}\right)'\left(\omega_i-\omega_{i-1}\right)\mid\mathcal{F}\right]$ is positive definite uniformly in N .

For the error terms, $\frac{1}{N}\sum_{i=2}^{N}\left(\omega_i-\omega_{i-1}\right)'\left(e_i-e_{i-1}\right)\to 0$ a.s. from (2)-(4) of Lemma 1. The second result of the theorem follows.

Lemma 2. *Let* $\eta_{i-1}=\zeta'\left(\omega_i-\omega_{i-1}\right)'\left(e_i-e_{i-1}\right)$, *where* ζ *be an arbitrary* $\left(p+k\right)\times 1$ *vector. Given Assumptions C2, C4 and C5,* $E\left[\left|\eta_{i-1}\right|^{2+\delta}\mid\mathcal{F}\right]<\Delta_*<\infty$ *a.s..*

Proof of Lemma 2. Write

$$
\begin{aligned}
& E\left[\left|\eta_{i-1}\right|^{2+\delta} \mid \mathcal{F}\right]=E\left[\left|\zeta'\left(\omega_i-\omega_{i-1}\right)'\left(e_i-e_{i-1}\right)\right|^{2+\delta} \mid \mathcal{F}\right] \\
& \leq E\left[\left\|\left(\omega_i-\omega_{i-1}\right)\zeta\right\|_2^{2+\delta}\left\|e_i-e_{i-1}\right\|_2^{2+\delta} \mid \mathcal{F}\right] \quad \text{(Submultiplicativity of the } \|.\|_2 \text{ norm)} \\
(8) \qquad & \leq E\left[\left(\left\|\omega_i\zeta\right\|_2+\left\|\omega_{i-1}\zeta\right\|_2\right)^{2+\delta}\left(\left\|e_i\right\|_2+\left\|e_{i-1}\right\|_2\right)^{2+\delta} \mid \mathcal{F}\right] \quad \text{(Triangle inequality)} \\
& = E\left[\left(\left\|\omega_i\zeta\right\|_2+\left\|\omega_{i-1}\zeta\right\|_2\right)^{2+\delta} \mid \mathcal{F}\right]E\left[\left(\left\|e_i\right\|_2+\left\|e_{i-1}\right\|_2\right)^{2+\delta} \mid \mathcal{F}\right] \quad \text{(Assumption C.5)} \\
& \leq 2^{2+2\delta}\left(E\left[\left\|\omega_i\zeta\right\|_2^{2+\delta} \mid \mathcal{F}\right]+E\left[\left\|\omega_{i-1}\zeta\right\|_2^{2+\delta} \mid \mathcal{F}\right]\right)\left(E\left[\left\|e_i\right\|_2^{2+\delta} \mid \mathcal{F}\right]+E\left[\left\|e_{i-1}\right\|_2^{2+\delta} \mid \mathcal{F}\right]\right) \\
& \qquad (c_r \text{ inequality}).
\end{aligned}
$$

We now focus on the term $E\left[\left\|\omega_i\zeta\right\|_2^{2+\delta} \mid \mathcal{F}\right]$. The other terms can be bounded in the same way. Notice that

