Online Appendix to *"Model Selection with Estimated Factors and Idiosyncratic Components"*

Jack Fosten*
University of East Anglia, UK

September 14, 2016

Abstract

This Online Appendix contains the proofs to Theorem 1 in the text, along with three Lemmas required in the proof. Additionally, there are some extra Monte Carlo and empirical results, which were mentioned but not reported in the main text for the sake of brevity.

1 Appendix A: Proofs of Main Results

Theorem 1. *Let Assumptions 1-5 hold and let the factors and factor loadings be estimated by Principal Components. For two models i and j, if model i corresponds to the true model such that the probability limit of $\left(\widehat{F}^i, \widehat{u}^i\right)$ is $\left(H^0 F_t^0, u_t^0\right)$ for all t, and for model j one or both of $\widehat{F}_t^j$ and $\widehat{u}_t^j$ has different probability limit, then:*

$$\lim_{N,T\to\infty} \Pr\left(IC\left(\widehat{F}^j, \widehat{u}^j\right) < IC\left(\widehat{F}^i, \widehat{u}^i\right)\right) = 0$$

as long as (i) $g(N,T) \to 0$ and (ii) $\min\left\{\sqrt{T}, N\right\} g(N,T) \to \infty$ as $N, T \to \infty$.

The proof of Theorem 1 makes use of the following three Lemmas on estimation error in the idiosyncratic components. As mentioned in the text, the number of factors, r, in the first-stage factor model for X_t is assumed to be fixed and known for these proofs.

Lemma 1. *Let Assumptions 1-5 hold and let the factors and idiosyncratic errors be estimated by Principal Components. Then for each $i = 1, ..., N$, and as $N, T \to \infty$:*

$$\frac{1}{T}\sum_{t=1}^{T} F_t\left(\widehat{u}_{it} - u_{it}\right) = O_p\left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{T}}\right\}\right) \tag{1}$$

Lemma 2. *Let Assumptions 1-5 hold and let the factors and idiosyncratic errors be estimated by Principal Components. Then for each $i = 1, ..., N$, and as $N, T \to \infty$:*

*Correspondence to: School of Economics, University of East Anglia, Norwich, NR4 7TJ, UK. E-mail address: j.fosten@uea.ac.uk. Telephone: +44 (0) 1603 59 1243.

$$\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{u}_{it}-u_{it}\right)\varepsilon_{t+h}=O_p\left(\max\left\{\frac{1}{N},\frac{1}{T}\right\}\right) \tag{2}$$

Lemma 3. *Let Assumptions 1-5 hold and let the factors and idiosyncratic errors be estimated by Principal Components. Then for each $i=1,...,N$, and as $N,T\to\infty$:*

$$\frac{1}{T}\sum_{t=1}^{T}u_{it}\left(\widehat{u}_{it}-u_{it}\right)=O_p\left(\max\left\{\frac{1}{N},\frac{1}{T}\right\}\right) \tag{3}$$

Each of these Lemmas makes use of the following identity:

$$\begin{aligned}
\widehat{u}_{it}-u_{it} &= \left(X_{it}-\widehat{\lambda}_i'\widehat{F}_t\right)-\left(X_{it}-\lambda_i'F_t\right)\\
&= \lambda_i'H^{-1}HF_t-\widehat{\lambda}_i'\widehat{F}_t\\
&= \lambda_i'H^{-1}\left(HF_t-\widehat{F}_t\right)+\left(H'^{-1}\lambda_i-\widehat{\lambda}_i\right)'\widehat{F}_t\\
&= \lambda_i'H^{-1}\left(HF_t-\widehat{F}_t\right)+\left(H'^{-1}\lambda_i-\widehat{\lambda}_i\right)'HF_t\\
&- \left(H'^{-1}\lambda_i-\widehat{\lambda}_i\right)'\left(HF_t-\widehat{F}_t\right)
\end{aligned} \tag{4}$$

where H is the rotation matrix described in Bai (2003) and Bai and Ng (2006).

Proof of Lemma 1. We can use Equation (4) to write for any $i=1,...,N$:

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}F_t\left(\widehat{u}_{it}-u_{it}\right) &= \frac{1}{T}\sum_{t=1}^{T}F_t\lambda_i'H^{-1}\left(HF_t-\widehat{F}_t\right)+\frac{1}{T}\sum_{t=1}^{T}F_t\left(H'^{-1}\lambda_i-\widehat{\lambda}_i\right)'HF_t\\
&-\frac{1}{T}\sum_{t=1}^{T}F_t\left(H'^{-1}\lambda_i-\widehat{\lambda}_i\right)'\left(HF_t-\widehat{F}_t\right)\\
&=\left(\frac{1}{T}\sum_{t=1}^{T}F_t\left(HF_t-\widehat{F}_t\right)'\right)H'^{-1}\lambda_i+\left(\frac{1}{T}\sum_{t=1}^{T}F_tF_t'\right)H'\left(H'^{-1}\lambda_i-\widehat{\lambda}_i\right)\\
&-\left(\frac{1}{T}\sum_{t=1}^{T}F_t\left(HF_t-\widehat{F}_t\right)'\right)\left(H'^{-1}\lambda_i-\widehat{\lambda}_i\right)
\end{aligned}$$

By Assumptions 1-5, the first term is $O_p\left(\max\left\{\frac{1}{N},\frac{1}{T}\right\}\right)$ using Lemma A.1 of Bai and Ng (2006) and as $H'^{-1}\lambda_i$ is $O_p(1)$. We also know that the final term is of smaller order since $\left(\widehat{\lambda}_i-H'^{-1}\lambda_i\right)$ is $o_p(1)$ by Bai (2003) Theorem 2. However, for the middle term:

$$\begin{aligned}
\left(\frac{1}{T}\sum_{t=1}^{T}F_tF_t'\right)H'\left(H'^{-1}\lambda_i-\widehat{\lambda}_i\right) &= O_p(1)\times\left(H'^{-1}\lambda_i-\widehat{\lambda}_i\right)\\
&= O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(\max\left\{\frac{1}{N},\frac{1}{T}\right\}\right)
\end{aligned}$$

Since Bai (2003) Theorem 2 shows that for the Principal Components estimator of $\widehat{\lambda}_i$ we have:

$$\widehat{\lambda}_i - H'^{-1}\lambda_i = H\frac{1}{T}\sum_{t=1}^{T} F_t u_{it} + O_p\left(\max\left\{\frac{1}{N}, \frac{1}{T}\right\}\right)$$

Now since we do not place any restriction on the relative rate of increase of T and N, $O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\max\left\{\frac{1}{N}, \frac{1}{T}\right\}\right) = O_p\left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{T}}\right\}\right)$ and therefore:

$$\frac{1}{T}\sum_{t=1}^{T} F_t\left(\widehat{u}_{it} - u_{it}\right) = O_p\left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{T}}\right\}\right)$$

as required. ■

Proof of Lemma 2. We can use Equation (4) to write for any $i = 1, ..., N$ that:

