NOT-FOR-PUBLICATION APPENDIX TO: IN-SAMPLE INFERENCE AND FORECASTING IN MISSPECIFIED FACTOR MODELS

by Marine Carrasco and Barbara Rossi

Proofs

Proof of Lemma 1.

$$E\left[\|S_{xx}-\Sigma_{xx}\|_F^2\right] = \sum_{i=1}^N \sum_{j=1}^N E\left(\left(S_{xx}(i,j)-\Sigma_{xx}(i,j)\right)^2\right)$$
$$= \frac{1}{T^2}\sum_{i=1}^N \sum_{j=1}^N E\left(\left(\sum_{t=1}^T x_{it}x_{jt} - E\left(x_{it}x_{jt}\right)\right)^2\right).$$

In the case where x_t are iid, we have

$$E\left[\|S_{xx}-\Sigma_{xx}\|_F^2\right] = \frac{1}{T}\sum_{i=1}^N \sum_{j=1}^N E\left(\left(x_{it}x_{jt} - E\left(x_{it}x_{jt}\right)\right)^2\right)$$

using the independence. The double sum is bounded under the assumption $E\left[\|x_t\|^4\right] < C$. In the mixing case, we apply Theorem 2 of Doukhan (1994, page 26). The lemma follows from Markov inequality.

Proof of Proposition 2. First, consider the regularization bias. The eigenvectors $\{\phi_j\}$ form a basis of $\mathbb{R}^N$, so that

$$\delta^\alpha - \delta = \sum_{j=1}^N (q_j-1)\langle \delta, \phi_j\rangle \phi_j.$$

Moreover,

$$E\left(\left(x_t'\delta^\alpha - x_t'\delta\right)^2\right) = \left(\delta^\alpha-\delta\right)' E\left(x_t x_t'\right)\left(\delta^\alpha-\delta\right)$$
$$= \left\|\Sigma_{xx}^{1/2}\left(\delta^\alpha-\delta\right)\right\|^2$$
$$= \sum_{j=1}^N (1-q_j)^2 \lambda_j^2 \langle \delta, \phi_j\rangle^2$$

Using Assumption 2, we have

$$E\left(\left(x_t'\delta^\alpha - x_t'\delta\right)^2\right) = \sum_{j=1}^N (1-q_j)^2 \lambda_j^{2\beta} \frac{\langle \delta, \phi_j\rangle^2}{\lambda_j^{2\beta-2}}$$
$$\le \sup_j (1-q_j)^2 \lambda_j^{2\beta} \sum_{j=1}^N \frac{\langle \delta, \phi_i\rangle^2}{\lambda_i^{2\beta-2}}$$
$$= \begin{cases} O\left(\alpha^{\min(\beta,2)}\right) \text{ for Ridge,} \\ O\left(\alpha^\beta\right) \text{ for SC and LF,} \end{cases}$$

where the last equality follows from Proposition 3.11 of Carrasco, Florens and Renault (2007).

Now, we turn our attention toward the estimation error. Let $S_{x\varepsilon} = X'\varepsilon/T$. Then, by (3),

$$\widehat{\delta}^{\alpha} = \left(S_{xx}^{\alpha}\right)^{-1} S_{xx}\delta + \left(S_{xx}^{\alpha}\right)^{-1} S_{x\varepsilon}$$

$$\begin{aligned} &\widehat{\delta}^{\alpha} - \delta^{\alpha} \\ &= \left(S_{xx}^{\alpha}\right)^{-1} S_{xx}\delta - \left(\Sigma_{xx}^{\alpha}\right)^{-1} \Sigma_{xx}\delta && (25)\\ &+ \left(S_{xx}^{\alpha}\right)^{-1} S_{x\varepsilon}. && (26)\end{aligned}$$

Note that

$$\begin{aligned} \left(S_{xx}^{\alpha}\right)^{-1} S_{x\varepsilon} &= \sum \frac{\hat{q}_j}{\hat{\lambda}_j^2} \left\langle S_{x\varepsilon}, \hat{\phi}_j \right\rangle \hat{\phi}_j \\ &= \sum \frac{\hat{q}_j}{\hat{\lambda}_j^2} \frac{\varepsilon' X \hat{\phi}_j}{T} \hat{\phi}_j \\ &= \sum \frac{\hat{q}_j}{\hat{\lambda}_j^2} \hat{\lambda}_j \frac{\varepsilon' \hat{\psi}_j}{T} \hat{\phi}_j \\ &= \sum \frac{\hat{q}_j}{\hat{\lambda}_j} \frac{\varepsilon' \hat{\psi}_j}{T} \hat{\phi}_j. \end{aligned}$$