$$
\begin{aligned}
& E\left[\left\|\omega_i\zeta\right\|_2^{2+\delta} \mid \mathcal{F}\right]=E\left[\left\|\left(z_i, v_i\right)\zeta+\left(0, F_T\Gamma_i\right)\zeta\right\|_2^{2+\delta} \mid \mathcal{F}\right] \\
& \leq E\left[\left(\left\|\left(z_i, v_i\right)\zeta\right\|_2+\left\|\left(0, F_T\Gamma_i\right)\zeta\right\|_2\right)^{2+\delta} \mid \mathcal{F}\right] \quad \text{(Triangle inequality)} \\
(9) \qquad & \leq 2^{1+\delta}\left(E\left[\left\|\left(z_i, v_i\right)\zeta\right\|_2^{2+\delta} \mid \mathcal{F}\right]+E\left[\left\|\left(0, F_T\Gamma_i\right)\zeta\right\|_2^{2+\delta} \mid \mathcal{F}\right]\right) \quad (c_r \text{ inequality}) \\
& = 2^{1+\delta}\left\|\zeta\right\|_2^{2+\delta}\left(E\left[\left\|\left(z_i, v_i\right)\right\|_2^{2+\delta} \mid \mathcal{F}\right]+E\left[\left\|\left(0, F_T\Gamma_i\right)\right\|_2^{2+\delta} \mid \mathcal{F}\right]\right) \quad \text{(Definition of } \|.\|_2 \text{ norm)} \\
& \leq 2^{1+\delta}\left\|\zeta\right\|_2^{2+\delta}\left(E\left[\left\|\left(z_i, v_i\right)\right\|_2^{2+\delta} \mid \mathcal{F}\right]+\left\|F_T\right\|_2^{2+\delta} E\left[\left\|\Gamma_i\right\|_2^{2+\delta} \mid \mathcal{F}\right]\right) \quad \text{(Definition of } \|.\|_2 \text{ norm)}
\end{aligned}
$$

The first term, $E\left[\left\|\left(z_i, v_i\right)\right\|_2^{2+\delta} \mid \mathcal{F}\right]$, is uniformly bounded by Assumption C.2. The second term is uniformly bounded because of Assumptions C.4 and C.5. Thus, $E\left[\left|\eta_i\right|^{2+\delta} \mid \mathcal{F}\right]$ is uniformly bounded.

The other terms can be bounded in the same way. The proof is now completed.

Proof of Theorem 2. We now verify the normality for the MFD estimator. Write

$$\sqrt{N}\left(\hat{\theta}-\theta_0\right)=\left(\frac{1}{N}\sum_{i=2}^{N}\left(\omega_i-\omega_{i-1}\right)'\left(\omega_i-\omega_{i-1}\right)\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=2}^{N}\left(\omega_i-\omega_{i-1}\right)'\left(e_i-e_{i-1}\right).$$

We have already shown in the proof of Theorem 1 that $\frac{1}{N}\sum_{i=2}^{N}\left(\omega_i-\omega_{i-1}\right)'\left(\omega_i-\omega_{i-1}\right)\to B\left(F_T\right)$ a.s. and it is positive definite uniformly in N.

For the other term, $\frac{1}{\sqrt{N}}\sum_{i=2}^{N}\left(\omega_i-\omega_{i-1}\right)'\left(e_i-e_{i-1}\right)$, we use Cramer-Wold device. Let ζ be an arbitrary $\left(p+k\right)\times 1$ vector and denote $\eta_{i-1}=\zeta'\left(\omega_i-\omega_{i-1}\right)'\left(e_i-e_{i-1}\right)$ as in Lemma 2. Then $\frac{1}{\sqrt{N}}\sum_{i=2}^{N}\zeta'\left(\omega_i-\omega_{i-1}\right)'\left(e_i-e_{i-1}\right)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N-1}\eta_i$. Lemma 2 shows that $E\left[\left|\eta_i\right|^{2+\delta}\mid\mathcal{F}\right]$ is uniformly bounded.

By construction, we have a *m*-dependent sequence $\left\{\eta_1,\ldots,\eta_{N-1}\right\}$, where $m=1$. We split the sample in h groups of k elements plus $r<k$ remainder terms as illustrated in the following display

$$\underbrace{\underbrace{\eta_1,\ldots,\eta_{k-1}}_{\mu_1=\eta_1+\cdots+\eta_{k-1}},\eta_k}_{group\ 1},\ldots,\underbrace{\underbrace{\eta_{(h-1)k+1},\ldots,\eta_{hk-1}}_{\mu_h=\eta_{(h-1)k+1}+\cdots+\eta_{hk-1}},\eta_{hk}}_{group\ h},\underbrace{\eta_{hk+1},\ldots,\eta_{hk+r}}_{remainders}.$$

and we will allow N to tend to infinity by allowing both h and k to tend to infinity in such a way that $\frac{hk}{N}\to 1$, $\frac{k}{N}\to 0$ and $\frac{h}{N}\to 0$.