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{u}_{it} - u_{it}\right)\varepsilon_{t+h} &= \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t+h}\lambda_i' H^{-1}\left(HF_t - \widehat{F}_t\right) + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t+h}\left(H'^{-1}\lambda_i - \widehat{\lambda}_i\right)' HF_t \\
&- \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t+h}\left(H'^{-1}\lambda_i - \widehat{\lambda}_i\right)'\left(HF_t - \widehat{F}_t\right) \\
&= \lambda_i' H^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\left(HF_t - \widehat{F}_t\right)\varepsilon_{t+h}\right) + \left(H'^{-1}\lambda_i - \widehat{\lambda}_i\right)' H\left(\frac{1}{T}\sum_{t=1}^{T} F_t\varepsilon_{t+h}\right) \\
&- \left(H'^{-1}\lambda_i - \widehat{\lambda}_i\right)'\left(\frac{1}{T}\sum_{t=1}^{T}\left(HF_t - \widehat{F}_t\right)\varepsilon_{t+h}\right)
\end{aligned}$$

By Assumptions 1-5, the first part is $O_p\left(\max\left\{\frac{1}{N}, \frac{1}{T}\right\}\right)$ by Lemma A.1 of Bai and Ng (2006) and as $\lambda_i' H^{-1} = O_p(1)$. The last part is of smaller order because $\widehat{\lambda}_i - H'^{-1}\lambda_i = o_p(1)$ by Bai (2003) Theorem 2. For the middle part,

$$\left(H'^{-1}\lambda_i - \widehat{\lambda}_i\right)' H\left(\frac{1}{T}\sum_{t=1}^{T} F_t\varepsilon_{t+h}\right) = \left[O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\max\left\{\frac{1}{N}, \frac{1}{T}\right\}\right)\right] \times O_p\left(\frac{1}{\sqrt{T}}\right)$$

since $\mathrm{E}[F_t\varepsilon_{t+h}] = 0$ implies that $\left(\frac{1}{T}\sum_{t=1}^{T} F_t\varepsilon_{t+h}\right) = O_p\left(\frac{1}{\sqrt{T}}\right)$ and as $\left(\widehat{\lambda}_i - H'^{-1}\lambda_i\right) = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\max\left\{\frac{1}{N}, \frac{1}{T}\right\}\right)$ by Bai (2003) Theorem 2. This part is therefore $O_p\left(\max\left\{\frac{1}{\sqrt{T}N}, \frac{1}{T}\right\}\right)$. Combining these three results shows that

$$\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{u}_{it} - u_{it}\right)\varepsilon_{t+h} = O_p\left(\max\left\{\frac{1}{N}, \frac{1}{T}\right\}\right)$$

as required. ■

Proof of Lemma 3. In a similar way to Lemmas 1 and 2, we can write for any $i = 1, ..., N$:

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} u_{it}\left(\widehat{u}_{it} - u_{it}\right) &= \frac{1}{T}\sum_{t=1}^{T} u_{it}\lambda_i' H^{-1}\left(HF_t - \widehat{F}_t\right) + \frac{1}{T}\sum_{t=1}^{T} u_{it}\left(H'^{-1}\lambda_i - \widehat{\lambda}_i\right)' HF_t \\
&- \frac{1}{T}\sum_{t=1}^{T} u_{it}\left(H'^{-1}\lambda_i - \widehat{\lambda}_i\right)'\left(HF_t - \widehat{F}_t\right) \\
&= \left(\frac{1}{T}\sum_{t=1}^{T} u_{it}\left(HF_t - \widehat{F}_t\right)'\right) H'^{-1}\lambda_i + \left(\frac{1}{T}\sum_{t=1}^{T} F_t' u_{it}\right) H'\left(H'^{-1}\lambda_i - \widehat{\lambda}_i\right) \\
&- \left(\frac{1}{T}\sum_{t=1}^{T} u_{it}\left(HF_t - \widehat{F}_t\right)'\right)\left(H'^{-1}\lambda_i - \widehat{\lambda}_i\right)
\end{aligned}$$

Now in this case, the first term is $O_p\left(\max\left\{\frac{1}{N}, \frac{1}{T}\right\}\right)$ by Assumptions 1-5 and Bai (2003) Lemma B.1, the final part is of smaller order as $\widehat{\lambda}_i - H'^{-1}\lambda_i$ is $o_p(1)$ and for the middle part we have:

$$\left(\frac{1}{T}\sum_{t=1}^{T} F_t' u_{it}\right) H'\left(H'^{-1}\lambda_i - \widehat{\lambda}_i\right) = O_p\left(\frac{1}{\sqrt{T}}\right) \times \left[O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\max\left\{\frac{1}{N}, \frac{1}{T}\right\}\right)\right]$$

since $\mathrm{E}[F_t' u_{it}] = 0$ implies that $\left(\frac{1}{T}\sum_{t=1}^{T} F_t' u_{it}\right) = O_p\left(\frac{1}{\sqrt{T}}\right)$ and $\left(\widehat{\lambda}_i - H'^{-1}\lambda_i\right) = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\max\left\{\frac{1}{N}, \frac{1}{T}\right\}\right)$ from Bai (2003) Theorem 2. This term is therefore $O_p\left(\max\left\{\frac{1}{\sqrt{T}N}, \frac{1}{T}\right\}\right)$. Combining the three results gives:

$$\frac{1}{T}\sum_{t=1}^{T} u_{it}\left(\widehat{u}_{it} - u_{it}\right) = O_p\left(\max\left\{\frac{1}{N}, \frac{1}{T}\right\}\right)$$

as required. ■

Proof of Theorem 1. We have two regression specifications i and j:

$$y_{t+h} = \beta^{i\prime}\widehat{F}_t^i + \alpha^{i\prime}\widehat{u}_t^i + \widehat{\varepsilon}_{t+h}^i$$

with $\widehat{F}_t^i$ and β^i of dimension $r_i \times 1$ and α^i and $\widehat{u}_t^i$ of dimension $m_i \times 1$; and:

$$y_{t+h} = \beta^{j\prime}\widehat{F}_t^j + \alpha^{j\prime}\widehat{u}_t^j + \widehat{\varepsilon}_{t+h}^j$$

with $\widehat{F}_t^j$ and β^j of dimension $r_j \times 1$ and α^j and $\widehat{u}_t^j$ of dimension $m_j \times 1$. For simplicity we will write these compactly as:

$$y_{t+h} = \theta^{i\prime}\widehat{Z}_t^i + \widehat{\varepsilon}_{t+h}^i$$

and

$$y_{t+h} = \theta^{j\prime}\widehat{Z}_t^j + \widehat{\varepsilon}_{t+h}^j$$

Where $\widehat{Z}_t^i = \left[\widehat{F}_t^{i\prime}, \widehat{u}_t^{i\prime}\right]^\prime$ and $\widehat{Z}_t^j = \left[\widehat{F}_t^{j\prime}, \widehat{u}_t^{j\prime}\right]^\prime$. Now in what follows we will relate the generated regressors $\widehat{F}_t^i$ and $\widehat{u}_t^i$ to their probability limits $H^i F_t^i$ and u_t^i where H^i is the relevant submatrix of H described in Bai and Ng (2006) and u_t^i is not rotated as the common component is identified without rotation. This gives the probability limit vectors $Z_t^i = \left[F_t^{i\prime} H^{i\prime}, u_t^{i\prime}\right]^\prime$ and $Z_t^j = \left[F_t^{j\prime} H^{j\prime}, u_t^{j\prime}\right]^\prime$.

