Econometrica Supplementary Material

SUPPLEMENT TO "PANEL DATA MODELS WITH INTERACTIVE FIXED EFFECTS": TECHNICAL DETAILS AND PROOFS

(*Econometrica*, Vol. 77, No. 4, July 2009, 1229–1279)

BY JUSHAN BAI

This document contains additional related results and omitted proofs.

APPENDIX B: TECHNICAL DETAILS

Unbalanced Panel

THE ESTIMATION PROCEDURE can be modified to handle unbalanced data. Stock and Watson (1998) presented a method for estimating unbalanced factor models based on the expectation–maximization (EM) algorithm.[1] The procedure is extended here to models with regressors. Two sets of iterations are needed: outer iterations and inner iterations. Outer iterations are those between β and the factor model, similar to balanced panels. Inner iterations are those within the factor model associated with the EM method. For cross section i, suppose we have observations for $t = 1, 2, \dots, T_i$ (missing observations could occur at the beginning of the sample or at both ends). Suppose λ_i and F_t are observable for the moment, then the least squares estimator for β is

$$\hat{\beta} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T_i} X_{it} X_{it}' \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T_i} X_{it} (Y_{it} - \lambda_i' F_t). \tag{49}$$

Assuming β is known, let $W_{it} = Y_{it} - X_{it}'\beta$. Then $W_{it} = \lambda_i' F_t + \varepsilon_{it}$ is a pure factor model with unbalanced panel. Let $T = \max\{T_1, T_2, \dots, T_N\}$ and define $I_{it} = 1$ for observable (i, t) and $= 0$, otherwise. The EM algorithm in Stock and Watson (1998) imputes the missing values at each stage of iteration using estimates from the prior stage. More specifically, let $\hat{\lambda}_i^{(h-1)}$ and $\hat{F}_t^{(h-1)}$ $(i = 1, \dots, N; t = 1, 2, \dots, T)$ be the estimates at stage $h - 1$. Let $W_{it}^{(h)} = W_{it}$ for $I_{it} = 1$ and $= \hat{\lambda}_i^{(h-1)\prime} \cdot \hat{F}_t^{(h-1)}$ for $I_{it} = 0$ (with starting value $W_{it}^{(0)} = 0$). Finally, let $W^{(h)} = (W_{it}^{(h)})$ be the $T \times N$ matrix. The h stage estimate for $\hat{F}^{(h)}$ is the first r eigenvectors associated with the first r largest eigenvalues of the matrix $W^{(h)} W^{(h)\prime}$, subject to the constraints $\hat{F}^{(h)\prime} \hat{F}^{(h)} / T = I$ and $\hat{\Lambda}^{(h)} = T^{-1} W^{(h)\prime} \hat{F}^{(h)}$. This process continues until convergence. Let λ_i^* and F_t^* be the final stage estimates; these values are then plugged into (49) to obtain a new estimate of β (outer iteration). With the new β, we recompute $W_{it} = Y_{it} - X_{it}\beta$ for $I_{it} = 1$, readying for another round of inner iterations. Within the inner iterations, the

[1] A comprehensive description of the EM algorithm as well as its application to pure factor models can be found in the monograph by McLachlan and Krishnan (1996).

 DOI: 10.3982/ECTA6135

starting value for $W_{it}^{(0)}$ when $I_{it}=0$ is now $W_{it}^{(0)}=\lambda_i^{*\prime}F_t^*$ (instead of zero for faster convergence), where λ_i^* and F_t^* are the converged values in the previous round of inner iterations. Note that convergence for the inner iterations is not necessary. In fact, inner iterations can be reduced to a single round of computation.

PROOF OF LEMMA A.1: From $\frac{1}{NT}\sum_{i=1}^N X_i'\varepsilon_i = o_p(1)$, it is sufficient to show $\sup_F \frac{1}{NT}\sum_{i=1}^N X_i'P_F\varepsilon_i = o_p(1)$. Using $P_F = FF'/T$,

$$\frac{1}{NT}\left\|\sum_{i=1}^N X_i P_F \varepsilon_i\right\| = \left\|\frac{1}{N}\sum_{i=1}^N\left(\frac{X_i'F}{T}\right)\frac{1}{T}\sum_{t=1}^T F_t\varepsilon_{it}\right\| \\ \leq \frac{1}{N}\sum_{i=1}^N\left\|\frac{X_i'F}{T}\right\|\cdot\left\|\frac{1}{T}\sum_{t=1}^T F_t\varepsilon_{it}\right\|.$$

Note that $T^{-1}\|X_i'F\|\leq T^{-1}\|X_i\|\cdot\|F\| = \sqrt{r}T^{-1/2}\|X_i\|\leq\sqrt{r}(\frac{1}{T}\sum_{t=1}^T\|X_{it}\|^2)^{1/2}$ because $T^{-1/2}\|F\|=\sqrt{r}$. Thus, using the Cauchy–Schwarz inequality, the above is bounded by

$$\sqrt{r}\left(\frac{1}{N}\sum_{i=1}^N\frac{1}{T}\sum_{t=1}^T\|X_{it}\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T}\sum_{t=1}^T F_t\varepsilon_{it}\right\|^2\right)^{1/2}.$$

The first expression is $O_p(1)$. It suffices to show that the second term is $o_p(1)$ uniformly in F. Now

$$\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T}\sum_{t=1}^T F_t\varepsilon_{it}\right\|^2 = \operatorname{tr}\left(\frac{1}{N}\sum_{i=1}^N\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T F_tF_s'\varepsilon_{it}\varepsilon_{is}\right) \\ = \operatorname{tr}\left(\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T F_tF_s'\frac{1}{N}\sum_{i=1}^N[\varepsilon_{it}\varepsilon_{is}-E(\varepsilon_{it}\varepsilon_{is})]\right) \\ + \operatorname{tr}\left(\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T F_tF_s'\frac{1}{N}\sum_{i=1}^N\sigma_{ii,ts}\right),$$

where $\sigma_{ii,ts}=E(\varepsilon_{it}\varepsilon_{is})$. The first expression is bounded by the Cauchy–Schwarz inequality:

$$\left(\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T\|F_t\|^2\|F_s\|^2\right)^{1/2}N^{-1/2} \\ \times\left(\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T\left[\frac{1}{\sqrt{N}}\sum_{i=1}^N[\varepsilon_{it}\varepsilon_{is}-E(\varepsilon_{it}\varepsilon_{is})]\right]^2\right)^{1/2}.$$

But $T^{-1}\sum_{t=1}^{T}\|F_t\|^2 = \|F'F/T\| = r$. Thus the above expression is equal to $rN^{-1/2}O_p(1)$. Next, $|\frac{1}{N}\sum_{i=1}^{N}\sigma_{ii,ts}| \le \tau_{ts}$ by Assumption C(ii). Again by the Cauchy–Schwarz inequality,

$$\begin{aligned}\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}F_tF_s'\frac{1}{N}\sum_{i=1}^{N}\sigma_{ii,ts}\right\| &\le \left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\|F_t\|^2\|F_s\|^2\right)^{1/2}\\ &\quad\times\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{ts}^2\right)^{1/2}\\ &= rT^{-1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{ts}^2\right)^{1/2}\\ &= rO(T^{-1/2}),\end{aligned}$$

where the last equality follows from $\tau_{ts}^2 \le M\tau_{ts}$ and Assumption C(ii). The proof for the remaining statements are the same, so are omitted. Note that we do not need bounded support for F_t and our optimization with respect to F_t does not need to be taken over bounded set. *Q.E.D.*

PROOF OF LEMMA A.2: Denote the term inside $\|\cdot\|^2$ as A. Then the left-hand side is equal to $E\,\mathrm{tr}(AA')$. Using $E\|F_t\|^4 \le M$ and Assumption C(iv), (i) follows readily. The proof of (ii) is similar. *Q.E.D.*

PROOF OF LEMMA A.3: (i) This part extends Lemma B.2 of Bai (2003). Using (42), it is easy to see that the first five terms are each $O_p(\hat\beta - \beta)$. In fact, the first, third, and fifth terms are $o_p(\hat\beta - \beta)$; the second and fourth terms are $O_p(\hat\beta - \beta)$. The next three terms are considered in Bai (2003) and each is shown to be $O_p(\delta_{NT}^{-2})$ in the absence of β. With the estimation of β, they are each shown to be $O_p(\hat\beta - \beta)O_p(\delta_{NT}^{-1}) + O_p(\delta_{NT}^{-2})$ due to Proposition A.1(ii) instead of Lemma A.1 of Bai (2003). But $O_p(\hat\beta - \beta)O_p(\delta_{NT}^{-1})$ is dominated by $O_p(\hat\beta - \beta)$, the order of the first five terms. Thus summing over the eight terms, we obtain part (i).

For part (ii),

$$\begin{aligned}\|T^{-1}\hat F'(\hat F - F^0H)\| &\le T^{-1}\|\hat F - F^0H\|^2 + \|H\|T^{-1}\|F^{0\prime}(\hat F - F^0H)\|\\ &= O_p(\hat\beta - \beta) + O_p(\delta_{NT}^{-2})\end{aligned}$$

by part (i) and Proposition A.1(ii). The proof of part (iii) is identical to part (i).

For (iv),

$$\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}}(\hat{F}-F^0H) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}X_i'(\hat{F}-F^0H) + \frac{1}{N}\sum_{i=1}^{N}\frac{X_i'\hat{F}}{T}\hat{F}'(\hat{F}-F^0H).$$

The first term on the right is an average of (iii) over i and thus is still that order of magnitude. The second term is bounded by $\frac{1}{N}\sum_{i=1}^{N}\|X_i/\sqrt{T}\|^2\sqrt{r}\|T^{-1}\hat{F}'(\hat{F}-F^0H)\| = O_p(1)\|T^{-1}\hat{F}'(\hat{F}-F^0H)\|$. Thus (iv) follows from part (ii). *Q.E.D.*

PROOF OF LEMMA A.4: Part (i) extends Lemma B.1 of Bai (2003). The proof is omitted as it is easier than the proof of part (ii) (a proof can be found in the working version). Now consider the proof of (ii). From (42) and denoting $G = (F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}$ for the moment,

$$T^{-1}N^{-1/2}\sum_{k=1}^{N}\varepsilon_k'(\hat{F}H^{-1}-F^0) = T^{-1}N^{-1/2}\sum_{k=1}^{N}\varepsilon_k'(I1+\cdots+I8)G = a1+\cdots+a8.$$

We show that the first four terms are each $T^{-1/2}O_p(\hat{\beta}-\beta)$:

$$\|a1\| \le T^{-1/2}\|G\|\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\left(\frac{1}{\sqrt{TN}}\sum_{k=1}^{N}\sum_{t=1}^{T}\varepsilon_{kt}X_{it}\right)\right\|\left(\frac{\|X_i\|^2}{T}\right)\right) \times \|\hat{\beta}-\beta\|^2 = T^{-1/2}\|\hat{\beta}-\beta)\|^2 O_p(1),$$

$$a2 = \frac{1}{NT}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\sum_{i=1}^{N}\varepsilon_k' X_i(\beta-\hat{\beta})\lambda_i\left(\frac{\Lambda'\Lambda}{N}\right)^{-1} = \frac{1}{\sqrt{T}}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{t=1}^{T}X_{it}\varepsilon_{kt}(\hat{\beta}-\beta)\lambda_i\left(\frac{\Lambda'\Lambda}{N}\right)^{-1} = T^{-1/2}O_p(\hat{\beta}-\beta),$$

$$\|a3\| \le T^{-1/2}\|G\|\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{t=1}^{T}\varepsilon_{kt}X_{it}\right\|\left(\frac{\|\varepsilon_i\|^2}{T}\right)\right)\|\hat{\beta}-\beta\| = T^{-1/2}O_p(\|\hat{\beta}-\beta\|),$$

$$\begin{aligned}a4 &= T^{-1/2}\left(\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{t=1}^{T}\varepsilon_{kt}F_t'\right)(\beta-\hat{\beta})'\left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i'\hat{F}}{T}\right)\right)G\\&= T^{-1/2}O_p(\hat{\beta}-\beta).\end{aligned}$$

For $a5$, let $W_i = X_i'\hat{F}/T$ and note that $\|W_i\|^2 \le \|X_i\|^2/T$:

$$\begin{aligned}a5 &= \frac{1}{NT}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\sum_{i=1}^{N}\varepsilon_k'\varepsilon_i(\beta-\hat{\beta})'W_iG\\&= \frac{1}{\sqrt{N}}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\varepsilon_{kt}\right)\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_{it}(\hat{\beta}-\beta)W_i\right)G\\&= N^{-1/2}O_p(\hat{\beta}-\beta).\end{aligned}$$

For $a6$,

$$\begin{aligned}a6 &= \frac{1}{NT^2}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\varepsilon_k'F^0\sum_{i=1}^{N}\lambda_i\varepsilon_i'\hat{F}G\\&= \frac{1}{NT^2}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\varepsilon_k'F^0\sum_{i=1}^{N}\lambda_i\varepsilon_i'F^0HG\\&\quad+\frac{1}{NT^2}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\varepsilon_k'F^0\sum_{i=1}^{N}\lambda_i\varepsilon_i'(\hat{F}-F^0H)G\\&= a6.1 + a6.2,\\a6.1 &= \frac{1}{\sqrt{N}T}\left(\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{t=1}^{T}F_t^{0\prime}\varepsilon_{kt}\right)\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_iF_t^{0\prime}\varepsilon_{it}\right)HG\\&\quad= O_p\big(T^{-1}N^{-1/2}\big),\\a6.2 &= T^{-1/2}\left(\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{t=1}^{T}F_t^{0\prime}\varepsilon_{kt}\right)\frac{1}{TN}\sum_{i=1}^{N}\lambda_i\varepsilon_i'(\hat{F}-F^0H)G,\\\|a6.2\| &\le T^{-1/2}O_p(1)\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|\left\|\frac{\varepsilon_i}{\sqrt{T}}\right\|\frac{\|\hat{F}-F^0H\|}{\sqrt{T}}\|G\|\\&= T^{-1/2}[O_p(\hat{\beta}-\beta)+O_p(\delta_{NT}^{-1})]\\&= T^{-1/2}(\hat{\beta}-\beta)+O_p(\delta_{NT}^{-2}).\end{aligned}$$