We now take the first $k-1$ terms in each of the h groups and add them up,

$$\mu_1=\eta_1+\cdots+\eta_{k-1};\ \mu_2=\eta_{k+1}+\cdots+\eta_{2k-1};\ \cdots;\ \mu_h=\eta_{(h-1)k+1}+\cdots+\eta_{hk-1}.$$

Notice that $\mu_1,\dots,\mu_h$ are independent by construction. For $j=1,\dots,h$,

$$\begin{aligned}
E\left[\mu_j \mid \mathcal{F}\right] &= E\left[\eta_{(j-1)k+1}+\cdots+\eta_{jk-1} \mid \mathcal{F}\right] = \sum_{p=1}^{k-1} E\left[\eta_{(j-1)k+p} \mid \mathcal{F}\right] \\
&= \sum_{p=1}^{k-1} E\left[\zeta'\left(\omega_{(j-1)k+p+1}-\omega_{(j-1)k+p}\right)'\left(e_{(j-1)k+p+1}-e_{(j-1)k+p}\right) \mid \mathcal{F}\right] \\
&= \zeta' \sum_{p=1}^{k-1} E\left[\left(\omega_{(j-1)k+p+1}-\omega_{(j-1)k+p}\right)' \mid \mathcal{F}\right] E\left[e_{(j-1)k+p+1}-e_{(j-1)k+p} \mid \mathcal{F}\right] = 0,
\end{aligned}$$

because $E\left[e_{(j-1)k+p+1}-e_{(j-1)k+p} \mid \mathcal{F}\right] = F_T\gamma - F_T\gamma = 0$ from assumptions C.1 and C.3. Moreover,

$$\begin{aligned}
E\left[\left|\mu_j\right|^{2+\delta} \mid \mathcal{F}\right] &= E\left[\left|\sum_{p=1}^{k-1}\eta_{(j-1)k+p}\right|^{2+\delta} \mid \mathcal{F}\right] \\
&\le E\left[\left(\sum_{p=1}^{k-1}\left|\eta_{(j-1)k+p}\right|\right)^{2+\delta} \mid \mathcal{F}\right] \quad \text{(Triangle inequality)} \\
&\le (k-1)^{1+\delta}\sum_{p=1}^{k-1} E\left[\left|\eta_{(j-1)k+p}\right|^{2+\delta} \mid \mathcal{F}\right] \quad (c_r \text{ inequality}) \\
&\le (k-1)^{2+\delta}\,\Delta_* \quad \text{(Lemma 2)},
\end{aligned}$$

where $E\left[\left|\eta_{(j-1)k+p}\right|^{2+\delta} \mid \mathcal{F}\right] < \infty$ has been shown above. Then, by Theorem A2, as $h\to\infty$,

(10) $$\frac{1}{\sigma_h}\sum_{j=1}^{h}\mu_j \to^{D} N(0,1) \; (\mathcal{F}\text{-stably}).$$

Notice that

(11) $$\frac{1}{\sqrt{N}}\sum_{j=1}^{N-1}\eta_j = \frac{1}{\sqrt{N}}\sum_{j=1}^{h}\mu_j + \frac{1}{\sqrt{N}}\sum_{j=1}^{h}\eta_{jk} + \frac{1}{\sqrt{N}}\left(\eta_{hk+1}+\cdots+\eta_{hk+r}\right)$$