Now we can rewrite the sum of square error functions $V(.)$ both for the estimated factor and idiosyncratic components, and for their probability limits as:

$$
\begin{aligned}
V\left(\widehat{F}^i, \widehat{u}^i\right) = V\left(\widehat{Z}^i\right) &= \frac{1}{T} y^\prime \widehat{M}^i y \\
V\left(\widehat{F}^j, \widehat{u}^j\right) = V\left(\widehat{Z}^j\right) &= \frac{1}{T} y^\prime \widehat{M}^j y \\
V\left(F^i H^{i\prime}, u^i\right) = V\left(Z^i\right) &= \frac{1}{T} y^\prime M^i y \\
V\left(F^j H^{j\prime}, u^j\right) = V\left(Z^j\right) &= \frac{1}{T} y^\prime M^j y
\end{aligned}
$$

where $\widehat{M}^i = I - \widehat{Z}^i \left(\widehat{Z}^{i\prime} \widehat{Z}^i\right)^{-1} \widehat{Z}^{i\prime}$, $M^i = I - Z^i \left(Z^{i\prime} Z^i\right)^{-1} Z^{i\prime}$ and similarly for $\widehat{M}^j$ and M^j. Therefore we can rewrite the statement in Theorem 1 as:

$$
\lim_{N,T\to\infty} \Pr\left(\ln\left[\frac{\frac{1}{T} y^\prime \widehat{M}^j y}{\frac{1}{T} y^\prime \widehat{M}^i y}\right] < \left(r_i + m_i - r_j - m_j\right) g(N,T)\right) = 0
$$

which we can manipulate in order to relate the estimated regressors back to the true (rotated) factors and idiosyncratic errors as follows:

$$
\lim_{N,T\to\infty} \Pr\left(\ln\left[\frac{\frac{1}{T} y^\prime M^j y}{\frac{1}{T} y^\prime M^i y}\right] + \ln\left[\frac{\frac{1}{T} y^\prime \widehat{M}^j y}{\frac{1}{T} y^\prime M^j y}\right] - \ln\left[\frac{\frac{1}{T} y^\prime \widehat{M}^i y}{\frac{1}{T} y^\prime M^i y}\right] < \left(r_i + m_i - r_j - m_j\right) g(N,T)\right) = 0 \tag{5}
$$

The second and third terms on the left of this expression are both estimation error terms involving the estimated factors and idiosyncratic components. We first show the convergence rate of these two terms as T and N grow large, using Lemmas 1-3. This proof deviates from those in Bai and Ng (2006) and Groen and Kapetanios (2013) as the matrices $\widehat{M}^i$ and $\widehat{M}^j$ do not just contain estimated factors, they additionally contain estimation error due to the idiosyncratic errors.

Consider the third term of Equation (5) for the model specification i. (That for j will follow the same argument).

$$
\begin{aligned}
\frac{1}{T} y^\prime \widehat{M}^i y - \frac{1}{T} y^\prime M^i y &= \frac{1}{T} y^\prime \left(\widehat{M}^i - M^i\right) y \\
&= \frac{1}{T} y^\prime \left(P^i - \widehat{P}^i\right) y
\end{aligned}
$$

where $\widehat{P}^i$ and P^i are projection matrices $\widehat{P}^i = \widehat{Z}^i \left(\widehat{Z}^{i\prime} \widehat{Z}^i\right)^{-1} \widehat{Z}^{i\prime}$ and $P^i = Z^i \left(Z^{i\prime} Z^i\right)^{-1} Z^{i\prime}$. Now make the following expansion:

$$\begin{aligned}\frac{1}{T}y'\left(P^i-\widehat{P}^i\right)y &= \frac{1}{T}y'\left(Z^i\left(Z^{i\prime}Z^i\right)^{-1}Z^{i\prime}-\widehat{Z}^i\left(\widehat{Z}^{i\prime}\widehat{Z}^i\right)^{-1}\widehat{Z}^{i\prime}\right)y\\
&= \frac{1}{T}y'\left(Z^i\left(Z^{i\prime}Z^i\right)^{-1}Z^{i\prime}\right.\\
&\left.-\left(\widehat{Z}^i-Z^i+Z^i\right)\left(\widehat{Z}^{i\prime}\widehat{Z}^i\right)^{-1}\left(\widehat{Z}^i-Z^i+Z^i\right)'\right)y\\
&= \frac{1}{T}y'Z^i\left[\left(Z^{i\prime}Z^i\right)^{-1}-\left(\widehat{Z}^{i\prime}\widehat{Z}^i\right)^{-1}\right]Z^{i\prime}y\\
&-\frac{1}{T}y'\left(\widehat{Z}^i-Z^i\right)\left(\widehat{Z}^{i\prime}\widehat{Z}^i\right)^{-1}Z^{i\prime}y\\
&-\frac{1}{T}y'Z^i\left(\widehat{Z}^{i\prime}\widehat{Z}^i\right)^{-1}\left(\widehat{Z}^i-Z^i\right)'y\\
&-\frac{1}{T}y'\left(\widehat{Z}^i-Z^i\right)\left(\widehat{Z}^{i\prime}\widehat{Z}^i\right)^{-1}\left(\widehat{Z}^i-Z^i\right)'y\\
&= I+II+III+IV\end{aligned}$$

Consider part II:

$$\begin{aligned}II &= \frac{1}{T}y'\left(\widehat{Z}^i-Z^i\right)\left(\widehat{Z}^{i\prime}\widehat{Z}^i\right)^{-1}Z^{i\prime}y\\
&= \frac{1}{T}y'\left(\widehat{Z}^i-Z^i\right)\left(\frac{1}{T}\widehat{Z}^{i\prime}\widehat{Z}^i\right)^{-1}\frac{1}{T}Z^{i\prime}y\end{aligned}$$