Next consider $a7$ and $a8$:

$$\begin{aligned}
a7 &= \frac{1}{NT}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\sum_{i=1}^{N}\varepsilon_k'\varepsilon_i\lambda_i'\left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \\
&= N^{-1/2}\frac{1}{T}\sum_{t=1}^{T}\left[\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\varepsilon_{kt}\right)\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_{it}\lambda_i'\right)\right]\left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \\
&= O_p\big(N^{-1/2}\big), \\
a8 &= \frac{1}{NT^2}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\sum_{i=1}^{N}\varepsilon_k'\varepsilon_i(\varepsilon_i'\hat{F})G \\
&= \frac{1}{NT^2}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\sum_{i=1}^{N}\varepsilon_k'\varepsilon_i(\varepsilon_i'F^0)HG \\
&\quad + \frac{1}{NT^2}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\sum_{i=1}^{N}\varepsilon_k'\varepsilon_i(\varepsilon_i'(\hat{F}-F^0H))G \\
&= b8 + c8, \\
b8 &= \frac{1}{NT}\sum_{i=1}^{N}\left[\left(\frac{1}{\sqrt{TN}}\sum_{k=1}^{N}\sum_{t=1}^{T}(\varepsilon_{kt}\varepsilon_{it} - E(\varepsilon_{kt}\varepsilon_{it}))\right)\right. \\
&\quad \times \left.\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\varepsilon_{is}F_s^0H\right)\right]G \\
&\quad + \frac{1}{\sqrt{NT}}\frac{1}{NT}\sum_{k=1}^{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\gamma_{ki,t}\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\varepsilon_{is}F_s^0HG \\
&= O_p(T^{-1}) + O_p\big((NT)^{-1/2}\big).
\end{aligned}$$

Ignoring G,

$$\begin{aligned}
c8 &= T^{-1/2}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sqrt{TN}}\sum_{k=1}^{N}\sum_{t=1}^{T}[\varepsilon_{kt}\varepsilon_{it} - E(\varepsilon_{kt}\varepsilon_{it})]\frac{\varepsilon_i'(\hat{F}-F^0H)}{T} \\
&\quad + \frac{1}{N^{3/2}T}\sum_{k=1}^{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\gamma_{ki,t}\frac{\varepsilon_i'(\hat{F}-F^0H)}{T} \\
&= c8.1 + c8.2,
\end{aligned}$$

$$\|c8.1\| \le T^{-1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{\sqrt{TN}}\sum_{k=1}^{N}\sum_{t=1}^{T}[\varepsilon_{kt}\varepsilon_{it}-E(\varepsilon_{kt}\varepsilon_{it})]\right]^2\right)^{1/2}$$
$$\times\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\|\varepsilon_i\|^2}{T}\right)^{1/2}\frac{\|\hat{F}-F^0H\|}{\sqrt{T}}$$
$$= T^{-1/2}O_p(\|\hat{\beta}-\beta\|)+T^{-1/2}O_p(\delta_{NT}^{-1})$$
$$= T^{-1/2}O_p(\|\hat{\beta}-\beta\|)+O_p(\delta_{NT}^{-2}),$$
$$\|c8.2\| \le \frac{1}{\sqrt{N}}\frac{\|\hat{F}-F^0H\|}{\sqrt{T}}\frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{N}|\gamma_{ki}|\frac{\|\varepsilon_i\|}{\sqrt{T}}$$
$$= \left[O_p(\|\hat{\beta}-\beta\|)+O_p(\delta_{NT}^{-1})\right]N^{-1/2}$$
$$= N^{-1/2}O_p(\|\hat{\beta}-\beta\|)+O_p(\delta_{NT}^{-2}).$$

Note that

$$EN^{-1}\sum_{k=1}^{N}\sum_{i=1}^{N}|\gamma_{ki}|\frac{\|\varepsilon_i\|}{\sqrt{T}}$$
$$\le \max_i E\left(\frac{\|\varepsilon_i\|}{\sqrt{T}}\right)N^{-1}\sum_{k=1}^{N}\sum_{i=1}^{N}|\gamma_{ki}| = O(1).$$

Part (iii) is derived from (ii) with division by $\sqrt{N}$. The presence of λ_k does not alter the results. A direct proof would be similar to that of (ii). The details are omitted.

Part (iv) is the same as (iii) with λ_k replaced by $(X_k'F^0/T)(F^{0\prime}F^0/T) = O_p(1)$. The first term on the right is an elaboration of the corresponding $O_p(N^{-1})$ term appearing in (iii). This elaborated expression will be used later. *Q.E.D.*

PROOF OF LEMMA A.5: Rewrite the left-hand side as

$$\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{k=1}^{N}X_i'(\varepsilon_k\varepsilon_k'-\Omega_k)\hat{F}G\lambda_i$$
$$-\frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i'\hat{F}}{T}\right)\frac{1}{NT^2}\sum_{k=1}^{N}\hat{F}'(\varepsilon_k\varepsilon_k'-\Omega_k)\hat{F}G\lambda_i$$
$$= I + II.$$

Adding and subtracting terms yields

$$I = \frac{1}{N^2T^2}\sum_{k=1}^{N}\sum_{i=1}^{N} X_i'(\varepsilon_k\varepsilon_k' - \Omega_k)F^0HG\lambda_i$$
$$+ \frac{1}{N^2T^2}\sum_{k=1}^{N}\sum_{i=1}^{N} X_i'(\varepsilon_k\varepsilon_k' - \Omega_k)(\hat{F} - F^0H)G\lambda_i.$$

The first term on the right is equal to

$$\left(\frac{1}{N^2T^2}\right)\sum_{i=1}^{N}\sum_{k=1}^{N}\left\{\sum_{t=1}^{T}\sum_{s=1}^{T} X_{it}[\varepsilon_{kt}\varepsilon_{ks} - E(\varepsilon_{kt}\varepsilon_{ks})]F_s^{0\prime}HG\lambda_i\right\}$$
$$= \frac{1}{T\sqrt{N}}\frac{1}{N}\sum_{i=1}^{N}\left[N^{-1/2}\sum_{k=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} X_{it}[\varepsilon_{kt}\varepsilon_{ks} - E(\varepsilon_{kt}\varepsilon_{ks})]F_s^{0\prime}\right]$$
$$\times HG\lambda_i$$
$$= O_p\left(\frac{1}{T\sqrt{N}}\right)$$

by Lemma A.2(ii). Denote

$$a_s = \left(\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{t=1}^{T} X_{it}[\varepsilon_{kt}\varepsilon_{ks} - E(\varepsilon_{kt}\varepsilon_{ks})]\right) = O_p(1).$$

Then the second term of I is

$$\frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{s=1}^{T} a_s(\hat{F}_s - F_s^0H)'G\lambda_i.$$

Notice

$$\left\|\frac{1}{T}\sum_{s=1}^{T} a_s(\hat{F}_s - F_s^0H)\right\| \le \left(\frac{1}{T}\sum_{s=1}^{T}\|a_s\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\|\hat{F}_s - F_s^0H\|^2\right)^{1/2}$$
$$= O_p(\hat{\beta} - \beta) + O_p(\delta_{NT}^{-1}).$$

Thus the second term of I is $(NT)^{-1/2}[O_p(\hat{\beta} - \beta) + O_p(\delta_{NT}^{-1})]$. Consider II:

$$\|II\| \le \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{X_i\hat{F}}{T}\right\|\|G\lambda_i\| \cdot \left\|\frac{1}{NT^2}\sum_{k=1}^{N}\hat{F}'(\varepsilon_k\varepsilon_k' - \Omega_k)\hat{F}\right\|$$

$$= O_p(1)\left\| \frac{1}{NT^2}\sum_{k=1}^{N}\hat{F}'(\varepsilon_k\varepsilon_k' - \Omega_k)\hat{F}\right\|.$$

But

$$\frac{1}{NT^2}\sum_{k=1}^{N}\hat{F}'(\varepsilon_k\varepsilon_k' - \Omega_k)\hat{F}$$

$$= H\frac{1}{NT^2}\sum_{k=1}^{N}F^{0\prime}(\varepsilon_k\varepsilon_k' - \Omega_k)F^0H$$

$$+ H\frac{1}{NT^2}\sum_{k=1}^{N}F^{0\prime}(\varepsilon_k\varepsilon_k' - \Omega_k)(\hat{F} - F^0H)$$

$$+ \frac{1}{NT^2}\sum_{k=1}^{N}(\hat{F} - F^0H)'(\varepsilon_k\varepsilon_k' - \Omega_k)F^0H$$

$$+ \frac{1}{NT^2}\sum_{k=1}^{N}(\hat{F} - F^0H)'(\varepsilon_k\varepsilon_k' - \Omega_k)(\hat{F} - F^0H)$$

$$= b1 + b2 + b3 + b4.$$

Now

$$b1 = H\left(\frac{1}{T^2N}\right)\sum_{k=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}F_sF_t'[\varepsilon_{kt}\varepsilon_{ks} - E(\varepsilon_{kt}\varepsilon_{ks})]H = O_p\left(\frac{1}{T\sqrt{N}}\right)$$

by Lemma A.2(i). Next

$$b2 = H\frac{1}{\sqrt{NT}}\frac{1}{T}\sum_{s=1}^{T}\left[\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{k=1}^{N}F_t^0[\varepsilon_{kt}\varepsilon_{ks} - E(\varepsilon_{kt}\varepsilon_{ks})]\right]$$
$$\times (\hat{F}_s - H'F_s^0).$$

Thus if we let $A_s = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{k=1}^{N}F_t^0[\varepsilon_{kt}\varepsilon_{ks} - E(\varepsilon_{kt}\varepsilon_{ks})]$, then

$$\|b2\| \le \|H\|\frac{1}{\sqrt{NT}}\left(\frac{1}{T}\sum_{s=1}^{T}\|A_s\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\|\hat{F}_s - H'F_s^0\|^2\right)^{1/2}$$

$$= \frac{1}{\sqrt{NT}}\big[O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-1})\big].$$

The term $b3$ has the same upper bound because it is the transpose of $b2$. The last term is

$$b4 = \frac{1}{\sqrt{N}}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{F}_t - H'F_t^0)(\hat{F}_s - H'F_s^0)' \times \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\varepsilon_{kt}\varepsilon_{ks} - E(\varepsilon_{kt}\varepsilon_{ks})]\right].$$

Thus by the Cauchy–Schwarz inequality,

$$\|b4\| \le \frac{1}{\sqrt{N}}\left(\frac{1}{T}\sum_{t=1}^{T}\|F_t - H'F_t^0\|^2\right) \times \left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\varepsilon_{kt}\varepsilon_{ks} - E(\varepsilon_{kt}\varepsilon_{ks})]\right]^2\right)^{1/2}$$
$$= \frac{1}{\sqrt{N}}O_p(\|\hat{\beta} - \beta\|^2) + \frac{1}{\sqrt{N}}O_p(\delta_{NT}^{-2}).$$

Now collecting terms yields the lemma. *Q.E.D.*

PROOF OF LEMMA A.6: First note that $\|(F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda/N)^{-1}\| = O_p(1)$. Next, $\|X_i' M_{\hat{F}}\Omega\hat{F}\| \le \|X_i' M_{\hat{F}}\|\|\Omega\hat{F}\|$, $\|X_i' M_{\hat{F}}\| \le \|X_i\|$, and $\|\Omega\hat{F}\| \le \lambda_{\max}(\Omega) \times \|\hat{F}\| = \lambda_{\max}(\Omega)\sqrt{rT}$, where $\lambda_{\max}(\Omega)$ is the largest eigenvalue of Ω and is bounded by assumption. The lemma follows from $\frac{1}{N}\sum_{i=1}^{N}(\|X_i\|/\sqrt{T})\|\lambda_i\| = O_p(1)$. *Q.E.D.*

PROOF OF LEMMA A.7: (i) The first two results of Lemma A.3 can be rewritten as

$$F^{0\prime}\hat{F}/T - (F^{0\prime}F^0/T)H = O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-2})$$

and

$$I - (\hat{F}'F^0/T)H = O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-2}).$$

Left multiply the first equation by H' and use the transpose of the second equation to obtain

$$I - H'(F^{0\prime}F^0/T)H = O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-2}).$$

Right multiplying by H' and left multiplying by H'^{-1}, we obtain

$$I - (F^{0\prime}F^0/T)HH' = O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-2}).$$

This is equivalent to (i).