We now focus on the term $\frac{1}{\sqrt{N}}\left(\eta_{hk+1}+\cdots+\eta_{hk+r}\right)$. Its expected value is zero and its variance is

$$\begin{aligned}\operatorname{var}\left[\frac{1}{\sqrt{N}}\left(\eta_{hk+1}+\cdots+\eta_{hk+r}\right)\mid\mathcal{F}\right] &= \frac{1}{N}\sum_{l=1}^{r}\operatorname{var}\left[\eta_{hk+l}\mid\mathcal{F}\right]+2\frac{1}{N}\sum_{l=1}^{r-1}\operatorname{cov}\left[\eta_{hk+l}\eta_{hk+l+1}\mid\mathcal{F}\right]\\ &\le \frac{1}{N}\sum_{l=1}^{r}\operatorname{var}\left[\eta_{hk+l}\mid\mathcal{F}\right]+2\frac{1}{N}\sum_{l=1}^{r-1}\left|\operatorname{cov}\left[\eta_{hk+l}\eta_{hk+l+1}\mid\mathcal{F}\right]\right|\\ &\le \frac{1}{N}\sum_{l=1}^{r}\operatorname{var}\left[\eta_{hk+l}\mid\mathcal{F}\right]+2\frac{1}{N}\sum_{l=1}^{r-1}\operatorname{var}\left[\eta_{hk+l}\mid\mathcal{F}\right]\operatorname{var}\left[\eta_{hk+l+1}\mid\mathcal{F}\right].\end{aligned}$$

It follows from Lemma 2 that

$$\operatorname{var}\left[\frac{1}{\sqrt{N}}\left(\eta_{hk+1}+\cdots+\eta_{hk+r}\right)\mid\mathcal{F}\right]\le\frac{1}{N}r\Delta_{*}+2\frac{1}{N}\left(r-1\right)\Delta_{*}^{2}\to 0$$

as h and k tend to infinity. Thus, $\frac{1}{\sqrt{N}}\left(\eta_{hk+1}+\cdots+\eta_{hk+r}\right)=o_p(1)$.

We then focus on the term $\frac{1}{\sqrt{N}}\sum_{j=1}^{h}\eta_{jk}$. It is easy to show that $E\left[\frac{1}{\sqrt{N}}\sum_{j=1}^{h}\eta_{jk}\mid\mathcal{F}\right]=0$ and for any $k>1$

$$\operatorname{var}\left[\frac{1}{\sqrt{N}}\sum_{j=1}^{h}\eta_{jk}\mid\mathcal{F}\right]=\frac{1}{N}\sum_{j=1}^{h}\operatorname{var}\left[\eta_{jk}\mid\mathcal{F}\right].$$

It follows from Lemma 2 that $\operatorname{var}\left[\frac{1}{\sqrt{N}}\sum_{j=1}^{h}\eta_{jk}\mid\mathcal{F}\right]\le\frac{h}{N}\Delta_{*}\to 0$ as h and k tend to infinity so that $\frac{1}{\sqrt{N}}\sum_{j=1}^{h}\eta_{jk}=o_p\left(1\right)$.

Finally, we consider the term $\frac{1}{\sqrt{N}}\sum_{j=1}^{h}\mu_j=\frac{\sigma_h}{\sqrt{N}}\left(\frac{1}{\sigma_h}\sum_{j=1}^{h}\mu_j\right)$.

Let

(12) $$\eta_N^2 = \sum_{\substack{i,j=1\\|j-i|\le 1}}^{N} E\left[\eta_i\eta_j \mid \mathcal{F}\right] = \sum_{i=1}^{N}\mathrm{var}\left[\eta_i \mid \mathcal{F}\right] + 2\sum_{i=1}^{N-1}\mathrm{cov}\left[\eta_i,\eta_{i+1} \mid \mathcal{F}\right],$$

and notice that

$$\eta_N^2 - \sigma_h^2 = \sum_{j=1}^{h}\mathrm{var}\left[\eta_{jk} \mid \mathcal{F}\right] + \sum_{p=1}^{r}\mathrm{var}\left[\eta_{jk+p} \mid \mathcal{F}\right] + 2\sum_{p=1}^{r-1}\mathrm{cov}\left[\eta_{jk+p}\eta_{jk+p+1} \mid \mathcal{F}\right].$$