The product of the second and last part of this expression $\left(\frac{1}{T}\widehat{Z}^{i\prime}\widehat{Z}^i\right)^{-1}\frac{1}{T}Z^{i\prime}y$ gives a column vector of dimension $(r_i+m_i)\times 1$ which is $O_p(1)$. The first part is a $1\times(r_i+m_i)$ row vector. Now re-write this part in terms of the estimates $\widehat{F}_t^i$ and $\widehat{u}_t^i$ using $\widehat{Z}_t^i=\left[\widehat{F}_t^{i\prime},\widehat{u}_t^{i\prime}\right]'$, or in matrix form $\widehat{Z}^i=\left[\widehat{F}^i,\widehat{u}^i\right]$, and substituting in the true model for y we have:

$$\begin{aligned}\frac{1}{T}y'\left(\widehat{Z}^i-Z^i\right) &= \frac{1}{T}y'\left[\left(\widehat{F}^i-F^iH^{i\prime}\right),\left(\widehat{u}^i-u^i\right)\right]\\
&= \frac{1}{T}\left(F^0\beta^0+u^0\alpha^0+\varepsilon\right)'\left[\left(\widehat{F}^i-F^iH^{i\prime}\right),\left(\widehat{u}^i-u^i\right)\right]\end{aligned}\tag{6}$$

Now the first r_i elements of this row vector corresponding to factor estimation error are results which have already been shown in the literature, namely:

$$\begin{aligned}\beta^{0\prime}\frac{1}{T}F^{0\prime}\left(\widehat{F}^i-F^iH^i\right) &= \beta^{0\prime}\frac{1}{T}\sum_{t=1}^{T}F_t^0\left(\widehat{F}_t^i-H^iF_t^i\right)'\\
&= O_p\left(\max\left\{\frac{1}{N},\frac{1}{T}\right\}\right)\end{aligned}\tag{7}$$

and

$$\frac{1}{T}\varepsilon'\left(\widehat{F}^i - F^i H^i\right) = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t+h}\left(\widehat{F}_t^i - H^i F_t^i\right)'$$
$$= O_p\left(\max\left\{\frac{1}{N},\frac{1}{T}\right\}\right) \tag{8}$$

both by Bai and Ng (2006) Lemmas A.1 (ii) and A.1 (vi), and

$$\alpha^{0\prime}\frac{1}{T}u^{0\prime}\left(\widehat{F}^i - F^i H^i\right) = \alpha^{0\prime}\frac{1}{T}\sum_{t=1}^{T}u_t^0\left(\widehat{F}_t^i - H^i F_t^i\right)'$$
$$= O_p\left(\max\left\{\frac{1}{N},\frac{1}{T}\right\}\right) \tag{9}$$

by Bai (2003) Lemma B.1 since the dimension of u_t^0 is m^0 which is finite. However, the remaining three terms are new to this paper and are analysed in the Lemmas above, namely:

$$\beta^{0\prime}\frac{1}{T}F^{0\prime}\left(\widehat{u}^i - u^i\right) = \beta^{0\prime}\frac{1}{T}\sum_{t=1}^{T}F_t^0\left(\widehat{u}_t^i - u_t^i\right)'$$
$$= O_p\left(\max\left\{\frac{1}{N},\frac{1}{\sqrt{T}}\right\}\right) \tag{10}$$

and

$$\frac{1}{T}\varepsilon'\left(\widehat{u}^i - u^i\right) = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t+h}\left(\widehat{u}_t^i - u_t^i\right)'$$
$$= O_p\left(\max\left\{\frac{1}{N},\frac{1}{T}\right\}\right) \tag{11}$$

and finally:

$$\alpha^{0\prime}\frac{1}{T}u^{0\prime}\left(\widehat{u}^i - u^i\right) = \alpha^{0\prime}\frac{1}{T}\sum_{t=1}^{T}u_t^0\left(\widehat{u}_t^i - u_t^i\right)'$$
$$= O_p\left(\max\left\{\frac{1}{N},\frac{1}{T}\right\}\right) \tag{12}$$

The same results are used in showing the other parts I, III and IV and as such we do not repeat them here. Combining all of these results yields:

$$\frac{1}{T}y'\widehat{M}^i y - \frac{1}{T}y' M^i y = O_p\left(\max\left\{\frac{1}{N},\frac{1}{\sqrt{T}}\right\}\right)$$

This differs to that of Groen and Kapetanios (2013), due to the consistency rate in Lemma 1. This result implies:

$$\frac{\frac{1}{T}y'\widehat{M^i}y}{\frac{1}{T}y'M^iy} = 1 + O_p\left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{T}}\right\}\right)$$

which in turn implies:

$$\ln\left[\frac{\frac{1}{T}y'\widehat{M^i}y}{\frac{1}{T}y'M^iy}\right] = O_p\left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{T}}\right\}\right)$$

And the same result holds for model j. Therefore we can rewrite the expression in Equation (5) to be:

$$\lim_{N,T\to\infty} \Pr\left(\ln\left[\frac{\frac{1}{T}y'M^jy}{\frac{1}{T}y'M^iy}\right] + O_p\left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{T}}\right\}\right) < (r_i + m_i - r_j - m_j)\, g(N,T)\right) = 0 \tag{13}$$

To show Theorem 1, we assume that model i is correct and that the probability limits of $\widehat{F}_t^i$ and $\widehat{u}_t^i$ are $H^0F_t^0$ and u_t^0 for all t, which because we impose orthogonality on the factors, means that model i contains the true number of variables with $r_i = r^0$ and $m_i = m^0$. This means that $M^iF^0 = 0$ and $M^iu^0 = 0$ so as in Groen and Kapetanios (2013), the denominator of the first part in (13) becomes:

$$\begin{aligned}\frac{1}{T}y'M^iy &= \frac{1}{T}\varepsilon'M^i\varepsilon \\ &= \sigma_\varepsilon^2 + O_p\left(\frac{1}{T}\right)\end{aligned}$$

The last line assumes homoskedasticity of ε_t, as in Cheng and Hansen (2015) for the Mallows criterion, and Groen and Kapetanios (2013) for the related information criteria. To assess statement (13) we now take two exhaustive cases in which the candidate model j is incorrectly specified:

Case 1: *The probability limits of $\widehat{F}_t^j$ and $\widehat{u}_t^j$ are such that: (i) $M^jF^0 = 0$ but $M^ju^0 \neq 0$ (Model j has correct factor specification but not all relevant idiosyncratic errors are included), (ii) $M^jF^0 \neq 0$ but $M^ju^0 = 0$ (not all relevant factors are included, but all relevant idiosyncratic errors are included) or (iii) $M^jF^0 \neq 0$ and $M^ju^0 \neq 0$ (model missing relevant factors and relevant idiosyncratic errors).*