(ii) We have

$$\|P_{\hat{F}} - P_{F^0}\|^2 = \operatorname{tr}[(P_{\hat{F}} - P_{F^0})^2] = 2\operatorname{tr}(I_r - \hat{F}'P_{F^0}\hat{F}/T).$$

Proposition 1(ii) already implies $I_r - \hat{F}'P_{F^0}\hat{F}/T = o_p(1)$. By rewriting $T^{-1}\hat{F}' \times F^0 = T^{-1}\hat{F}(F^0 - \hat{F}H^{-1}) + H$, we can easily show, using earlier lemmas that $I_r - \hat{F}'P_{F^0}\hat{F}/T = O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-2})$. The details are omitted. *Q.E.D.*

PROOF OF LEMMA A.8: First consider $\frac{1}{\sqrt{NT}}\sum_{i=1}^N X_i(M_{F^0} - M_{\hat{F}})\varepsilon_i$. Note that $M_{F^0} - M_{\hat{F}} = P_{\hat{F}} - P_{F^0}$ and $P_{\hat{F}} = \hat{F}\hat{F}'/T$. By adding and subtracting terms,

$$\begin{aligned}
&\frac{1}{\sqrt{NT}}\sum_{i=1}^N \frac{X_i'\hat{F}}{T}\hat{F}'\varepsilon_i - \frac{1}{\sqrt{NT}}\sum_{i=1}^N X_i'P_{F^0}\varepsilon_i \\
&\quad = \frac{1}{\sqrt{NT}}\sum_{i=1}^N \frac{X_i'(\hat{F} - F^0H)}{T}H'F^{0\prime}\varepsilon_i \\
&\qquad + \frac{1}{\sqrt{NT}}\sum_{i=1}^N \frac{X_i'(\hat{F} - F^0H)}{T}(\hat{F} - F^0H)'\varepsilon_i \\
&\qquad + \frac{1}{\sqrt{NT}}\sum_{i=1}^N \frac{X_i'F^0H}{T}(\hat{F} - F^0H)'\varepsilon_i \\
&\qquad + \frac{1}{\sqrt{NT}}\sum_{i=1}^N \frac{X_i'F^0}{T}\left[HH' - \left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right]F^{0\prime}\varepsilon_i \\
&\quad = a + b + c + d.
\end{aligned}$$

Consider a. Note that $(\hat{F}_s - H'F_s^0)'H'F_t^0$ is scalar and thus commutable with X_{it}:

$$a = \frac{1}{T}\sum_{s=1}^T(\hat{F}_s - H'F_s^0)'H'\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T F_t^0 X_{is}\varepsilon_{it}\right).$$

Thus

$$\begin{aligned}
\|a\| &\le \left[\frac{1}{T}\sum_{s=1}^T \|\hat{F}_s - H'F_s^0\|^2\right]^{1/2}\|H\| \\
&\quad \times \left[\frac{1}{T}\sum_{s=1}^T \left\|\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T F_t^0 X_{is}\varepsilon_{it}\right)\right\|^2\right]^{1/2} \\
&= \left[O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-1})\right]O_p(1) = o_p(1).
\end{aligned}$$

Similarly,

$$b = T^{1/2}\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}(\hat{F}_s - H'F_s^0)'(\hat{F}_t - H'F_t^0)\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_{is}\varepsilon_{it}\right)$$

and

$$\begin{aligned}\|b\| &\le \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t - H'F_t^0\|^2\right)\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_{it}\varepsilon_{it}\right\|^2\right)^{1/2}\\ &= \sqrt{T}\big[O_p(\|\hat{\beta}-\beta\|^2) + O_p(\delta_{NT}^{-2})\big]O_p(1).\end{aligned}$$

Consider c:

$$\begin{aligned}c &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{X_i'F^0}{T}HH'(\hat{F}H^{-1} - F^0)'\varepsilon_i\\ &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{X_i'F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}(\hat{F}H^{-1} - F^0)'\varepsilon_i\\ &\quad + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{X_i'F^0}{T}\left[HH' - \left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right](\hat{F}H^{-1} - F^0)'\varepsilon_i\\ &= c1 + c2.\end{aligned}$$

Denote $Q = HH' - (F^{0\prime}F^0/T)^{-1}$ for the moment. We show $c2 = o_p(1)$, that is,

$$\begin{aligned}c2 &= \sqrt{NT}\left(\frac{1}{NT}\sum_{i=1}^{N}\left[\varepsilon_i'(\hat{F}H^{-1} - F^0)\otimes\left(\frac{X_i'F^0}{T}\right)\right]\right)\text{vec}(Q)\\ &= \sqrt{NT}\big[(NT)^{-1/2}(\|\hat{\beta}-\beta\|) + O_p(N^{-1}) + N^{-1/2}O_p(\delta_{NT}^{-2})\big]\text{vec}(Q)\end{aligned}$$

by the argument of Lemma A.4(iii) and (iv). By Lemma A.7, $\text{vec}(Q) = O_p(\|\hat{\beta}-\beta\|) + O_p(\delta_{NT}^{-2})$. Thus $c2 = O_p(\hat{\beta}-\beta) + \sqrt{T/N}O_p(\delta_{NT}^{-2}) + \sqrt{T}\times O_p(\delta_{NT}^{-4}) \xrightarrow{p} 0$ if $T/N^3 \to 0$.

By Lemma A.4(iv), switching the role of i and k, we get

$$c1 = (\sqrt{NT}/N)\psi_{NT} + O_p(\hat{\beta}-\beta) + \sqrt{T}O_p(\delta_{NT}^{-2}),$$

where

$$\psi_{NT} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\frac{X_i'F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_k\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{kt}\right), \tag{50}$$

which is $O_p(1)$. To see this, let $A_i = (X_i'F_0/T)(F^{0\prime}F^0/T)^{-1}$ and $B_k = (\Lambda'\Lambda/N)^{-1}\lambda_k$. Then

$$\psi_{NT} = \frac{1}{T}\sum_{t=1}^{T}\left(N^{-1/2}\sum_{i=1}^{N} A_i\varepsilon_{it}\right)\left(N^{-1/2}\sum_{k=1}^{N} B_k\varepsilon_{kt}\right) = O_p(1).$$

For d, again let $Q = HH' - (F^{0\prime}F^0/T)^{-1}$. Then

$$d = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[\varepsilon_i'F^0 \otimes \left(\frac{X_i'F^0}{T}\right)\right]\operatorname{vec}(Q) \tag{51}$$

$$= \left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} F_t^0\varepsilon_{it} \otimes \left(\frac{X_i'F^0}{T}\right)\right)\operatorname{vec}(Q) = O_p(1)\operatorname{vec}(Q), \tag{52}$$

which is $o_p(1)$ because $\operatorname{vec}(Q) = O_p(\|\hat\beta - \beta\|) + O_p(\delta_{NT}^{-2})$ by Lemma A.7. In summary, ignore dominated terms:

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_i'(M_{F^0} - M_{\hat F})\varepsilon_i = \left(\frac{\sqrt{NT}}{N}\right)\psi_{NT} + \sqrt{T}O_p(\|\hat\beta - \beta^0\|^2) + O_p(\hat\beta - \beta) + \sqrt{T}O_p(\delta_{NT}^{-2}). \tag{53}$$

Let $V_i = \frac{1}{N}\sum_{k=1}^{N} a_{ik}X_k$. Then replacing X_i with V_i, the same argument leads to

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} V_i'(M_{F^0} - M_{\hat F})\varepsilon_i = \left(\frac{\sqrt{NT}}{N}\right)\psi_{NT}^* + \sqrt{T}O_p(\|\hat\beta - \beta^0\|^2) + O_p(\hat\beta - \beta) + \sqrt{T}O_p(\delta_{NT}^{-2}), \tag{54}$$

where $\psi_{NT}^* = O_p(1)$ is defined as

$$\psi_{NT}^* = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\frac{V_i'F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_k\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{kt}\right), \tag{55}$$

which is also $O_p(1)$ for the same reasoning as for ψ_{NT}. Combining (53) and (54), and defining $\xi_{NT}^\dagger = \psi_{NT} - \psi_{NT}^*$, we obtain the lemma:

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[X_i'M_{\hat F} - \frac{1}{N}\sum_{k=1}^{N} a_{ik}X_k'M_{\hat F}\right]\varepsilon_i$$

$$= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[X_i'M_{F^0} - \frac{1}{N}\sum_{k=1}^{N} a_{ik}X_k'M_{F^0}\right]\varepsilon_i$$

$$-\left(\frac{\sqrt{NT}}{N}\right)(\psi_{NT}-\psi^*_{NT})+\sqrt{T}O_p(\|\hat\beta-\beta^0\|^2)$$
$$+O_p(\|\hat\beta-\beta^0\|)+\sqrt{T}O_p(\delta_{NT}^{-2}). \qquad Q.E.D.$$

PROOF OF LEMMA A.9: (i) We have

$$D(\hat F)-D(F^0)=\frac{1}{NT}\sum_{i=1}^N X_i'(M_{\hat F}-M_{F^0})X_i - \frac{1}{T}\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N X_i'(M_{\hat F}-M_{F^0})X_k a_{ik}\right].$$

The norm of the first term on the right is bounded by

$$\left\|\frac{1}{NT}\sum_{i=1}^N X_i'(P_{\hat F}-P_{F^0})X_i\right\|\le\frac{1}{N}\sum_{i=1}^N\left(\frac{\|X_i\|^2}{T}\right)\|P_{\hat F}-P_{F^0}\|=o_p(1)$$

by Lemma A.7(ii). The proof that the second term is $o_p(1)$ is the same.

(ii) By Lemma A.6, $\zeta_{NT}=O_p(1)$, and by Lemma A.8, $\xi^\dagger_{NT}=O_p(1)$ so that $\xi_{NT}=D(\hat F)^{-1}\xi^\dagger_{NT}=O_p(1)$. Thus Corollary 1 implies that $\sqrt{NT}(\hat\beta-\beta)=O_p(\sqrt{N/T})+O_p(\sqrt{T/N})$. That is, $\hat\beta-\beta=O_p(\frac{1}{N})+O_p(\frac{1}{T})=O_p(\delta_{NT}^{-2})$. The proof for (i) implies that $\|D(\hat F)-D(F^0)\|\le O_p(1)\|P_{\hat F}-P_{F^0}\|$. But by Lemma A.7(ii), $\|P_{\hat F}-P_{F^0}\|=O_p(\|\hat\beta-\beta\|^{1/2})+O_p(\delta_{NT}^{-1})=O_p(\delta_{NT}^{-1})$. In summary, $\|D(\hat F)-D(F^0)\|\le O_p(\delta_{NT}^{-1})$. It follows that $\sqrt{T/N}\|D(\hat F)-D(F^0)\|\le\sqrt{T/N}O_p(\delta_{NT}^{-1})=o_p(1)$ if $T/N^2\to 0$. The proof of (iii) is the same.

(iv) From $\xi_{NT}=D(\hat F)^{-1}\xi^\dagger_{NT}$, $\sqrt{T/N}[D(\hat F)-D(F^0)]=o_p(1)$ by part (ii), and $\xi^\dagger_{NT}=O_p(1)$, it suffices to show $\sqrt{T/N}[D(F^0)^{-1}\xi^\dagger_{NT}-B]=o_p(1)$. Let $A_{ik}=\lambda_k'\otimes[(X_i-V_i)'F^0/T]$ and $G=(F^{0\prime}F^0/T)^{-1}(\Lambda'\Lambda/N)^{-1}$. Then $E\|A_{ik}\|^2\le M$ and $\|G\|=O_p(1)$. We have

$$D(F^0)\xi^\dagger_{NT}-B = -D(F^0)^{-1}\frac{1}{N}\sum_{i=1}^N\sum_{k=1}^N\left[A_{ik}\frac{1}{T}\sum_{t=1}^T(\varepsilon_{it}\varepsilon_{kt}-\sigma_{ik,tt})\right]\mathrm{vec}(G).$$

Assumption C(iv) implies that the above is $O_p(T^{-1/2})$. Thus $\sqrt{T/N}[D(F^0)^{-1}\times\xi^\dagger_{NT}-B]=O_p(N^{-1/2})=o_p(1)$.