Dividing by N,

$$\frac{1}{N}\eta_N^2 - \frac{1}{N}\sigma_h^2 = \frac{h}{N}\frac{1}{h}\sum_{j=1}^{h}\mathrm{var}\left[\eta_{jk} \mid \mathcal{F}\right] + \frac{r}{N}\frac{1}{r}\sum_{p=1}^{r}\mathrm{var}\left[\eta_{jk+p} \mid \mathcal{F}\right] + 2\frac{r}{N}\frac{1}{r}\sum_{p=1}^{r-1}\mathrm{cov}\left[\eta_{jk+p}\eta_{jk+p+1} \mid \mathcal{F}\right].$$

We now show that each term on the right hand side converges to zero in probability. For the first term, $\left|\frac{h}{N}\frac{1}{h}\sum_{j=1}^{h}\mathrm{var}\left[\eta_{jk} \mid \mathcal{F}\right]\right| \le \frac{h}{N}\frac{1}{h}\sum_{j=1}^{h}E\left[\left|\eta_{jk}\right|^2 \mid \mathcal{F}\right] \to 0$, because $E\left[\left|\eta_{jk}\right|^2 \mid \mathcal{F}\right]$ is uniformly bounded by Lemma 2 and $h/N \to 0$ from the construction. Similarly we can show that $\left|\frac{r}{N}\frac{1}{r}\sum_{p=1}^{r}\mathrm{var}\left[\eta_{jk+p} \mid \mathcal{F}\right]\right| \to 0$ and $\left|\frac{r}{N}\frac{1}{r}\sum_{p=1}^{r-1}\mathrm{cov}\left[\eta_{jk+p}\eta_{jk+p+1} \mid \mathcal{F}\right]\right| \to 0$. Then we obtain that $\left|\frac{1}{N}\eta_N^2 - \frac{1}{N}\sigma_h^2\right| \to 0$ in probability.

The final step consists in showing that the limit $\lim_{N\to\infty}\left(\eta_N^2 / N\right)$ is well defined a.s.. Write

$$\frac{1}{N}\eta_N^2 = \frac{1}{N}\sum_{i=1}^{N}\mathrm{var}\left[\eta_i \mid \mathcal{F}\right] + 2\frac{1}{N}\sum_{i=1}^{N-1}\mathrm{cov}\left[\eta_i,\eta_{i+1} \mid \mathcal{F}\right],$$

and notice that

$$\frac{1}{N}\sum_{i=1}^{N-1}\mathrm{cov}\left[\eta_i,\eta_{i+1} \mid \mathcal{F}\right] \le \frac{1}{N}\sum_{i=1}^{N-1}\left|\mathrm{cov}\left[\eta_i,\eta_{i+1} \mid \mathcal{F}\right]\right| \le \frac{1}{N}\sum_{i=1}^{N-1}\mathrm{var}\left[\eta_i \mid \mathcal{F}\right]^{1/2}\mathrm{var}\left[\eta_{i+1} \mid \mathcal{F}\right]^{1/2}.$$