In any of these three cases (i)-(iii), the numerator of the first term in Equation (13) is:

$$\begin{aligned}\frac{1}{T}y'M^jy &= \frac{1}{T}\varepsilon'M^j\varepsilon + \frac{1}{T}\left(F^0\beta^0 + u^0\alpha^0\right)' M^j \left(F^0\beta^0 + u^0\alpha^0\right) \\ &= \sigma_\varepsilon^2 + \tau_1 + O_p\left(\frac{1}{T}\right)\end{aligned}$$

where $\tau_1 > 0$ and the form of τ_1 depends on whether we are in Case 1 (i), (ii) or (iii). Therefore:

$$\frac{1}{T}y'M^jy - \frac{1}{T}y'M^iy = \tau_1 + O_p\left(\frac{1}{T}\right)$$

which implies that

$$\ln\left[\frac{\frac{1}{T}y'M^jy}{\frac{1}{T}y'M^iy}\right] \geq \tau_2 > 0$$

for some known τ_2. Therefore using Equation (13), the statement in Theorem 1 will hold in Case 1 as long as we can show that:

$$\lim_{N,T\to\infty} \Pr\left(\tau_2 + O_p\left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{T}}\right\}\right) < (r_i + m_i - r_j - m_j)\, g(N,T)\right) = 0 \tag{14}$$

and since $(r_i + m_i - r_j - m_j)$ is finite, this statement is true when $g(N,T) \to 0$. Since this is stated in Condition (i) of Theorem 1, this proves what was required in Case 1. Turning to Case 2:

Case 2: *The probability limits of $\widehat{F}_t^j$ and $\widehat{u}_t^j$ are such that both $M^jF^0 = 0$ and $M^ju^0 = 0$, but more than the relevant variables are included with either (i) $r_j = r^0$ but $m_j > m^0$ (correct factor specification but too many idiosyncratic errors included), (ii) $r_j > r^0$ but $m_j = m^0$ (too many factors specified, but correct idiosyncratic error specification) or (iii) $r_j > r^0$ and $m_j > m^0$ (too many factors and idiosyncratic errors included).*

In this case, the numerator of the first term in (13) is:

$$\begin{aligned}\frac{1}{T}y'M^jy &= \frac{1}{T}\varepsilon'M^j\varepsilon + \frac{1}{T}\left(F^0\beta^0 + u^0\alpha^0\right)'M^j\left(F^0\beta^0 + u^0\alpha^0\right) \\ &= \sigma_\varepsilon^2 + O_p\left(\frac{1}{T}\right)\end{aligned}$$

Therefore:

$$\frac{1}{T}y'M^jy - \frac{1}{T}y'M^iy = O_p\left(\frac{1}{T}\right)$$

which implies that:

$$\ln\left[\frac{\frac{1}{T}y'M^jy}{\frac{1}{T}y'M^iy}\right] = O_p\left(\frac{1}{T}\right)$$

Therefore using (13), the statement in Theorem 1 will hold in Case 2 as long as we can show that:

$$\lim_{N,T\to\infty} \Pr\left(O_p\left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{T}}\right\}\right) < (r_i + m_i - r_j - m_j)\, g(N,T)\right) = 0 \tag{15}$$

Now since model i is the correct model with $r_i = r^0$ and $m_i = m^0$, each of Case 2 (i), (ii) and (iii) imply that $(r_i + m_i - r_j - m_j) < 0$. Therefore this statement holds when $g(N,T)\min\left\{\sqrt{T}, N\right\} \to \infty$ and the right hand side diverges to $-\infty$ at a quicker rate than the estimation error. Since this corresponds to Condition (ii) stated in Theorem 1, this shows what was required in Case 2. ■

Finally, we also offer a short proof of the statement in Remark 1, which follows immediately as a special case of the proof of Theorem 1.

Proof of Remark 1. For the case where $\beta^0 = 0$, it is immediate to see that the expression in

Equation (6) simplifies to:

$$\begin{aligned}
\frac{1}{T} y' \left(\widehat{Z}^i - Z^i \right) &= \frac{1}{T} y' \left[\left(\widehat{F}^i - F^i H^{i\prime} \right), \left(\widehat{u}^i - u^i \right) \right] \\
&= \frac{1}{T} \left(u^0 \alpha^0 + \varepsilon \right)' \left[\left(\widehat{F}^i - F^i H^{i\prime} \right), \left(\widehat{u}^i - u^i \right) \right]
\end{aligned}$$

This, in turn, implies that the terms analysed in Equations (7) and (10) of the proof of Theorem 1 disappear, leaving only the terms in Equations (8), (9), (11) and (12). Now, since all of these terms have the rate $O_p\left(\max\left\{\frac{1}{N}, \frac{1}{T}\right\}\right)$, it follows that we can replace the $O_p\left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{T}}\right\}\right)$ rate in Equations (13), (14) and (15), which leads directly to Condition (ii') displayed in Remark 1. ■

2 Appendix B: Additional Monte Carlo Results

2.1 MSD Results for Specifications $r = 1$, $m^0 = 1, 2$ in the Text

Figure A1: MSD^u for different information criteria over 1,000 Monte Carlo replications when the true $r = 1$ and $m^0 = 1$

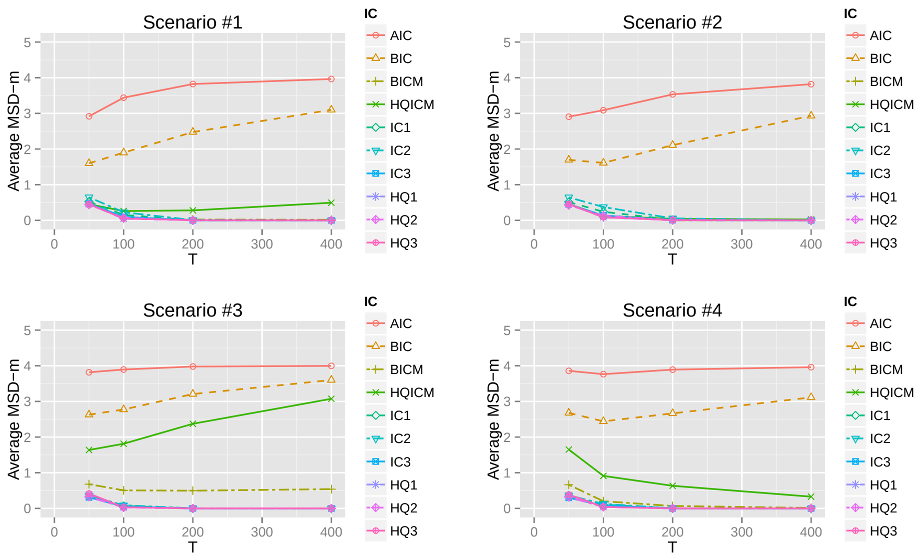


Figure A2: MSD^u for different information criteria over 1,000 Monte Carlo replications when the true $r = 1$ and $m^0 = 2$