(v) Comparing ζ_{NT} and C, and in view of (iii), it suffices to show that the difference

$$\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}} \Omega \hat{F}\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i$$
$$-\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{F^0} \Omega F^0\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i$$

multiplied by $\sqrt{N/T}$ is $o_p(1)$. Using $ab - cd = a(b-c) + (a-c)b$, by adding and subtracting terms, we first consider

$$\frac{1}{NT}\sum_{i=1}^{N} X_i (M_{\hat{F}} - M_{F^0})\Omega \hat{F} G_i,$$

where $G_i = (F^{0\prime}\hat{F}/T)^{-1}(\Lambda'\Lambda)^{-1}\lambda_i$. Note that $\|\Omega\hat{F}\| \le \lambda_{\max}(\Omega)\|\hat{F}\| = O(\sqrt{T})$, where $\lambda_{\max}(\Omega)$ is the largest eigenvalue of Ω and is $O(1)$. Further, $\|X_i(M_{\hat{F}} - M_{F^0})\| \le \|X_i\|\|P_{\hat{F}} - P_{F^0}\| = \|X_i\| O_p(\delta_{NT}^{-1})$. Thus

$$\left\|\frac{1}{NT}\sum_{i=1}^{N} X_i (M_{\hat{F}} - M_{F^0})\Omega \hat{F} G_i\right\|$$
$$\le O(1)\|P_{\hat{F}} - P_{F^0}\|\frac{1}{N}\sum_{i=1}^{N}\big(T^{-1/2}\|X_i\|\big)\|G_i\| = O_p(\delta_{NT}^{-1}).$$

But $\sqrt{N/T}O_p(\delta_{NT}^{-1}) = o_p(1)$ if $N/T^2 \to 0$. Next consider

$$\frac{1}{NT}\sum_{i=1}^{N} X_i M_{F^0}\Omega\left[\hat{F}\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1} - F^0\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right]\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i.$$

From $\|X_i M_{F^0}\Omega\| \le \lambda_{\max}(\Omega)\|X_i M_{F^0}\| \le \lambda_{\max}(\Omega)\|X_i\|$ and $\frac{1}{N}\sum_{i=1}^{N} T^{-1/2}\|X_i\| \times \|\Lambda'\Lambda/N\lambda_i\| = O_p(1)$, it suffices to show

$$\begin{aligned} &T^{-1/2}[\hat{F}(F^{0\prime}\hat{F}/T)^{-1} - F^0(F^{0\prime}F^0/T)^{-1}] \\ &\quad = T^{-1/2}(\hat{F} - P_{F^0}\hat{F})(F^{0\prime}\hat{F}/T)^{-1} \\ &\quad = O_p(\delta_{NT}^{-1}). \end{aligned}$$

But $T^{-1/2}\|\hat{F} - P_{F^0}\hat{F}\| = \|P_{\hat{F}} - P_{F^0}\| = O_p(\delta_{NT}^{-1})$, proving the lemma. *Q.E.D.*

PROOF OF LEMMA A.10: By definition, $\hat{\Lambda}' = \frac{1}{T}\hat{F}'(Y - X\hat{\beta})$, where $Y = (Y_1, \dots, Y_N)$ is $T \times N$ and X is $T \times N \times p$ (three-dimensional matrix), so that

$X\hat{\beta}$ is $T \times N$ (readers may consider β is a scalar so that X is simply $T \times N$). Thus from $Y - X\hat{\beta} = F^0 \Lambda' + \varepsilon - X(\hat{\beta} - \beta^0)$,

$$\hat{\Lambda}' = T^{-1}\hat{F}'F^0\Lambda' + T^{-1}\hat{F}'\varepsilon - T^{-1}\hat{F}'X(\hat{\beta} - \beta^0).$$

From $F^0 = F^0 - \hat{F}H^{-1} + \hat{F}H^{-1}$ and using $\hat{F}'\hat{F}/T = I$, we have

$$\hat{\Lambda}' - H^{-1}\Lambda' = T^{-1}\hat{F}'(F^0 - \hat{F}H^{-1})\Lambda' + T^{-1}\hat{F}'\varepsilon - T^{-1}\hat{F}'X(\hat{\beta} - \beta). \tag{56}$$

Thus

$$N^{-1/2}\|\hat{\Lambda}' - H^{-1}\Lambda'\| \le \sqrt{r}\frac{\|\hat{F}^0 - \hat{F}H^{-1}\|}{\sqrt{T}}\frac{\|\Lambda\|}{\sqrt{N}} + T^{-1/2}\left\|\frac{1}{\sqrt{NT}}\hat{F}'\varepsilon\right\| + \sqrt{r}\frac{\|X\|}{\sqrt{NT}}\|\hat{\beta} - \beta\|.$$

The first term is $O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-1})$ by Proposition A.1(ii); the second term is $O_p(T^{-1/2})$; the third term is $O_p(\|\hat{\beta} - \beta\|)$ in view $\|X\|/\sqrt{NT} = O_p(1)$. Thus $N^{-1/2}\|\hat{\Lambda}' - H^{-1}\Lambda'\| = O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-1})$. This is equivalent to (i).

For (ii), left multiplying Λ on each side and then dividing by N gives

$$N^{-1}(\hat{\Lambda}' - H^{-1}\Lambda')\Lambda = T^{-1}\hat{F}'(F^0 - \hat{F}H^{-1})(\Lambda'\Lambda/N) + (TN)^{-1}\hat{F}'\varepsilon\Lambda - (TN)^{-1}\hat{F}'X(\hat{\beta} - \beta)\Lambda.$$

The first term on the right is $O_p(\|\hat{\beta} - \beta\|) + O_p(\delta_{NT}^{-2})$ by Lemma A.3. The second term is

$$(TN)^{-1}(\hat{F} - F^0H)'\varepsilon\Lambda + (TN)^{-1}HF^{0\prime}\varepsilon\Lambda = a + b.$$

But a is the left-hand side of Lemma A.4(iii), thus having the desired result. Term b is simply $(TN)^{-1/2}O_p(1)$, also as desired. Finally,

$$\|(TN)^{-1}\hat{F}'X(\hat{\beta} - \beta)\Lambda\| \le \sqrt{r}\|X/\sqrt{TN}\| \cdot \|\Lambda/\sqrt{N}\| \cdot \|\hat{\beta} - \beta\| = O_p(\|\hat{\beta} - \beta\|)$$

proving (ii). By adding and subtracting terms, (iii) follows from (i) and (ii). Part (iv) follows from $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ and (iii). Part (v) follows from (59) below and Lemma A.3. For part (vi), multiply (59) by $\|T^{-1/2}X_i\|$ on each side and then take the sum. The bound is the same as in (v). *Q.E.D.*

PROOF OF LEMMA A.11: The denominator of B is $D(F^0)$. Equation (61) shows that $\sqrt{T/N}[\hat{D}_0 - D(F^0)] = o_p(1)$. Thus it is sufficient to consider the numerator only. We shall prove

$$\left(\frac{\sqrt{T}}{N}\right)\left[\frac{1}{N}\sum_{i=1}^{N}\frac{X_i'\hat{F}}{T}\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i\hat{\sigma}_i^2 - \frac{1}{N}\sum_{i=1}^{N}\frac{X_i'F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i\sigma_i^2\right] = o_p(1) \tag{57}$$

and

$$\left(\frac{\sqrt{T}}{N}\right)\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{V}_i'\hat{F}}{T}\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i\hat{\sigma}_i^2 - \frac{1}{N}\sum_{i=1}^{N}\frac{V_i'F^0}{T}\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\lambda_i\sigma_i^2\right] = o_p(1). \tag{58}$$

Consider (57). There are four items being estimated, namely F, $\Lambda'\Lambda/N$, λ_i, and σ_i^2. Using the identity $\hat{a}\hat{b}\hat{c}\hat{d} - abcd = (\hat{a}-a)\hat{b}\hat{c}\hat{d} + a(\hat{b}-b)\hat{c}\hat{d} + ab(\hat{c}-c)\hat{d} + abc(\hat{d}-d)$, the first corresponding term is

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\frac{X_i'(\hat{F}-F^0H)}{T}\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i\hat{\sigma}_i^2\right\| \le \frac{\|\hat{F}-F^0H\|}{\sqrt{T}}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\|X_i\|}{\sqrt{T}}\left\|\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i\hat{\sigma}_i^2\right\|\right) = O_p(\delta_{NT}^{-1}).$$

The second corresponding term is $O_p(\delta_{NT}^{-1})$, which follows from Lemma A.10(iv). The term HH' arises in the interim, which just matches $(F^{0\prime}F^0/T)^{-1}$ by Lemma A.7 and $HH' - (F^{0\prime}F^0/T)^{-1} = O_p(\delta_{NT}^{-1})$.

For the third corresponding term, from (56),

$$\begin{aligned}\hat{\lambda}_i - H^{-1}\lambda_i &= T^{-1}\hat{F}'(F^0 - \hat{F}H^{-1})\lambda_i + T^{-1}\hat{F}'\varepsilon_i - T^{-1}\hat{F}'X_i'(\hat{\beta}-\beta) \\ &= T^{-1}\hat{F}'(F^0 - \hat{F}H^{-1})\lambda_i + T^{-1}(\hat{F} - \hat{F}^0H)'\varepsilon_i \\ &\quad + T^{-1}HF^{0\prime}\varepsilon_i - T^{-1}\hat{F}'X_i(\hat{\beta}-\beta).\end{aligned} \tag{59}$$

This means that the corresponding third term is also split into four expressions. Each expression can be easily shown to be dominated by $O_p(\delta_{NT}^{-1})$.

Next

$$\hat{\varepsilon}_{it} = \varepsilon_{it} + X_{it}'(\hat{\beta}-\beta) + (\hat{F}_t - H'F_t^0)H^{-1}\lambda_i + \hat{F}_t'(\hat{\lambda}_i - H^{-1}\lambda_i). \tag{60}$$

It is easy to show that $\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2 - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2 = O_p(\delta_{NT}^{-1})$. Furthermore, $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2 - \sigma_i^2 = \frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}^2 - E(\varepsilon_{it}^2)] = O_p(T^{-1/2})$. In summary, (57) is equal to $\sqrt{T/N}O_p(\delta_{NT}^{-1}) = o_p(1)$ if $T/N^2 \to 0$.

Consider (58). The only difference between (58) and (57) is X_i replaced by $\hat{V}_i$. Thus it is sufficient to prove

$$\left(\frac{\sqrt{T}}{N}\right)\frac{1}{N}\sum_{i=1}^{N}\frac{(\hat{V}_i - V_i)'F^0}{T}A_i = o_p(1),$$

where $A_i = (F^{0\prime}F^0/T)^{-1}(\Lambda'\Lambda/N)^{-1}\lambda_i\sigma_i^2 = O_p(1)$:

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\frac{(\hat{V}_i - V_i)'F^0}{T}A_i\right\| \le \left(\frac{1}{N}\sum_{i=1}^{N}T^{-1/2}\|\hat{V}_i - V_i\|\|A_i\|\right)\left\|T^{-1/2}F^0\right\|.$$

Now $\hat{V}_i - V_i = \frac{1}{N}\sum_{k=1}^{N}(\hat{a}_{ik} - a_{ik})X_k$, where

$$\begin{aligned}\hat{a}_{ik} - a_{ik} = {} & (\hat{\lambda}_i - H^{-1}\lambda_i)'(\hat{\Lambda}'\hat{\Lambda}/N)^{-1}\hat{\lambda}_k \\ & + \lambda_i' H'^{-1}[(\hat{\Lambda}'\hat{\Lambda}/N)^{-1} - H'(\Lambda'\Lambda/N)^{-1}H]\hat{\lambda}_k \\ & + \lambda_i'(\Lambda'\Lambda/N)^{-1}H(\hat{\lambda}_k - H^{-1}\lambda_k).\end{aligned}$$

Thus

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}T^{-1/2}\|\hat{V}_i - V_i\|\|A_i\| \\ &\le \left(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\lambda}_i - H^{-1}\lambda_i\|\|A_i\|\right)\left\|\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\right\|\left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\lambda}_k\|\left\|\frac{X_k}{\sqrt{T}}\right\|\right) \\ &\quad + \left\|\left[\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1} - H'\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}H\right]\right\|\left(\frac{1}{N}\sum_{i=1}^{N}\|H^{-1}\lambda_i\|\|A_i\|\right) \\ &\quad \times\left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\lambda}_k\|\left\|\frac{X_k}{\sqrt{T}}\right\|\right) \\ &\quad + \frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}H\right\|\|A_i\|\frac{1}{N}\sum_{k=1}^{N}\|(\hat{\lambda}_k - H^{-1}\lambda_k)\|\left\|\frac{X_k}{\sqrt{T}}\right\|.\end{aligned}$$

Each term on the right is bounded $O_p(\delta_{NT}^{-1})$ by Lemmas A.10. Thus (58) is equal to $\sqrt{T/N}O_p(\delta_{NT}^{-1})$, which is $o_p(1)$ if $T/N^2 \to 0$. *Q.E.D.*

PROOF OF LEMMA A.12: We only analyze terms involving the difference $\hat{\Omega} - \Omega$, because expressions that involve other estimates were analyzed in the proof of Lemma A.11. Consider

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N} X_i' M_{\hat{F}}(\hat{\Omega} - \Omega)\hat{F}\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i \\
&\quad = \frac{1}{NT}\sum_{i=1}^{N} X_i'(\hat{\Omega} - \Omega)\hat{F}\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i \\
&\qquad + \frac{1}{NT}\sum_{i=1}^{N}\frac{X_i'\hat{F}}{T}\hat{F}'(\hat{\Omega} - \Omega)\hat{F}\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i = a + b, \\
&a = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} X_{it}\hat{F}_t'\left(\frac{1}{N}\sum_{k=1}^{N}\hat{\varepsilon}_{kt}^2 - \sigma_{k,t}^2\right)\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i, \\
&\|a\| \le \left[\frac{1}{T}\sum_{t=1}^{T}\left(\left\|\frac{1}{N}\sum_{i=1}^{N} X_{it}\hat{F}_t'\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i\right\|\right)^2\right]^{1/2} \\
&\qquad\quad \times\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{k=1}^{N}\hat{\varepsilon}_{kt}^2 - \sigma_{k,t}^2\right)^2\right]^{1/2}.
\end{aligned}$$

But $\frac{1}{N}\sum_{k=1}^{N}\hat{\varepsilon}_{kt}^2 - \sigma_{k,t}^2 = \frac{1}{N}\sum_{k=1}^{N}[\hat{\varepsilon}_{kt}^2 - \varepsilon_{kt}^2] + \frac{1}{N}\sum_{k=1}^{N}[\varepsilon_{kt}^2 - \sigma_{k,t}^2] = \frac{1}{N}\sum_{k=1}^{N}[\hat{\varepsilon}_{kt}^2 - \varepsilon_{kt}^2] + O_p(N^{-1/2})$. Moreover, $\frac{1}{N}\sum_{k=1}^{N}[\hat{\varepsilon}_{kt}^2 - \varepsilon_{kt}^2] = O_p(\delta_{NT}^{-1})$ and so is the average over t. Thus $a = O_p(\delta_{NT}^{-1})$. Next

$$\begin{aligned}
\|b\| &\le T^{-1}\|\hat{F}'(\hat{\Omega} - \Omega)\hat{F}\|\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{X_i'\hat{F}}{T}\right\|\left\|\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i\right\| \\
&= T^{-1}\|\hat{F}'(\hat{\Omega} - \Omega)\hat{F}\|O_p(1).
\end{aligned}$$