Since

$$\operatorname{var}\left[\eta_i \mid \mathcal{F}\right] = E\left[\eta_i^{\,2} \mid \mathcal{F}\right] = E\left[\left|\zeta'\left(\omega_{i-1}-\omega_{i-2}\right)'\left(e_{i-1}-e_{i-2}\right)\right|^2 \mid \mathcal{F}\right]$$

and the right-hand side is uniformly bounded by Lemma 2 with $\delta = 0$. It follows that $\frac{1}{N}\eta_N^2$ is absolutely convergent. Thus,

$$\begin{aligned}
&\left|\frac{1}{N}\sigma_h^2 - \lim_{N\to\infty}\frac{1}{N}\eta_N^2\right| = \left|\frac{1}{N}\sigma_h^2 - \frac{1}{N}\eta_N^2 + \frac{1}{N}\eta_N^2 - \lim_{N\to\infty}\frac{1}{N}\eta_N^2\right| \\
&\le \left|\frac{1}{N}\sigma_h^2 - \frac{1}{N}\eta_N^2\right| + \left|\frac{1}{N}\eta_N^2 - \lim_{N\to\infty}\frac{1}{N}\eta_N^2\right| \to 0
\end{aligned}$$

and we can conclude that

$$\frac{1}{\sqrt{N}}\sum_{j=1}^{N-1}\eta_j = \frac{\sigma_h}{\sqrt{N}}\left(\frac{1}{\sigma_h}\sum_{j=1}^{h}\mu_j\right) + o_p(1) \to^D \eta \times N(0,1) \tag{13}$$

as $N \to \infty$. The fact that the convergence is stable follows from (10) and Theorem 1 of Aldous and Eagleson (1978), $\left(\frac{1}{\sigma_h}\sum_{j=1}^{h}\mu_j, \frac{\sigma_h}{\sqrt{N}}\right) \to \left(N(0,1), \eta\right)$ stably, so that the convergence in (13) is also stable.

Notice that

$$
\begin{aligned}
\lim_{N\to\infty}\left(\frac{1}{N}\eta_N^2\right) &= \lim_{N\to\infty}\frac{1}{N}\sum_{i=2}^{N-1}\zeta' E\left[\left(2\omega_i-\omega_{i-1}-\omega_{i+1}\right)'\left(F_T\Sigma_{\gamma_i}F_T'+\Sigma_{\varepsilon_i}\right)\left(2\omega_i-\omega_{i-1}-\omega_{i+1}\right)\mid\mathcal{F}\right]\zeta \\
&\quad+\lim_{N\to\infty}\frac{1}{N}\zeta' E\left[\left(\omega_N-\omega_{N-1}\right)'\left(F_T\Sigma_{\gamma_N}F_T'+\Sigma_{\varepsilon_N}\right)\left(\omega_N-\omega_{N-1}\right)\mid\mathcal{F}\right]\zeta \\
&\quad+\lim_{N\to\infty}\frac{1}{N}\zeta' E\left[\left(\omega_2-\omega_1\right)'\left(F_T\Sigma_{\gamma_1}F_T'+\Sigma_{\varepsilon_1}\right)\left(\omega_2-\omega_1\right)\mid\mathcal{F}\right]\zeta \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{i=2}^{N-1}\zeta' E\left[\left(2\omega_i-\omega_{i-1}-\omega_{i+1}\right)'\left(F_T\Sigma_{\gamma_i}F_T'+\Sigma_{\varepsilon_i}\right)\left(2\omega_i-\omega_{i-1}-\omega_{i+1}\right)\mid\mathcal{F}\right]\zeta \\
&= \zeta' A\left(F_T\right)\zeta.
\end{aligned}
$$

Therefore,

$$
\frac{1}{\sqrt{N}}\sum_{j=1}^{N-1}\left(\omega_{j-1}-\omega_j\right)'\left(e_{j-1}-e_j\right)\to A\left(F_T\right)^{1/2}\times N\left(0,I_{p+k}\right)\ \text{(stably).}
$$