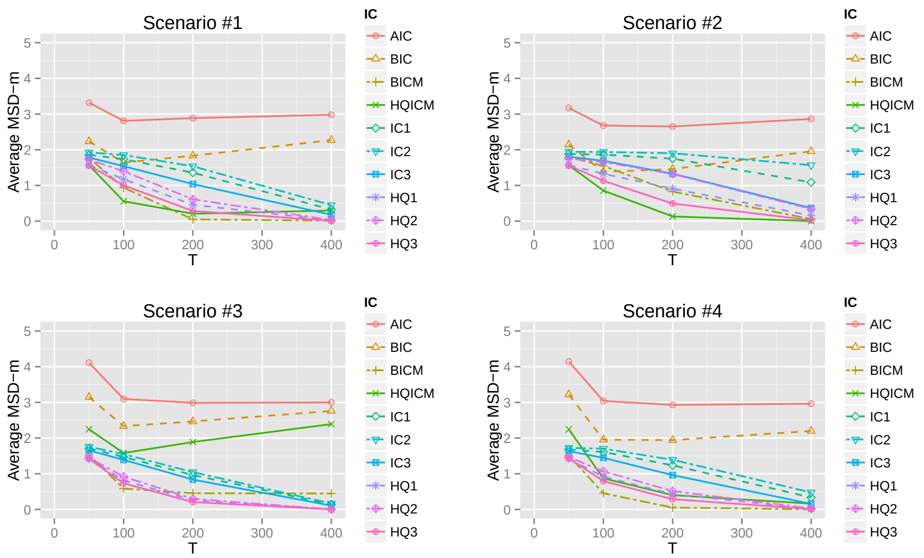


Figure A3: MSD^r for different information criteria over 1,000 Monte Carlo replications when the true $r = 1$ and $m^0 = 1$

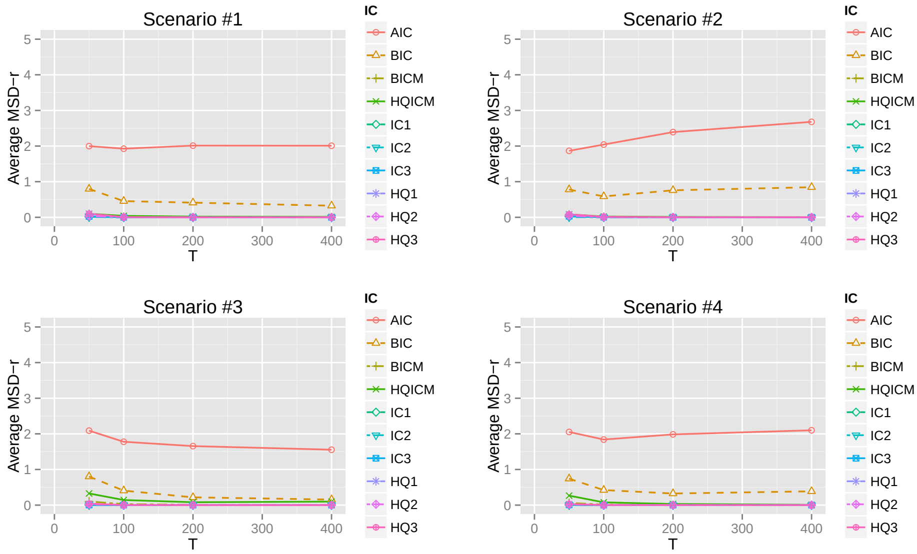
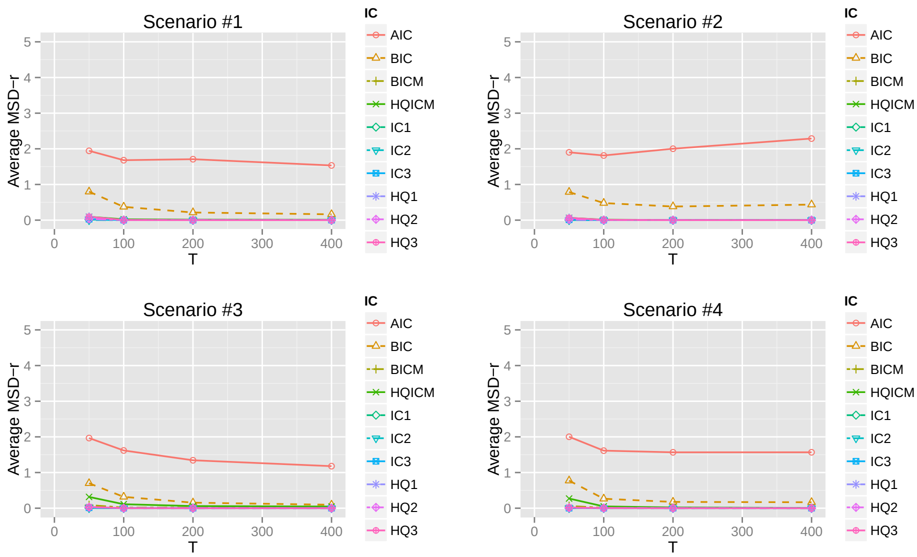
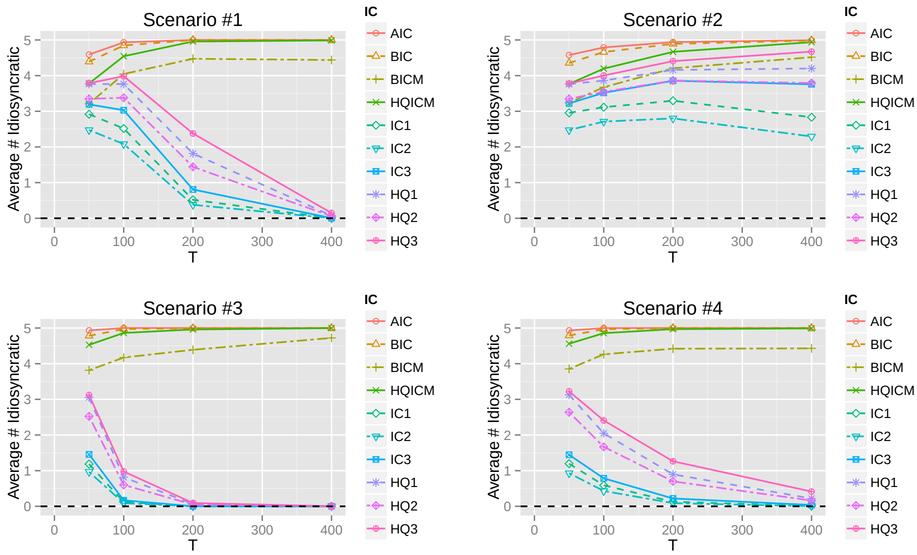


Figure A4: MSD^r for different information criteria over 1,000 Monte Carlo replications when the true $r = 1$ and $m^0 = 2$