But $\frac{1}{T}\|\hat{F}'(\hat{\Omega} - \Omega)\hat{F}\| = \frac{1}{T}\|\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'(\frac{1}{N}\sum_{k=1}^{N}\hat{\varepsilon}_{kt}^2 - \sigma_{k,t}^2)\| \le \sqrt{r}\{\frac{1}{T}\sum_{t=1}^{T}[\frac{1}{N} \times \sum_{k=1}^{N}(\hat{\varepsilon}_{kt}^2 - \sigma_{k,t}^2)]^2\}^{1/2} = O_p(\delta_{NT}^{-1})$, that is, $b = O_p(\delta_{NT}^{-1})$. Thus $\sqrt{N/T}(\hat{C} - C) = \sqrt{N/T}O_p(\delta_{NT}^{-1}) \to 0$ if $N/T^2 \to 0$. *Q.E.D.*

PROOF OF PROPOSITION 2: (i) Because $D(F^0) \xrightarrow{p} D_0$, it suffices to prove $\hat{D}_0 - D(F^0) \xrightarrow{p} 0$, where

$$D(F^0) = \frac{1}{NT}\sum_{i=1}^{N} X_i' M_{F^0} X_i' - \frac{1}{TN^2}\sum_{i=1}^{N}\sum_{k=1}^{N} X_i' M_{F^0} X_k a_{ik}$$

and $\hat{D}_0$ is the same as $D(F^0)$ with F^0 and a_{ik} replaced by $\hat{F}$ and $\hat{a}_{ik}$. The proof of Proposition A.2 shows that $\|\frac{1}{NT}\sum_{i=1}^{N} X_i'(M_{\hat{F}} - M_{F^0})X_i\| = O_p(1)\|P_{\hat{F}} - P_{F^0}\| \le O_p(\|\hat{\beta} - \beta\|^{1/2}) + O_p(\delta_{NT}^{-1}) = O_p(\delta_{NT}^{-1})$ by Lemma A.7(ii). It remains to show

$$\delta = \frac{1}{TN^2}\sum_{i=1}^{N}\sum_{k=1}^{N} X_i' M_{\hat{F}} X_k [\hat{a}_{ik} - a_{ik}] = o_p(1).$$

Noticing $a_{ik} = \lambda_k'(\Lambda'\Lambda/N)^{-1}\lambda_i$, and adding and subtracting terms yields

$$\begin{aligned}
\hat{a}_{ik} - a_{ik} &= (\hat{\lambda}_k - H^{-1}\lambda_k)'(\hat{\Lambda}'\hat{\Lambda}/N)^{-1}\hat{\lambda}_i \\
&\quad + \lambda_k' H'^{-1}[(\hat{\Lambda}'\hat{\Lambda}/N)^{-1} - H'(\Lambda'\Lambda/N)^{-1}H]\hat{\lambda}_i \\
&\quad + \lambda_k'(\Lambda'\Lambda/N)^{-1}H(\hat{\lambda}_i - H^{-1}\lambda_i) \\
&= b_{ik} + c_{ik} + d_{ik}.
\end{aligned}$$

Decompose δ into $\delta_1 + \delta_2 + \delta_3$, where δ_1, δ_2, and δ_3 are defined the same way as δ but with $\hat{a}_{ik} - a_{ik}$ replaced by b_{ik}, c_{ik}, and d_{ik}, respectively. From $T^{-1}\|X_i' M_{\hat{F}} X_k\| \le \|T^{-1/2}X_i\|\|T^{-1/2}X_k\|$,

$$\begin{aligned}
\|\delta_1\| &\le \left(\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}X_i\|\left\|\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1}\hat{\lambda}_i\right\|\right) \\
&\quad \times \left(\frac{1}{N}\sum_{k=1}^{N}\|T^{-1/2}X_k\|\|\hat{\lambda}_k - H^{-1}\lambda_k\|\right).
\end{aligned}$$

By Lemma A.10(v), $\|\delta_1\| = O_p(\delta_{NT}^{-1}) + O_p(\|\hat{\beta} - \beta\|) = o_p(1)$. Next,

$$\begin{aligned}
\|\delta_2\| &\le \left(\frac{1}{N}\sum_{i=1}^{N}\|T^{-1/2}X_i\|\|\hat{\lambda}_i\|\right)\left(\frac{1}{N}\sum_{k=1}^{N}\|T^{-1/2}X_k\|\|\lambda_k\|\|H^{-1}\|\right) \\
&\quad \times \left\|\left(\frac{\hat{\Lambda}'\hat{\Lambda}}{N}\right)^{-1} - H'\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}H\right\|,
\end{aligned}$$

which is $O_p(\delta_{NT}^{-2}) + O_p(\|\hat{\beta} - \beta\|)$ by Lemma A.10(iii). Finally, $\delta_3 = o_p(1)$ using the same argument for δ_1. In summary, $\hat{D}_0 - D(F^0) = o_p(1)$. In fact, we obtain a stronger result $\hat{D}_0 - D(F^0) = O_p(\delta_{NT}^{-1})$. Thus

$$(61) \qquad \sqrt{T/N}[\hat{D}_0 - D(F^0)] = \sqrt{T/N} O_p(\delta_{NT}^{-1}) = o_p(1)$$

provided that $T/N^2 \to 0$. Similarly

$$\sqrt{N/T}[\hat{D}_0 - D(F^0)] = \sqrt{N/T} O_p(\delta_{NT}^{-1}) = o_p(1)$$

provided that $N/T^2 \to 0$. These two results are used in the bias-corrected estimators.

Consider $\hat{D}_1$. Let $D_1^* = \frac{1}{NT} \sum_{i=1}^N \sigma_i^2 \sum_{t=1}^T Z_{it} Z_{it}'$. From $D_1^* \xrightarrow{p} D_1$, we only need to show $\hat{D}_1 - D_1^* \xrightarrow{p} 0$:

$$\begin{aligned}
\hat{D}_1 - D_1^* &= \frac{1}{N} \sum_{i=1}^N (\hat{\sigma}_i^2 - \sigma_i^2) \frac{1}{T} \sum_{t=1}^T \hat{Z}_{it} \hat{Z}_{it}' \\
&\quad + \frac{1}{N} \sum_{i=1}^N \sigma_i^2 \frac{1}{T} \sum_{t=1}^T (\hat{Z}_{it} \hat{Z}_{it}' - Z_{it} Z_{it}') \\
&= a + b, \\
\|a\| &\le \frac{1}{N} \sum_{i=1}^N |(\hat{\sigma}_i^2 - \sigma_i^2)| \frac{1}{T} \sum_{t=1}^T \|\hat{Z}_{it}\|^2.
\end{aligned}$$

From

$$\hat{\varepsilon}_{it} = \varepsilon_{it} + X_{it}'(\hat{\beta} - \beta) + (\hat{F}_t - H' F_t^0) H^{-1} \lambda_i + \hat{F}_t'(\hat{\lambda}_i - H^{-1} \lambda_i)$$

and

$$\begin{aligned}
\hat{\lambda}_i - H^{-1} \lambda_i &= T^{-1} \hat{F}'(F^0 - \hat{F} H^{-1}) \lambda_i \\
&\quad + T^{-1}(\hat{F} - \hat{F}^0 H)' \varepsilon_i + T^{-1} H F^{0\prime} \varepsilon_i - T^{-1} \hat{F}' X_i (\hat{\beta} - \beta)
\end{aligned}$$

(see the proofs of Lemmas A.10 and A.11), we can write

$$\frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_{it}^2 - \frac{1}{T} \sum_{t=1}^T \varepsilon_{it}^2 = O_p(\delta_{NT}^{-1}) v_i,$$

where $O_p(\delta_{NT}^{-1})$ does not depend on i and where v_i is such that $\frac{1}{N} \sum_{i=1}^N |v_i|^2 = O_p(1)$. Now $\hat{\sigma}_i^2 - \sigma_i^2 = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_{it}^2 - \frac{1}{T} \sum_{t=1}^T \varepsilon_{it}^2 + \frac{1}{T} \sum_{t=1}^T [\varepsilon_{it}^2 - \sigma_i^2] = O_p(\delta_{NT}^{-1}) v_i +$

$T^{-1/2}w_i$, where $w_i = T^{-1/2}\sum_{t=1}^{T}[\varepsilon_{it}^2 - \sigma_i^2] = O_p(1)$. Thus

$$\|a\| \le O_p(\delta_{NT}^{-1})\frac{1}{N}\sum_{i=1}^{N}|v_i|\left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{Z}_{it}\|^2\right)$$
$$+ T^{-1/2}\frac{1}{N}\sum_{i=1}^{N}|w_i|\left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{Z}_{it}\|^2\right) = O_p(\delta_{NT}^{-1}).$$

The proof of b being $o_p(1)$ is the same as that of part (i); the factor σ_i^2 does not affect the proof.

The proof of $\hat{D}_2$ being consistent for $\hat{D}_2$ is similar and thus is omitted.

Consider $\hat{D}_3$. Let $D_3^* = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Z_{it}Z_{it}'\sigma_{i,t}^2$. From $D_3^* \xrightarrow{p} D_3$, it is sufficient to show $\hat{D}_3 - D_3^* = o_p(1)$:

$$\hat{D}_3 - D_3^* = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{Z}_{it}\hat{Z}_{it}'(\hat{\varepsilon}_{it}^2 - \varepsilon_{it}^2)$$
$$+ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{Z}_{it}\hat{Z}_{it}' - Z_{it}Z_{it})\varepsilon_{it}^2$$
$$+ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{it}Z_{it}'(\varepsilon_{it}^2 - \sigma_{i,t}^2).$$

The first term is bounded by

$$\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\hat{Z}_{it}\|^4\right)^{1/2}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\varepsilon}_{it}^2 - \varepsilon_{it}^2)^2\right)^{1/2},$$

so it is easy to show $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\varepsilon}_{it}^2 - \varepsilon_{it}^2)^2 = o_p(1)$. The second term on the right is essentially analyzed in part (i); the extra factor ε_{it}^2 does not affect the analysis. The third term being $o_p(1)$ is due to the law of large numbers, as in White's heteroskedasticity estimator. Thus $\hat{D}_2 - D_3^* = o_p(1)$. *Q.E.D.*

On Instrumental Variable Interpretation in Section 6

The estimator can be interpreted as an instrumental variable (IV) estimator with Z_i as the IV. Left multiplying Z_i' on each side of

$$Y_i = X_i\beta + F\lambda_i + \varepsilon_i,$$

we obtain, noting $Z_i'F = 0$,

$$Z_i'Y_i = Z_i'X_i\beta + Z_i'\varepsilon.$$

Summing over i and solving for β, we obtain the instrumental variable estimator

$$\hat{\beta}_{\text{IV}} = \left(\sum_{i=1}^{N} Z_i' X_i\right)^{-1} \sum_{i=1}^{N} Z_i' Y_i.$$

Moreover, it is easy to show $\sum_{i=1}^{N} Z_i' X_i = \sum_{i=1}^{N} Z_i' Z_i$. Thus the instrumental variable estimator has the same form as the asymptotic representation of the interactive-effects estimator. It follows that the latter estimator is an asymptotic IV estimator with Z_i as instruments.

A Useful Lemma for Section 8

LEMMA A.13: *The following identities hold* (*i.e.*, $\dot{\varepsilon}_i$ *can be replaced by* ε_i):

$$\sum_{i=1}^{N} \dot{Z}_i(\hat{F})' \dot{\varepsilon}_i \equiv \sum_{i=1}^{N} \dot{Z}_i(\hat{F})' \varepsilon_i, \tag{62}$$

$$\sum_{i=1}^{N} \dot{Z}_i' \dot{\varepsilon}_i \equiv \sum_{i=1}^{N} \dot{Z}_i' \varepsilon_i. \tag{63}$$

PROOF: First note that

$$\dot{\varepsilon}_i = \varepsilon_i - \iota_T \bar{\varepsilon}_{i\cdot} - \bar{\varepsilon} + \iota_T \bar{\varepsilon}_{\cdot\cdot},$$

where $\bar{\varepsilon} = (\bar{\varepsilon}_{\cdot 1}, \bar{\varepsilon}_{\cdot 2}, \dots, \bar{\varepsilon}_{\cdot T})'$ does not depend on i. From the constraint $\sum_{t=1}^{T} \hat{F}_t = \hat{F}' \iota_T = 0$, we have $M_{\hat{F}} \iota_T = \iota_T$. Also, $\dot{X}_i' \iota_T = 0$ for all i. It follows that $\dot{Z}_i' \iota_T = 0$ in view of

$$\dot{Z}_i(\hat{F}) = \dot{X}_i' M_{\hat{F}} - \frac{1}{N} \sum_{k=1}^{N} a_{ik} \dot{X}_k' M_{\hat{F}}.$$

From $\sum_{i=1}^{N} \dot{X}_i = 0$ and $\sum_{i=1}^{N} a_{ik} = 0$, we have $\sum_{i=1}^{N} \dot{Z}_i = 0$. It follows that $\sum_{i=1}^{N} \dot{Z}_i(\hat{F})' \bar{\varepsilon} = 0$. Thus $\sum_{i=1}^{N} \dot{Z}_i(\hat{F})' \dot{\varepsilon}_i = \sum_{i=1}^{N} \dot{Z}_i(\hat{F})' \varepsilon_i$, proving (62). We have used the fact that $\sum_{i=1}^{N} a_{ik} = 0$, which follows from $\sum_{i=1}^{N} \lambda_i = 0$. Noting that $F^{0\prime} \iota_T = 0$, due to the restriction (29), the proof of (63) is identical to that of (62). *Q.E.D.*