Proof of Theorem 3. Let $g_{i-1}=\left(\omega_i-\omega_{i-1}\right)'\left(e_i-e_{i-1}\right)$ and notice that $E\left[g_{i-1}\mid\mathcal{F}\right]=0$ and define

$$
\hat{g}_{i-1}=\left(\omega_i-\omega_{i-1}\right)'\left(y_i-\omega_i\hat{\theta}-y_{i-1}+\omega_{i-1}\hat{\theta}\right)
$$

where $\hat{\theta}=\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix}$. Notice that

$$
\begin{aligned}
\hat{g}_{i-1} &= \left(\omega_i-\omega_{i-1}\right)'\left(y_i-\omega_i\hat{\theta}-y_{i-1}+\omega_{i-1}\hat{\theta}\right) \\
&= \left(\omega_i-\omega_{i-1}\right)'\left(\omega_i\left(\theta_0-\hat{\theta}\right)+e_i-\omega_{i-1}\left(\theta_0-\hat{\theta}\right)-e_{i-1}\right) \\
&= \left(\omega_i-\omega_{i-1}\right)'\left(\omega_i-\omega_{i-1}\right)\left(\theta_0-\hat{\theta}\right)+\left(\omega_i-\omega_{i-1}\right)'\left(e_i-e_{i-1}\right) \\
&= \left(\omega_i-\omega_{i-1}\right)'\left(\omega_i-\omega_{i-1}\right)\left(\theta_0-\hat{\theta}\right)+g_{i-1}.
\end{aligned}
$$

So that $\frac{1}{N}\sum_{i=1}^{N-1}\hat{g}_i\hat{g}_i'$ can be written as

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N-1}\hat{g}_i\hat{g}_i' &= \frac{1}{N}\sum_{i=1}^{N-1}(\omega_{i+1}-\omega_i)'(\omega_{i+1}-\omega_i)\left(\theta_0-\hat{\theta}\right)\left(\theta_0-\hat{\theta}\right)'(\omega_{i+1}-\omega_i)'(\omega_{i+1}-\omega_i) \\ &+\frac{1}{N}\sum_{i=1}^{N-1}g_i\left(\theta_0-\hat{\theta}\right)'(\omega_{i+1}-\omega_i)'(\omega_{i+1}-\omega_i) \\ &+\frac{1}{N}\sum_{i=1}^{N-1}(\omega_{i+1}-\omega_i)'(\omega_{i+1}-\omega_i)\left(\theta_0-\hat{\theta}\right)g_i' \\ &+\frac{1}{N}\sum_{i=1}^{N-1}g_ig_i'.\end{aligned} \tag{14}$$

For the first term on right-hand-side above, we notice that

$$\begin{aligned}&\left\|\frac{1}{N}\sum_{i=1}^{N-1}(\omega_{i+1}-\omega_i)'(\omega_{i+1}-\omega_i)\left(\theta_0-\hat{\theta}\right)\left(\theta_0-\hat{\theta}\right)'(\omega_{i+1}-\omega_i)'(\omega_{i+1}-\omega_i)\right\|_2 \\ &\le \frac{1}{N}\sum_{i=1}^{N-1}\left\|(\omega_{i+1}-\omega_i)'(\omega_{i+1}-\omega_i)\left(\theta_0-\hat{\theta}\right)\left(\theta_0-\hat{\theta}\right)'(\omega_{i+1}-\omega_i)'(\omega_{i+1}-\omega_i)\right\|_2 \quad \text{(Triangle inequality)}. \\ &\le \left\|\theta_0-\hat{\theta}\right\|_2^2\frac{1}{N}\sum_{i=1}^{N-1}\left\|\omega_{i+1}-\omega_i\right\|_2^4 \quad \text{(Submultiplicativity)}\end{aligned}$$

Moreover,

$$\begin{aligned}E\left[\left\|\omega_{i+1}-\omega_i\right\|_2^{4+\delta}\mid\mathcal{F}\right] &\le E\left[\left(\left\|\omega_{i+1}\right\|_2+\left\|\omega_i\right\|_2\right)^{4+\delta}\mid\mathcal{F}\right] \quad \text{(Triangle inequality)} \\ &\le 2^3\left(E\left[\left\|\omega_{i+1}\right\|_2^{4+\delta}\mid\mathcal{F}\right]+E\left[\left\|\omega_i\right\|_2^{4+\delta}\mid\mathcal{F}\right]\right) \quad (c_r \text{ inequality}).\end{aligned}$$