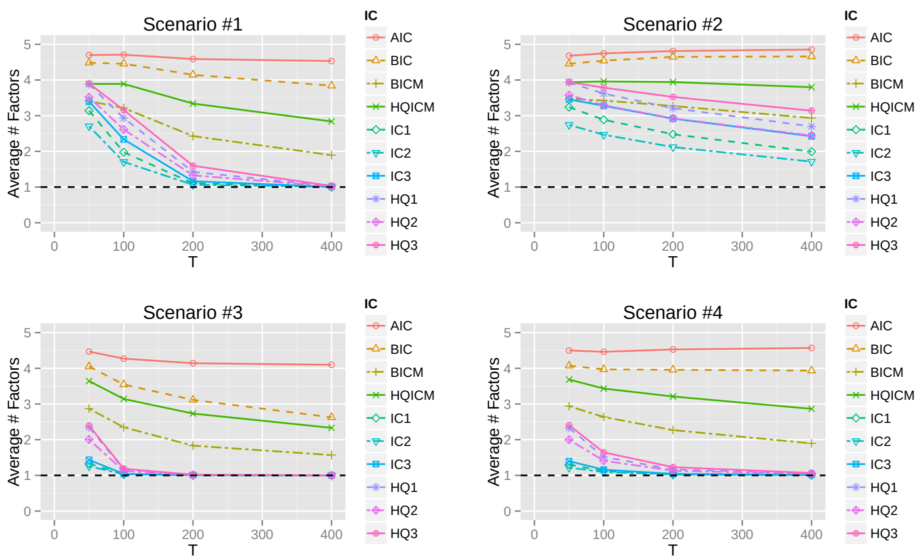
2.2 Additional Results for $r = 1$, $m^0 = 0$

Figure A5: Average number of selected idiosyncratic components ($\widehat{m}^0$) for different information criteria over 1,000 Monte Carlo replications when the true $r = 1$ and $m^0 = 0$



Notes: The horizontal dashed line represents the true number of idiosyncratic components $m^0 = 0$.

Figure A6: Average number of selected factors ($\widehat{r}$) for different information criteria over 1,000 Monte Carlo replications when the true $r = 1$ and $m^0 = 0$



Notes: The horizontal dashed line represents the true number of factors $r = 1$.

Figure A7: MSD^u for different information criteria over 1,000 Monte Carlo replications when the true $r = 1$ and $m^0 = 0$

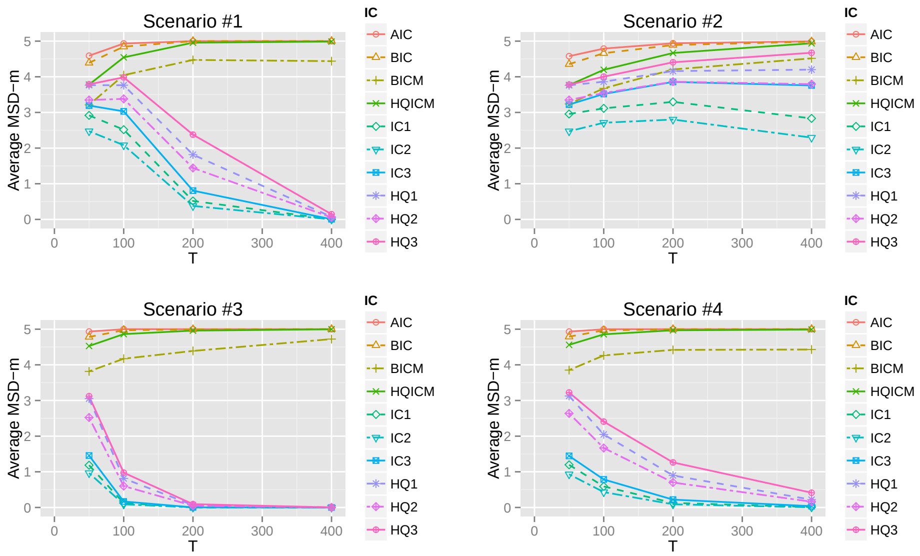
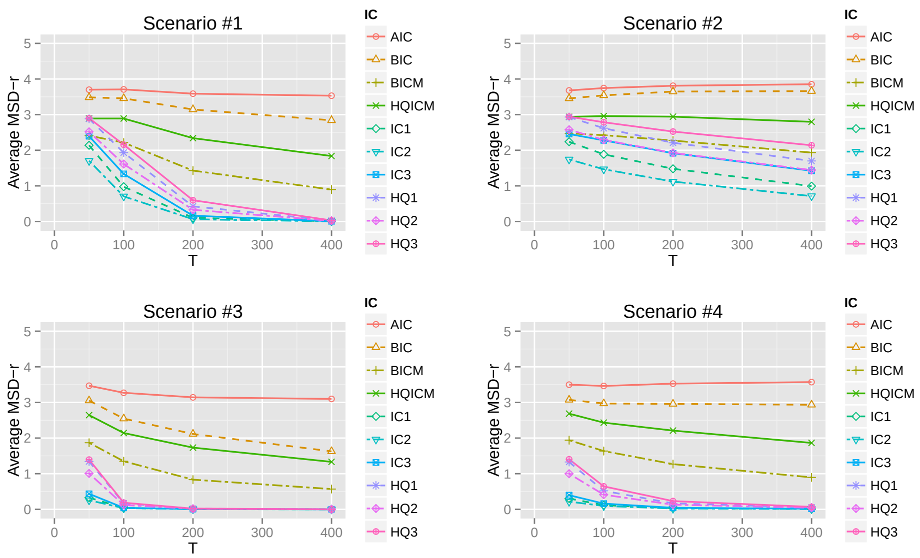


Figure A8: MSD^r for different information criteria over 1,000 Monte Carlo replications when the true $r = 1$ and $m^0 = 0$



3 Appendix C: Additional Empirical Results

3.1 In-Sample Results with Other Selection Criteria

Tables A1 and A2, below, correspond to Tables 3 and 4 in the text, but where model selection has taken place using the $HQICM$ Criterion of Groen and Kapetanios (2013):

Table A1: In-Sample Model Selection Results: 1988-1998 pre-Euro era.

	Selection Criterion: $HQICM$	
	$h = 1$	$h = 12$
	Countries ($\widehat{m}$)	Countries ($\widehat{m}$)
AUS	- (0)	CHE (1)
AUT	- (0)	- (0)
BEL	- (0)	DNK (1)
CAN	- (0)	FIN, ESP, BEL (3)
DNK	- (0)	BEL (1)
FIN	- (0)	- (0)
FRA	- (0)	- (0)
DEU	- (0)	- (0)
ITA	- (0)	FIN, CAN (2)
JPN	- (0)	ITA, BEL (2)
KOR	- (0)	DNK, ESP, CHE (3)
NLD	- (0)	- (0)
NOR	- (0)	- (0)
ESP	- (0)	FIN (1)
SWE	- (0)	FIN (1)
CHE	- (0)	- (0)
GBR	- (0)	FIN (1)

Notes: For each country, the column "Countries ($\widehat{m}$)" displays the identity of selected spillover countries, and the number of these selected countries in parentheses. Model selection is performed using the $HQICM$ criterion of Groen and Kapetanios (2013). The results are displayed only for the horizons $h = 1$ and $h = 12$, with other results available upon request.