APPENDIX C: ADDITIONAL RESULTS ON TESTING ADDITIVE VERSUS INTERACTIVE EFFECTS

PROOF OF (34): Under the i.i.d. assumption, $E(\varepsilon_i\varepsilon_j')=0$ and $E\varepsilon_i\varepsilon_i'=\sigma^2 I_T$. Thus

$$E(\eta\psi')=\sigma^2\frac{1}{NT}\sum_{i=1}^N X_i'M_F[X_i-\iota_T\bar{X}_{i.}-\bar{X}+\iota_T\bar{X}_{..}]$$
$$=\sigma^2\frac{1}{NT}\sum_{i=1}^N X_i'M_F X_i-\sigma^2\frac{1}{T}\bar{X}'M_F\bar{X}$$

from $M_F\iota_T=0$, because F contains ι_T as one of its columns. Next,

$$E\xi\psi'=\sigma^2\frac{1}{NT}\sum_{i=1}^N\left[\frac{1}{N}\sum_{k=1}^N a_{ik}X_k'M_F\right][X_i-\iota_T\bar{X}_{i.}-\bar{X}+\iota_T\bar{X}_{..}]$$
$$=\sigma^2\frac{1}{TN^2}\sum_{i=1}^N\sum_{k=1}^N a_{ik}X_k'M_F X_i$$
$$-\sigma^2\frac{1}{NT}\sum_{k=1}^N\left[\frac{1}{N}\sum_{i=1}^N a_{ik}\right]X_k'M_F\bar{X}$$
$$=\sigma^2\frac{1}{TN^2}\sum_{i=1}^N\sum_{k=1}^N a_{ik}X_k'M_F X_i-\frac{1}{T}\sigma^2\bar{X}'M_F\bar{X},$$

because $\sum_{i=1}^N a_{ik}=1$ under the null hypothesis. This follows from $\lambda_i=(1,\alpha_i)'$ with $\sum_i\alpha_i=0$. Thus

$$E[(\eta-\xi)\psi']=\sigma^2\frac{1}{NT}\sum_{i=1}^N X_i'M_F X_i$$
$$-\frac{1}{T}\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N a_{ik}X_k'M_F X_i=\sigma^2 D(F^0).$$

Q.E.D.

C.1. *Time-Invariant versus Time-Varying Individual Effects*

Consider the null hypothesis of the fixed-effects model

$$Y_{it}=X_{it}'\beta+\lambda_i+\varepsilon_{it}, \tag{64}$$

where λ_i is an unobservable scalar. The alternative hypothesis is that the fixed effects are time-varying,

$$Y_{it} = X_{it}'\beta + \lambda_i F_t + \varepsilon_{it}, \tag{65}$$

where F_t is also an unobservable scalar. This is a single factor interactive-effects model. If $F_t = 1$ for all t, the fixed-effects model is obtained.

The interactive-effects estimator for β is consistent under both models (64) and (65), but is less efficient than the least squares dummy-variable estimator for model (64), as the latter imposes the restriction $F_t = 1$ for all t. Nevertheless, the fixed-effects estimator is inconsistent under model (65). The principle of the Hausman test is applicable here.

The least squares dummy-variable estimator is

$$\sqrt{NT}(\hat{\beta}_{\mathrm{FE}} - \beta) = \left(\frac{1}{NT}\sum_{i=1}^{N} X_i' M_T X_i\right)^{-1} \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_i' M_T \varepsilon_i,$$

where $M_T = I_T - \iota_T \iota_T'/T$. For the interactive model, the estimator is

$$\begin{aligned}
&\sqrt{NT}(\hat{\beta}_{\mathrm{IE}} - \beta) \\
&\quad = D(F^0)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[X_i' M_{F^0} - \frac{1}{N}\sum_{k=1}^{N} a_{ik} X_k' M_{F^0}\right]\varepsilon_i + o_p(1).
\end{aligned}$$

Let

$$\eta = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} X_i' M_{F^0}\varepsilon_i, \qquad \xi = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{k=1}^{N} a_{ik} X_k' M_{F^0}\right]\varepsilon_i. \tag{66}$$

By Proposition A.3,

$$\sqrt{NT}(\hat{\beta}_{\mathrm{IE}} - \beta) = D(F^0)^{-1}(\eta - \xi) + o_p(1). \tag{67}$$

Under the null hypothesis, $F^0 = \iota_T$, and thus $M_T = M_{F^0}$ and

$$\sqrt{NT}(\hat{\beta}_{\mathrm{FE}} - \beta) = A^{-1}\eta,$$

where $A = (\frac{1}{NT}\sum_{i=1}^{N} X_i' M_T X_i)$.

The variances of the two estimators (the conditional variance to be precise) are

$$\begin{aligned}
&\mathrm{var}(\sqrt{NT}(\hat{\beta}_{\mathrm{FE}} - \beta)) = \sigma^2 A^{-1}, \\
&\mathrm{var}(\sqrt{NT}(\hat{\beta}_{\mathrm{IE}} - \beta)) = \sigma^2 D(F^0)^{-1},
\end{aligned}$$

respectively. To show that the variance of the difference in estimators is equal to the difference in variances, that is,

$$\text{var}(\hat{\beta}_{\text{IE}} - \hat{\beta}_{\text{FE}}) = \text{var}(\hat{\beta}_{\text{IE}}) - \text{var}(\hat{\beta}_{\text{FE}}),$$

it suffices to show

$$E(\eta\xi') = E(\xi\xi'). \tag{68}$$

The proof is given below. Note that $E\xi\xi'$ is positive definite, that is, $A - D(F^0) = E\xi\xi'$ is positive definite. This implies that $\text{var}(\sqrt{NT}(\hat{\beta}_{\text{IE}} - \hat{\beta}_{\text{FE}})) = \sigma^2[D(F^0)^{-1} - A^{-1}]$ is a matrix of full rank (positive definite). Thus

$$J = NT\sigma^2(\hat{\beta}_{\text{IE}} - \hat{\beta}_{\text{FE}})'[D(F^0)^{-1} - A^{-1}]^{-1}(\hat{\beta}_{\text{IE}} - \hat{\beta}_{\text{FE}}) \xrightarrow{d} \chi^2_p.$$

Replacing $D(F^0)$ and σ^2 by their consistent estimators, the above is still true. Proposition 2 shows that $D(F^0)$ is consistently estimated by $\hat{D}_0$. Let $\hat{\sigma}^2 = \frac{1}{L}\sum_{i=1}^N \sum_{t=1}^T \hat{\varepsilon}_{it}^2$, where $L = NT - (N+T) - p + 1$. Then $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$.

C.2. *Homogeneous versus Heterogeneous Time Effects*

For the purpose of comparison, the usual time effects are called homogeneous time effects since they are the same across individuals:

$$Y_{it} = X_{it}\beta + F_t + \varepsilon_{it},$$

where F_t is an unobservable scalar. The heterogeneous time-effects model is

$$Y_{it} = X_{it}\beta + \lambda_i F_t + \varepsilon_{it},$$

which is a simple interactive-effects model with $r = 1$. The least squares dummy-variable method for the homogeneous effects gives

$$\sqrt{NT}(\hat{\beta}_{\text{FE}} - \beta) = B^{-1}\psi,$$

where $B = (\frac{1}{NT}\sum_{i=1}^N (X_i - \bar{X})'(X_i - \bar{X}))$ and $\psi = \frac{1}{\sqrt{NT}}\sum_{i=1}^N (X_i - \bar{X})'\varepsilon_i$, and $\bar{X} = \frac{1}{N}\sum_{i=1}^N X_i$, is a $T \times 1$ vector. The interactive-effects estimator has the same representation as in (67). Under the null hypothesis of the homogeneous time effect, we have $\lambda_i = 1$ for all i and hence $a_{ik} = 1$. It follows that

$$\text{var}(\eta - \xi) = \sigma^2 D(F^0) = \sigma^2 \frac{1}{NT}\sum_{i=1}^N X_i' M_{F^0} X_i - \sigma^2 \frac{1}{T}\bar{X}' M_{F^0}\bar{X}.$$

It is shown below that

$$E\eta\psi' = \text{var}(\eta - \xi) = \sigma^2 D(F^0), \quad E(\xi\psi') = 0. \tag{69}$$

This implies that

$$\operatorname{var}(\hat{\beta}_{\mathrm{IE}} - \hat{\beta}_{\mathrm{FE}}) = \operatorname{var}(\hat{\beta}_{\mathrm{IE}}) - \operatorname{var}(\hat{\beta}_{\mathrm{FE}}).$$

Thus Hausman's test takes the form

$$J = NT\sigma^2(\hat{\beta}_{\mathrm{IE}} - \hat{\beta}_{\mathrm{FE}})'[D(F^0)^{-1} - B^{-1}]^{-1}(\hat{\beta}_{\mathrm{IE}} - \hat{\beta}_{\mathrm{FE}}) \xrightarrow{d} \chi^2_p.$$

The above still holds with $D(F^0)$ and σ^2 replaced by $\hat{D}_0$ and $\hat{\sigma}^2$.

PROOF OF (68): Under i.i.d. assumptions for ε_{it}, using $E\varepsilon_i\varepsilon_j' = 0$ for $i \neq j$ and $E\varepsilon_i\varepsilon_i' = \sigma^2 I_T$,

$$E(\eta\xi') = \sigma^2 \frac{1}{N^2T}\sum_{i=1}^N\sum_{k=1}^N a_{ik}X_k'M_F X_i,$$

$$\begin{aligned}
E(\xi\xi') &= \frac{1}{N^3T}\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N\sum_{\ell=1}^N a_{ik}X_k'M_F E(\varepsilon_i\varepsilon_j)M_F X_\ell a_{j\ell} \\
&= \sigma^2\frac{1}{N^2T}\sum_{k=1}^N\sum_{\ell=1}^N\left(\frac{1}{N}\sum_{i=1}^N a_{ik}a_{i\ell}\right)X_k'M_F X_\ell \\
&= \sigma^2\frac{1}{N^2T}\sum_{k=1}^N\sum_{\ell=1}^N a_{k\ell}X_k'M_F X_\ell = E\eta\xi'
\end{aligned}$$

since $\frac{1}{N}\sum_{i=1}^N a_{ik}a_{i\ell} = a_{k\ell}$. *Q.E.D.*

PROOF OF (69): We have

$$\begin{aligned}
E(\eta\phi') &= \sigma^2\frac{1}{NT}\sum_{i=1}^N X_i'M_F[X_i - \bar{X}] \\
&= \sigma^2\frac{1}{NT}\sum_{i=1}^N X_i'M_F X_i - \sigma^2\frac{1}{T}\bar{X}'M_F\bar{X},
\end{aligned}$$

$$\begin{aligned}
E(\xi\psi') &= \sigma^2\frac{1}{N^2T}\sum_{i=1}^N\sum_{k=1}^N a_{ik}X_k'M_F[X_i - \bar{X}] \\
&= \sigma^2\frac{1}{T}\bar{X}'M_F(\bar{X} - \bar{X}) = 0.
\end{aligned}$$

Note that $a_{ik} = 1$ for all i and k under the null hypothesis. *Q.E.D.*

C.3. *The Number of Factors*

In this section we argue why the number of factors can be consistently estimated, and discuss how to use this fact to discern additive and interactive-effects. For pure factor models, Bai and Ng (2002) showed that the number of factors can be consistently estimated based on the information criterion approach. Their analysis can be amended to our current setting. Details will not be presented to avoid repetition, but intuition will be given.

We assume that $r \le \overline{k}$, where $\overline{k}$ is given. Suppose r is unknown, but we entertain k factors in the estimation. It can be shown that as long as $k \ge r$, we have $\hat{\beta}_{\mathrm{IE}}^{(k)} - \beta = O_p(1/\sqrt{NT})$, where the superscript k indicates k factors are estimated. Let $\hat{u}_{it}(k) = Y_{it} - X_{it}'\hat{\beta}_{\mathrm{IE}}^{(k)}$ and $\hat{\varepsilon}_{it}(k) = \hat{u}_{it}(k) - \hat{\lambda}_i(k)'\hat{F}_t(k)$. Then

$$\hat{u}_{it}(k) = \lambda_i' F_t + \varepsilon_{it} + O_p(1/\sqrt{NT}).$$

Thus $\hat{u}_{it}$ has a pure factor model; the $O_p(1/\sqrt{NT})$ error will not affect the analysis of Bai and Ng (2002). This means that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2(k) - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2 = O_p\left(\frac{1}{\min[N,T]}\right).$$

Since $\overline{k} \ge r$, the above is true when k is replaced by $\overline{k}$. Thus,

$$\hat{\sigma}^2(k) - \hat{\sigma}^2(\overline{k}) = O_p(1/\min[N,T]),$$

where $\hat{\sigma}^2(k) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2(k)$.