By the same procedure of (9) of Lemma 2, we can show that the terms $E\left[\left\|\omega_i\right\|_2^{4+\delta}\mid\mathcal{F}\right]$ are uniformly bounded. Thus

$$\left\|\theta_0-\hat{\theta}\right\|_2\frac{1}{N}\sum_{i=1}^{N-1}\left\|\omega_{i+1}-\omega_i\right\|_2^4 \to^P 0$$

because $\hat{\theta} \to \theta_0$ a.s. as N tends to infinity. Similarly, we can show that the second and third terms on the right hand side of (14) tend to zero in probability as N tends to infinity. Thus we can write that as $N \to \infty$

$$\left\| \frac{1}{N} \sum_{i=1}^{N-1} \hat{g}_i \hat{g}_i ' - \frac{1}{N} \sum_{i=1}^{N-1} g_i g_i ' \right\|_2 \to^P 0 .$$

Similarly, we can show that

$$\left\| \frac{1}{N} \sum_{i=2}^{N-1} \hat{g}_i \hat{g}_{i-1} ' - \frac{1}{N} \sum_{i=2}^{N-1} g_i g_{i-1} ' \right\|_2 \to^P 0 \quad \text{and} \quad \left\| \frac{1}{N} \sum_{i=2}^{N-1} \hat{g}_{i-1} \hat{g}_i ' - \frac{1}{N} \sum_{i=2}^{N-1} g_{i-1} g_i ' \right\|_2 \to^P 0 .$$

Therefore, the result follows.

References

Aldous, D. J. and G. K. Eagleson (1978). "On mixing and stability of limit theorems." The Annals of Probability **6**(2): 325-331.

Andrews, D. W. K. (2005). "Cross-section regression with common shocks." Econometrica **73**(5): 1551-1585.

Cabrera, M. O., A. Rosalsky and A. Volodin (2012). "Some theorems on conditional mean convergence and conditional almost sure convergence for randomly weighted sums of dependent random variables." Test **21**: 369-385.

Daley, D. J. and D. Vere-Jones (1988). An introduction to the theory of point processes. New York, Springer.

Dedecker, J. and F. Merlevede (2002). "Necessary and sufficient conditions for the conditional central limit theorem." Annals of Probability **30**(3): 1044-1081.

Forchini, G., B. Jiang and B. Peng (2015). "Common Shocks in Panels with Endogenous Regressors." Mimeo.

Forchini, G. and B. Peng (2015). "A conditional approach to panel data models with common shocks." mimeo.

Kuersteiner, G. M. and I. R. Prucha (2013). "Limit theory for panel data models with cross sectional dependence and sequential exogeneity." Journal of Econometrics **174**: 107-126.

Magnus, J. R. and H. Neudecker (1999). Matrix differential calculus with applications in statistics and econometrics. Chichester, John Wiley.

Majerek, D., W. Nowak and W. Zięba (2005). "Conditional strong law of large numbers." International Journal of Pure and Applied Mathematics **20**(2): 143-156.

Pesaran, M. H. (2006). "Estimation and inference in large heterogeneous panels with a multifactor error structure." Econometrica **74**(4): 967-1012.

Prakasa Rao, B. L. S. (2009). "Conditional independence, conditional mixing and conditional association." Annals of the Institute of Statistical Mathematics **61**: 441-460.

Rényi, A. (1963). "On stable sequences of events." Sankhya: The Indian Journal of Statistics, Series A **25**(3): 293-302.

Yuan, D. M., L. R. Wei and L. Lei (2014). "Conditional central limit theorems for a sequence of conditional independent random variables." Journal of the Korean Mathematical Society **51**(1): 1-15.