Table A2: In-Sample Model Selection Results: 1999-2015 post-Euro era.

	Selection Criterion: $HQICM$	
	$h = 1$	$h = 12$
	Countries ($\widehat{m}$)	Countries ($\widehat{m}$)
AUS	- (0)	- (0)
CAN	- (0)	- (0)
DNK	- (0)	- (0)
EUR	- (0)	- (0)
JPN	- (0)	- (0)
KOR	- (0)	- (0)
NOR	- (0)	- (0)
SWE	- (0)	- (0)
CHE	- (0)	CAN (1)
GBR	- (0)	- (0)

Notes: For each country, the column "Countries ($\widehat{m}$)" displays the identity of selected spillover countries, and the number of these selected countries in parentheses. Model selection is performed using the $HQICM$ criterion of Groen and Kapetanios (2013). The results are displayed only for the horizons $h = 1$ and $h = 12$, with other results available upon request.

3.2 Results for Specification Including the Global Recession

Tables A3 and A4 correspond to Tables 3 and 4 in the main text, but where model selection takes place in the specification involving interactions of the idiosyncratic components with the global recession dummy:

$$s_{i,t+h} - s_{it} = \mu_1 + \mu_2 D_t + \alpha_{1i}\widehat{u}_{it} + \alpha_{2i}\widehat{u}_{it}D_t + \sum_{j \neq i} \alpha_{1j}\widehat{u}_{jt} + \sum_{j \neq i} \alpha_{2j}\widehat{u}_{jt}D_t + \varepsilon_{i,t+h}$$

The spillover effects, $\widehat{u}_{jt}$, have the same 3-letter country label as above, but the interaction terms $\widehat{u}_{jt}D_t$ are labelled "AUS-R", "CAN-R" and so on.

Table A3: In-Sample Model Selection Results: 1988-1998 pre-Euro era.

	Selection Criterion: HQ_3	
	$h = 1$	$h = 12$
	Countries ($\widehat{m}$)	Countries ($\widehat{m}$)
AUS	- (0)	ITA, FIN, BEL (3)
AUT	- (0)	NOR (1)
BEL	- (0)	DNK (1)
CAN	- (0)	FIN, ITA, ESP (3)
DNK	- (0)	BEL (1)
FIN	- (0)	- (0)
FRA	- (0)	NOR (1)
DEU	- (0)	NOR (1)
ITA	- (0)	FIN, ESP (2)
JPN	- (0)	ITA, BEL (2)
KOR	- (0)	DNK, ESP, CHE (3)
NLD	- (0)	NOR (1)
NOR	- (0)	FIN (1)
ESP	- (0)	FIN (1)
SWE	- (0)	FIN, ITA (2)
CHE	- (0)	- (0)
GBR	- (0)	FIN (1)
	Selection Criterion: BIC	
	$h = 1$	$h = 12$
	Countries ($\widehat{m}$)	Countries ($\widehat{m}$)
AUS	CHE (1)	CHE, ITA, FIN, BEL, ESP (5)
AUT	CAN, BEL, DNK (3)	NOR, DNK, BEL (3)
BEL	DNK, CAN (2)	DNK, NOR, CHE-R, SWE (4)
CAN	ESP, CHE (2)	AUS, FIN, SWE-R, ITA, ESP (5)
DNK	JPN (1)	NOR, SWE, SWE-R, BEL (4)
FIN	ITA (1)	KOR, NOR, SWE (3)
FRA	DNK, JPN, BEL (3)	NOR, CHE-R, DNK, BEL, SWE (5)
DEU	FIN, BEL, ITA (3)	NOR, DNK, NLD, BEL (4)
ITA	FIN (1)	FIN, ESP, CHE-R, SWE, CHE (5)
JPN	AUS-R, ITA (2)	ITA, BEL (2)
KOR	ITA, FIN (2)	DNK, ESP, CHE, SWE (4)
NLD	CAN, BEL, DNK (3)	NOR, DNK, BEL (3)
NOR	DNK (1)	DNK, CHE-R, FIN, ESP (4)
ESP	FIN (1)	FIN, DEU, ITA, GBR (4)
SWE	CHE (1)	FIN, ITA, CAN (3)
CHE	DNK, BEL, CAN (3)	NOR, DNK, SWE (3)
GBR	KOR-R (1)	FIN, CHE-R, FRA, SWE (4)

Notes: For each country, the column "Countries ($\widehat{m}$)" displays the identity of selected spillover countries, and the number of these selected countries in parentheses. Model selection is performed using the HQ_3 criterion (upper panel) and BIC criterion (lower panel). The results are displayed only for the horizons $h = 1$ and $h = 12$, with other results available upon request.

Table A4: In-Sample Model Selection Results: 1999-2015 post-Euro era.

	Selection Criterion: HQ_3	
	$h = 1$	$h = 12$
	Countries ($\widehat{m}$)	Countries ($\widehat{m}$)
AUS	- (0)	- (0)
CAN	- (0)	SWE (1)
DNK	- (0)	SWE (1)
EUR	- (0)	SWE (1)
JPN	- (0)	- (0)
KOR	- (0)	- (0)
NOR	- (0)	SWE (1)
SWE	- (0)	- (0)
CHE	- (0)	CAN (1)
GBR	- (0)	- (0)
	Selection Criterion: BIC	
	$h = 1$	$h = 12$
	Countries ($\widehat{m}$)	Countries ($\widehat{m}$)
AUS	JPN-R, EUR, DNK (3)	SWE, CHE (2)
CAN	JPN-R (1)	SWE, GBR-R (2)
DNK	CAN (1)	SWE, JPN, AUS-R (3)
EUR	CAN (1)	SWE, JPN, AUS-R (3)
JPN	CAN, AUS-R (2)	SWE, AUS, GBR (3)
KOR	JPN-R (1)	CHE, AUS (2)
NOR	- (0)	SWE, KOR, AUS-R (3)
SWE	NOR (1)	NOR, AUS, AUS-R, CHE (4)
CHE	CAN (1)	JPN, SWE (2)
GBR	- (0)	SWE, CHE (2)

Notes: For each country, the column "Countries ($\widehat{m}$)" displays the identity of selected spillover countries, and the number of these selected countries in parentheses. Model selection is performed using the HQ_3 criterion (upper panel) and BIC criterion (lower panel). The results are displayed only for the horizons $h = 1$ and $h = 12$, with other results available upon request.

References

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica 74*(4), 1133–1150.

Cheng, X. and B. E. Hansen (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics 186*(2), 280–293.

Groen, J. J. J. and G. Kapetanios (2013). Model selection criteria for factor-augmented regressions. *Oxford Bulletin of Economics and Statistics 75*(1), 37–63.