If $k < r$, unless $\lambda_i' F_t$ are uncorrelated with the regressors, and $E(\lambda_i) = 0$ and $E(F_t) = 0$, β cannot be consistently estimated. In any case, F^0 cannot be consistently estimated since F^0 is $T \times r$ and $\hat{F}(k)$ is only $T \times k$. The consequence of inconsistency is

$$\hat{\sigma}^2(k) - \hat{\sigma}^2(\overline{k}) > c > 0$$

for some $c > 0$, not depending on N and T. This implies that any penalty function that converges to zero but is of greater magnitude than $O_p(1/\min[N,T])$ will lead to consistent estimation of the number of factors. In particular,

$$\mathrm{CP}(k) = \hat{\sigma}^2(k) + \hat{\sigma}^2(\overline{k})[k(N+T) - k^2]\frac{\log(NT)}{NT}$$

or

$$\mathrm{IC}(k) = \log\hat{\sigma}^2(k) + [k(N+T) - k^2]\frac{\log(NT)}{NT}$$

will work. That is, let $\hat{k} = \arg\min_{k\leq\bar{k}} \mathrm{CP}(k)$ or $\hat{k} = \arg\min_{k\leq\bar{k}} \mathrm{IC}(k)$. Then $P(\hat{k} = r) \to 1$ as $N, T \to 0$. Although the usual Bayesian information criterion (BIC) only assumes either $T \to \infty$ or $N \to \infty$ but not both, the $\mathrm{IC}(k)$ has the same form as the BIC criterion, as there is a total of NT observations. With k factors, the number of parameters is $k(N + T) - k^2 + p$, where k^2 reflects the restriction $F'F/T = I$ and $\Lambda'\Lambda =$ diagonal, but p does not vary with k, so can be excluded in the penalty function. The CP criterion is similar to Mallows' C_p.

Ignoring k^2 for a moment (since it is dominated by $k(N + T)$ for large N and T), the penalty function in $\mathrm{IC}(k)$ is $k \cdot g(N, T)$, where $g(N, T) = (N + T)\frac{\log(NT)}{NT}$. Clearly, the penalty function goes to zero as $N, T \to 0$, unless $N = \exp(T)$ or $T = \exp(N)$ (these are the rare situations where BIC breaks down; Bai and Ng (2002) suggested several alternative criteria). In addition, $g(N, T)$ is of larger magnitude than $1/\min[N, T]$ since $g(N, T) * \min[N, T] \to \infty$. These two properties of a penalty function imply consistency, as shown by Bai and Ng (2002).

Given that the number of factors can be consistently estimated, we can determine whether an additive model or interactive model is more appropriate. Suppose the null hypothesis postulates time-invariant fixed effects as $Y_{it} = X_{it}'\beta + \lambda_i + \varepsilon_{it}$. Then

$$Y_{it} - \bar{Y}_{i\cdot} = (X_{it} - \bar{X}_{i\cdot})'\beta + \varepsilon_{it} - \bar{\varepsilon}_{i\cdot}.$$

Under the time-varying fixed-effects model $Y_{it} = X_{it}'\beta + \lambda_i F_t + \varepsilon_{it}$, we have

$$Y_{it} - \bar{Y}_{i\cdot} = (X_{it} - \bar{X}_{i\cdot})'\beta + \lambda_i(F_t - \bar{F}) + \varepsilon_{it} - \bar{\varepsilon}_{i\cdot}.$$

Under the null hypothesis, no factor exists, and under the alternative, there exists one factor.

The same argument works for the fixed time-effects model, in which we use $Y_{it} - \bar{Y}_{\cdot t}$ as the left-hand side variable and $X_{it} - \bar{X}_{\cdot t}$ as the right-hand side variable.

Next consider the additive versus the interactive model:

$$Y_{it} = X_{it}'\beta + \mu + \alpha_i + \xi_t + \varepsilon_{it}$$

or

$$\dot{Y}_{it} = \dot{X}_{it}'\beta + \dot{\varepsilon}_{it},$$

where $\dot{Y}_{it}$ and $\dot{X}_{it}$ were defined previously. Therefore, the transformed data exhibit no factors. Under the interactive model (32), the transformed data obey

$$\dot{Y}_{it} = \dot{X}_{it}'\beta + \lambda_i' F_t + \dot{\varepsilon}_{it}.$$

The factor structure is unscathed by the transformation and the number of factors is still two.

APPENDIX D: ADDITIONAL RESULTS FOR SECTION 10

We now explain the meaning of $D(F^0) > 0$ and argue that it can be segregated into some intuitive and reasonable conditions. To simplify notation and for ease of discussion, we assume the only regressors are time invariant or common (no X_{it}), that is,

$$X_i = (\iota_T x_i', W), \quad \beta' = (\gamma', \delta').$$

The condition $D(F^0) > 0$ implies the following four restrictions:

I. *Genuine interactive-effects*: F^0 or its rotation does not contain ι_T; Λ or its rotation does not contain ι_N. Otherwise, we are back in the environment of Hausman and Taylor, and instrumental variables must be used to identify β. In algebraic notation

$$\frac{1}{T}\iota_T' M_{F^0} \iota_T > 0 \quad \text{and} \quad \frac{1}{N}\iota_N' M_\Lambda \iota_N > 0.$$

II. *No multicollinearity between W and F^0*: The following matrix is positive definite:

$$\frac{1}{T} W' M_{F^0} W > 0.$$

Without this assumption, even if F^0 is observable, we cannot identify β and Λ due to multicollinearity.

III. *No multicollinearity between $\underline{x}$ and Λ*:

$$\frac{1}{N}\underline{x}' M_\Lambda \underline{x} > 0.$$

This is required for identification of β and F^0.

IV. *Identification of the grand mean, if it exists*: At least one of the following holds:

$$\frac{1}{N}(\underline{x}, \iota_N)' M_\Lambda (\underline{x}, \iota_N) > 0, \tag{70}$$

$$\frac{1}{T}(\iota_T, W)' M_{F^0} (\iota_T, W) > 0. \tag{71}$$

That is, either $\underline{x}$ does not contain ι_N or W does not contain ι_T. If both contain the constant regressor, there will be two grand mean parameters, and thus they are not identifiable.

To see that $D(F^0) > 0$ implies the above four conditions, we simply compute

$$D(F) = \begin{bmatrix} ((1/N)\underline{x}'M_\Lambda\underline{x})(\iota_T'M_F\iota_T/T) & ((1/N)\underline{x}'M_\Lambda\iota_N)(\iota_T'M_FW/T) \\ (W'M_F\iota_T/T)((1/N)\iota_N'M_\Lambda\underline{x}) & ((1/N)\iota_N'M_\Lambda\iota_N)(W'M_FW/T) \end{bmatrix}.$$

For a positive definite matrix, the diagonal block matrices must be positive definite. This leads to the first three conditions immediately. To see that $D(F^0) > 0$ also implies IV, we use a contradiction argument. Suppose neither of the matrices in (70) and (71) is positive definite, and since they are semipositive definite, their determinants must be zero. Then it is not difficult to show that the determinant of $D(F^0)$ is also zero. This contradicts $D(F^0) > 0$.

More interestingly, the four conditions above are also sufficient for $D(F^0) > 0$, a consequence of the following lemma.

LEMMA 14: *Let A be a $q \times q$ symmetric matrix. Assume the $(q+1)\times(q+1)$ matrix*

$$\bar{A} = \begin{bmatrix} A & \alpha \\ \alpha' & \tau \end{bmatrix} > 0$$

is positive definite, so $A > 0$ and $\tau > 0$ (a scalar). Suppose

$$\bar{B} = \begin{bmatrix} \nu & b' \\ b & B \end{bmatrix} \geq 0, \quad \textit{with} \quad \nu > 0, B > 0,$$

is semipositive definite, where B is $\ell \times \ell$ and ν is scalar. Then the following $(q+\ell)\times(q+\ell)$ matrix is positive definite:

$$\bar{A}\diamond\bar{B} = \begin{bmatrix} A\nu & \alpha b' \\ b\alpha' & \tau B \end{bmatrix} > 0.$$

The role of $\bar{A}$ and $\bar{B}$ can be reversed. The lemma only requires one of them to be positive definite, not both. Now suppose (70) holds. Let $\bar{A} = \frac{1}{N}(\underline{x}, \iota_N)'M_\Lambda(\underline{x}, \iota_N)$ and $\bar{B} = \frac{1}{T}(\iota_T, W)'M_{F^0}(\iota_T, W)$, and $\bar{A} > 0$. In addition, $A = \frac{1}{N}\underline{x}'M_\Lambda\underline{x} > 0$, $\tau = \iota_N'M_\Lambda\iota_N > 0$, $\nu = \frac{1}{T}\iota_T'M_{F^0}\iota_T > 0$, and $B = W'M_{F^0}W/T > 0$, all following from the first three conditions. Thus the assumptions of Lemma 14 hold. It follows that $\bar{A}\diamond\bar{B} > 0$, but $\bar{A}\diamond\bar{B} = D(F^0)$. Thus the four conditions imply $D(F^0) > 0$. We now summarize the result.

LEMMA 15: *The matrix $D(F^0) > 0$ if and only if the above four conditions hold.*

REMARK ON LEMMA 14: $\bar{B}$ need not be positive definite. For example, for $\ell = 1$, $\bar{B}$ can be the 2×2 matrix with each entry being 1. Then $\bar{A}\diamond\bar{B} = \bar{A} > 0$. The lemma holds if $\bar{A} \geq 0$ with $A > 0$ and $\tau > 0$, but $\bar{B} > 0$ (reversing the role of $\bar{A}$ and $\bar{B}$). Moreover, from $\bar{A}\diamond\bar{B} > 0$, one can deduce the condition of the lemma (or the conditions reversing the role of $\bar{A}$ and $\bar{B}$). In this sense, the condition is necessary and sufficient. The operator $\diamond$ is analogous to the Hadamard product, which requires equal sizes for $\bar{A}$ and $\bar{B}$, and is defined as componentwise multiplication. We are not aware of any matrix result in this nature. The lemma can be proved for $\ell = 1$ and for arbitrary q then with induction over ℓ (the proof is available from the author). *Q.E.D.*

PROOF OF PROPOSITION 3: As in the proof of Proposition 1, denote the true value by (β^0, F^0). Recall that the objective function can be written as $S_{NT}(\beta, F) = \tilde{S}_{NT}(\beta, F) + o_p(1)$, where

$$\tilde{S}_{NT}(\beta, F) = (\beta - \beta^0)' D(F)(\beta - \beta^0) + \theta' B \theta,$$

where $B = [(\Lambda'\Lambda/N)^{-1} \otimes I_T] > 0$, and θ is a function of (β, F) such that

$$\theta = \text{vec}(M_F F^0) + B^{-1} \frac{1}{NT} \sum_{i=1}^{N} (\lambda_i \otimes M_F X_i)(\beta - \beta^0); \tag{72}$$

see the proof of Proposition 1 in Appendix A. Since $D(F)$ is semipositive definite for any F and since B is positive definite,

$$\tilde{S}_{NT}(\beta, F) \geq 0$$

for all (β, F). On the other hand, $\tilde{S}_{NT}(\beta^0, F^0) = 0$. We show that (β^0, F^0) is the unique point at which $\tilde{S}_{NT}(\beta, F)$ achieves its minimum, where uniqueness with respect to F^0 is up to a rotation (identification restrictions on F and Λ in fact fix the rotation). Letting

$$(\beta^*, F^*) = \arg\min \tilde{S}_{NT}(\beta, F),$$

we can show $(\beta^*, F^*) = (\beta^0, F^0)$. Since $\tilde{S}_{NT}(\beta^*, F^*) = 0$, we must have

$$(\beta^* - \beta^0)' D(F^*)(\beta^* - \beta^0) = 0 \quad \text{and} \quad \theta^* = \theta(\beta^*, F^*) = 0.$$

If $D(F^*)$ is of full rank, then $\beta^* - \beta^0 = 0$. In this case, from $0 = \theta^* = \text{vec}(M_{F^*} F^0)$, we have $F^* = F^0$. Only when $D(F^*)$ is not full rank is it possible for $\beta^* \neq \beta^0$. The matrix $D(F^*)$ will not be full rank if F^* or its rotation contains the column ι_T or contains a column of W. We show that this is not possible under $D(F^0) > 0$. If F^* contains the column ι_T, then

$$D(F^*) = \begin{bmatrix} 0 & 0 \\ 0 & ((1/N)\iota_N' M_\Lambda \iota_N)(W' M_{F^*} W)/T \end{bmatrix}$$

and it follows that

$$0 = (\beta^* - \beta^0)D(F^*)(\beta^* - \beta^0) = a(\delta^* - \delta^0)'(W'M_{F^*}W/T)(\delta^* - \delta^0),$$

where $a = \frac{1}{N}\iota_N' M_\Lambda \iota_N > 0$. The above implies that

$$M_{F^*}W(\delta^* - \delta^0) = 0$$

since $x'x = 0$ implies $x = 0$. Therefore,

$$\begin{aligned} M_{F^*}X_i(\beta^* - \beta^0) &= (M_{F^*}\iota_T x_i', M_{F^*}W)(\beta^* - \beta^0) \\ &= (0, M_{F^*}W)(\beta^* - \beta^0) = M_{F^*}W(\delta^* - \delta^0) = 0. \end{aligned}$$

Thus, by (72), $0 = \theta^* = \text{vec}(M_{F^*}F^0)$. It follows that $F^* = F^0$. Thus F^* cannot contain ι_T since F^0 does not contain ι_T, a contradiction. Next, suppose that F^* contains at least one column of W. Partition $W = (W_1, W_2)$ and suppose, without loss of generality, that F^* contains W_2. Then $M_{F^*}W = (M_{F^*}W_1, 0)$ and

$$D(F^*) = \begin{bmatrix} ((1/N)\underline{x}'M_\Lambda\underline{x})(\iota_T'M_{F^*}\iota_T/T) & ((1/N)\underline{x}'M_\Lambda\iota_N)(\iota_T'M_{F^*}W_1/T) & 0 \\ (W_1'M_{F^*}\iota_T/T)((1/N)\iota_N'M_\Lambda\underline{x}) & ((1/N)\iota_N'M_\Lambda\iota_N)(W_1'M_{F^*}W_1/T) & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Under $\frac{1}{T}\iota_T'M_{F^*}\iota_T > 0$, the first 2×2 block of $D(F^*)$ is positive definite by Lemma 14. Partition $\delta = (\delta_1', \delta_2')'$ so $\beta = (\gamma', \delta_1', \delta_2')'$. Partition β^* and β^0 correspondingly. From

$$(\beta^* - \beta^0)'D(F^*)(\beta^* - \beta^0) = 0$$

we have $\gamma^* - \gamma^0 = 0$ and $\delta_1^* - \delta_1^0 = 0$. Thus $\beta^* - \beta^0 = (0', 0', \delta_2^{*\prime} - \delta_2^{0\prime})$. Together with $M_{F^*}W_2 = 0$, we have

$$M_{F^*}X_i(\beta^* - \beta^0) = (M_F^*\iota_T x_i', M_{F^*}W_1, 0)(\beta^* - \beta^0) = 0.$$

In view of (72), $0 = \theta^* = \text{vec}(M_{F^*}F^0)$. It follows that $F^* = F^0$, again a contradiction. In summary, under the assumption that $D(F^0) > 0$, the optimal solution of $\tilde{S}_{NT}(\beta, F)$ is achieved uniquely at (β^0, F^0). This implies that $\hat{\beta}$ is a consistent estimator for β^0; see the proof of Proposition 1 in Appendix A. *Q.E.D.*

APPENDIX E: FINITE SAMPLE PROPERTIES VIA SIMULATIONS

Data are generated according to

$$Y_{it} = X_{it,1}\beta_1 + X_{it,2}\beta_2 + a\lambda_i'F_t + \varepsilon_{it},$$

$\lambda_i = (\lambda_{i1}, \lambda_{i2})'$, and $F_t = (F_{t1}, F_{t2})$. The regressors are generated according to

$$X_{it,1} = \mu_1 + c_1\lambda_i'F_t + \iota'\lambda_i + \iota'F_t + \eta_{it,1},$$
$$X_{it,2} = \mu_2 + c_2\lambda_i'F_t + \iota'\lambda_i + \iota'F_t + \eta_{it,2}$$

with $\iota' = (1, 1)$. The variables λ_{ij}, F_{tj}, and $\eta_{it,j}$ are all i.i.d. $N(0, 1)$. The important parameters are

$$(\beta_1, \beta_2) = (1, 3).$$

We set $c_1 = c_2 = \mu_1 = \mu_2 = 1$ and $a = 1$. We first consider the case of

$$\varepsilon_{it} \text{ i.i.d. } N(0, 4)$$

and then extend it to correlated errors.

To estimate $(\hat{\beta}_{\text{IE}}, \hat{F})$, consider the iteration scheme in (11) and (12). A starting value for β or F is needed. The least squares objective function is not globally convex, so there is no guarantee that an arbitrary starting value will lead to the global optimal solution. Two natural choices exist. The first is the simple least squares estimator of β, ignoring the interactive-effects. The second is the principal components estimator for F, ignoring the regressors. If λ_i and F_t have unusually large nonzero means (arbitrarily stretching the model), the first choice can fail, but the second choice leads to the optimal solution. This is because as the interactive-effects become dominant, it makes sense to estimate the factor structure first. In this case, using the within-group estimator β as a starting value will also work. To minimize the chance of local minimum, both choices are used. Upon convergence, we choose the estimator that gives a smaller value of the objective function. Iterations based on (11) and (12) have difficulty achieving convergence for models with time-invariant and common regressors.

A more robust iteration scheme (having a much better convergence property) is the following: given F and Λ, compute $\hat{\beta}(F, \Lambda) = (\sum_{i=1}^N X_i'X_i)^{-1} \times \sum_{i=1}^N X_i'(Y_i - F\lambda_i)$, and given β, compute F and Λ from the pure factor model $W_i = F\lambda_i + e_i$ with $W_i = Y_i - X_i\beta$. This iteration scheme only requires a single matrix inverse $(\sum_{i=1}^N X_i'X_i)^{-1}$, so there is no need to update during iteration, unlike the scheme of $\hat{\beta}(F) = (\sum_{i=1}^N X_i'M_F X_i)^{-1} \sum_{i=1}^N X_i'M_F Y_i$. Furthermore, if $N > T$, we do principal components analysis using WW' $(T \times T)$, and if $N < T$, we use $W'W$ $(N \times N)$ to speed up computation. The same product $\hat{F}\hat{\lambda}_i$ is achieved, no matter which matrix is used. For the model associated with Table I, the iteration method in the previous paragraph has many realizations that do not converge to global optimum, but for the iteration scheme here, all lead to global solutions.

TABLE III

VARIOUS ESTIMATORS; I.I.D. ERRORS

		Within-Group Estimator				Infeasible Estimator				Interactive-Effects Estimator			
N	T	Mean $\beta_1 = 1$	SD	Mean $\beta_2 = 3$	SD	Mean β_1	SD	Mean β_2	SD	Mean β_1	SD	Mean β_2	SD
100	3	1.363	0.145	3.364	0.145	0.990	0.161	3.008	0.158	1.022	0.236	3.025	0.229
100	5	1.382	0.096	3.382	0.098	1.000	0.089	3.000	0.086	1.021	0.133	3.021	0.129
100	10	1.388	0.064	3.393	0.063	0.998	0.055	3.002	0.054	1.011	0.071	3.014	0.067
100	20	1.396	0.043	3.399	0.042	0.997	0.034	3.002	0.035	1.002	0.040	3.006	0.040
100	50	1.399	0.027	3.400	0.027	1.000	0.021	3.001	0.021	1.002	0.024	3.003	0.024
100	100	1.399	0.020	3.399	0.020	1.000	0.015	2.999	0.015	1.001	0.017	3.000	0.017
3	100	1.360	0.150	3.361	0.136	0.996	0.090	2.999	0.093	1.039	0.240	3.032	0.231
5	100	1.384	0.098	3.380	0.095	1.003	0.071	2.998	0.070	1.025	0.132	3.019	0.128
10	100	1.389	0.062	3.393	0.063	0.998	0.046	3.002	0.048	1.009	0.066	3.011	0.069
20	100	1.394	0.043	3.395	0.042	0.999	0.034	3.001	0.035	1.004	0.041	3.006	0.041
50	100	1.399	0.027	3.398	0.028	1.000	0.021	3.000	0.021	1.002	0.024	3.002	0.024

For comparison, we also compute two additional estimators: (i) the usual within-group estimator $\hat{\beta}_{\text{LSDU}}$ and (ii) the infeasible estimator $\hat{\beta}(F)$, assuming F is observable.

From Table III, we can draw several conclusions. First, the within estimator is biased and inconsistent. Biases become more severe when the interactive-effects are magnified by setting a larger a. For example, if $a = 10$, the biases are also almost ten times larger (not reported). The infeasible estimator and the interactive-effects estimator are virtually unaffected by the value of a. Second, both the feasible and the interactive-effects estimators are unbiased and consistent. The interactive-effects estimator is less efficient than the infeasible estimator, as can be seen from the larger standard errors, which is consistent with the theory. Third, even with small N and T, the interactive-effects estimator performs quite well, and as both N and T increase and the standard deviation becomes smaller.

Table IV gives results for cross-sectionally correlated ε_{it}. For cross-sectional data in reality, a large value of $|i - j|$ does not necessarily mean the correlation between ε_{it} and ε_{jt} is small. Nevertheless, for the purpose of introducing cross-section correlation, ε_{it} is generated as AR(1) for each fixed t such that

$$\varepsilon_{it} = \rho\varepsilon_{i-1,t} + e_{it},$$

where $\rho = 0.7$. Once cross-section correlation is introduced, the data can be permuted cross-sectionally if wanted, but the results do not depend on any particular permutation. We generate stationary data by discarding the first 100 observations. This implies that $\text{var}(\varepsilon_{it}) = \sigma_e^2/(1 - \rho^2) \approx 4$ for $\sigma_e^2 = 2$ and $\rho = 0.7$. Thus the variance of ε_{it} is approximately the same as the variance for Table III. Theorem 1 claims that for $N \gg T$, cross-section correlation does not

TABLE IV

VARIOUS ESTIMATORS; CROSS-SECTIONALLY CORRELATED ERRORS

		Within-Group Estimator				Infeasible Estimator				Interactive-Effects Estimator			
N	T	Mean $\beta_1=1$	SD	Mean $\beta_2=3$	SD	Mean β_1	SD	Mean β_2	SD	Mean β_1	SD	Mean β_2	SD
100	3	1.368	0.136	3.366	0.142	1.005	0.176	2.996	0.172	1.062	0.235	3.061	0.242
100	5	1.381	0.094	3.382	0.092	0.995	0.092	2.999	0.093	1.064	0.152	3.069	0.157
100	10	1.390	0.061	3.393	0.061	0.998	0.056	3.002	0.058	1.053	0.105	3.057	0.107
100	20	1.397	0.043	3.395	0.042	1.001	0.039	2.999	0.038	1.033	0.078	3.031	0.076
100	50	1.397	0.026	3.400	0.026	0.999	0.023	3.001	0.022	1.010	0.046	3.013	0.046
100	100	1.399	0.020	3.399	0.019	1.000	0.016	3.000	0.016	1.006	0.030	3.005	0.030
3	100	1.368	0.110	3.370	0.105	1.002	0.089	2.999	0.091	1.176	0.166	3.181	0.171
5	100	1.382	0.075	3.385	0.076	1.000	0.070	3.000	0.070	1.222	0.117	3.218	0.117
10	100	1.394	0.053	3.392	0.056	1.002	0.050	2.998	0.049	1.237	0.089	3.238	0.090
20	100	1.396	0.040	3.395	0.041	1.000	0.038	2.999	0.037	1.227	0.089	3.227	0.088
50	100	1.399	0.027	3.398	0.027	1.001	0.024	3.000	0.023	1.072	0.116	3.071	0.117

[a]Matlab programs are available from the author.

affect consistency. On the other hand, for small N (no matter how large is T), the estimates are inconsistent. The simulation results are consistent with those predictions.

Table V presents results for panel models with lagged dependent variables

$$Y_{it} = \rho Y_{i,t-1} + X_{it,1}\beta_1 + X_{it,2}\beta_2 + \lambda_i' F_t + \varepsilon_{it}$$
$$((\rho, \beta_1, \beta_2) = (0.75, 1, 3);\ t = 2, 3, \dots, T;\ i = 1, 2, \dots, N),$$

TABLE V

MODELS WITH LAGGED DEPENDENT VARIABLES AND HETEROKEDASTICITY

		Interactive-Effects Estimator						Within Estimator					
N	T	Mean $\rho=0.75$	SD	Mean $\beta_1=1$	SD	Mean $\beta_2=3$	SD	Mean $\rho=0.75$	SD	Mean $\beta_1=1$	SD	Mean $\beta_2=3$	SD
100	5	0.733	0.029	1.125	0.161	3.114	0.163	0.713	0.032	1.345	0.111	3.314	0.117
100	10	0.739	0.012	1.090	0.103	3.087	0.097	0.735	0.012	1.383	0.059	3.375	0.059
100	20	0.748	0.004	1.017	0.042	3.018	0.041	0.744	0.006	1.392	0.039	3.390	0.038
100	30	0.749	0.003	1.009	0.030	3.009	0.030	0.746	0.005	1.395	0.031	3.394	0.032
100	50	0.750	0.002	1.004	0.022	3.004	0.022	0.748	0.003	1.397	0.025	3.396	0.024
100	100	0.750	0.001	1.002	0.015	3.002	0.016	0.749	0.002	1.398	0.017	3.399	0.018
5	100	0.743	0.011	1.072	0.131	3.068	0.125	0.748	0.008	1.380	0.087	3.382	0.083
10	100	0.747	0.007	1.024	0.074	3.029	0.073	0.749	0.006	1.391	0.055	3.394	0.055
20	100	0.749	0.003	1.009	0.039	3.009	0.038	0.749	0.004	1.396	0.036	3.397	0.037
30	100	0.749	0.003	1.005	0.030	3.005	0.030	0.749	0.003	1.395	0.031	3.396	0.031
50	100	0.750	0.002	1.002	0.022	3.003	0.022	0.749	0.003	1.397	0.024	3.399	0.024

where $X_{it,1}$, $X_{it,2}$, λ_i, and F_t are generated as in previous tables. We also allow time serial heteroskedasticity in ε_{it}, with $\varepsilon_{it} \sim N(0, 1)$ for t odd and $\sim N(0, 4)$ for t even. For small T, the estimated parameters are biased primarily due to heteroskedasticity. Under large T, the estimator performs quite well. The within-group estimator for ρ is not heavily biased, but the estimators for β_1 and β_2 are biased for all combinations of N and T.

REFERENCES

MCLACHLAN, G. J., AND T. KRISHNAN (1996): *The EM Algorithm and Extensions*. New York: Wiley.

STOCK, J. H., AND M. W. WATSON (1998): "Diffusion Indexes," Working Paper 6702, NBER.

Dept. of Economics, New York University, 19 West 4th Street, New York, NY 10012, U.S.A., SEM, Tsinghua University, CEMA, Central University of Finance and Economics, Beijing, China; jushan.bai@nyu.edu.

Manuscript received October, 2005; final revision received January, 2009.