Moreover, by stationarity

$$E\left(\left(x_t' \left(S_{xx}^{\alpha}\right)^{-1} S_{x\varepsilon}\right)^2\right) = \frac{1}{T} E\left\| X \left(S_{xx}\right)^{-1} S_{x\varepsilon}\right\|^2.$$

We have

$$\begin{aligned} X \left(S_{xx}^{\alpha}\right)^{-1} S_{x\varepsilon} &= \sum \frac{\hat{q}_j}{\hat{\lambda}_j} \frac{\varepsilon' \hat{\psi}_j}{T} X \hat{\phi}_j \\ &= \sum \hat{q}_j \frac{\varepsilon' \hat{\psi}_j}{T} \hat{\psi}_j. \\ \frac{1}{T}\left\| X \left(S_{xx}^{\alpha}\right)^{-1} S_{x\varepsilon}\right\|^2 &= \sum \hat{q}_j^2 \frac{\left(\varepsilon' \hat{\psi}_j\right)^2}{T^2}, \\ \frac{1}{T} E\left\| X \left(S_{xx}^{\alpha}\right)^{-1} S_{x\varepsilon}\right\|^2 &= \sum_j E\left[\hat{q}_j^2 \frac{E\left[\left(\varepsilon' \hat{\psi}_j\right)^2 |X\right]}{T^2}\right] \end{aligned}$$

Using the homoskedasticity and the normalization of the eigenvectors, we have

$$E\left[\frac{\left(\varepsilon' \hat{\psi}_j\right)^2}{T} |X\right] = \sigma_{\varepsilon}^2 \frac{\hat{\psi}_j' \hat{\psi}_j}{T} = \sigma_{\varepsilon}^2.$$

Moreover, $\sum_j \hat{q}_j^2 \leq \frac{c}{\alpha}$ where c is some constant for LF, R, and SC (see for instance Carrasco, Florens, and Renault, 2007). Hence,

$$E\left(\left(x_t'\left(S_{xx}^{\alpha}\right)^{-1} S_{x\varepsilon}\right)^2\right) = O\left(\frac{1}{\alpha T}\right).$$

It follows from Markov's inequality that $x_t'\left(S_{xx}^{\alpha}\right)^{-1} S_{x\varepsilon} = O_p\left(1/\sqrt{\alpha T}\right)$.

Now we analyze the term $\left(S_{xx}^{\alpha}\right)^{-1} S_{xx}\delta - \left(\Sigma_{xx}^{\alpha}\right)^{-1} \Sigma_{xx}\delta$. We can expand δ with respect to both orthonormal series ϕ_1, ϕ_2, ...and $\hat{\phi}_1$, $\hat{\phi}_2$, ...obtaining $\delta = \sum_{j=1}^N \langle \delta, \phi_j \rangle \phi_j = \sum_{j=1}^N \left\langle \delta, \hat{\phi}_j \right\rangle \hat{\phi}_j$. $\left(\Sigma_{xx}^{\alpha}\right)^{-1}$ is defined as the population version of $\left(S_{xx}^{\alpha}\right)^{-1}$. Using (2), we have

$$\left(\Sigma_{xx}^{\alpha}\right)^{-1} v = \sum_{j=1}^N \frac{q_j}{\lambda_j^2} \langle v, \phi_j \rangle \phi_j.$$

For R and LF, Proposition 3.14 of Carrasco et al. (2007) combined with Lemma 1 permits to conclude that

$$\begin{aligned}
&\left\|\left(S_{xx}^{\alpha}\right)^{-1} S_{xx}\delta - \left(\Sigma_{xx}^{\alpha}\right)^{-1} \Sigma_{xx}\delta\right\| \\
&\leq \frac{c}{\alpha} \left\|S_{xx} - \Sigma_{xx}\right\| \left\|\delta^{\alpha} - \delta\right\| \left(1 + \epsilon_n\right) \\
&= \begin{cases} O_p\left(\dfrac{\alpha^{\min\left(\frac{\beta-1}{2},1\right)}}{\alpha\sqrt{T}}\right) & \text{for R,} \\ O_p\left(\dfrac{\alpha^{\frac{\beta-1}{2}}}{\alpha\sqrt{T}}\right) & \text{for LF.} \end{cases}
\end{aligned}$$

where $\epsilon_n = 0$ for R and $= \left\|S_{xx} - \Sigma_{xx}\right\|$ for LF. So we see that without an extra assumption on β, the term (25) dominates the term (26).

For SC, the discontinuity of q_j makes the analysis more complicated. We have

$$\begin{aligned}
&\left(S_{xx}^{\alpha}\right)^{-1} S_{xx}\delta - \left(\Sigma_{xx}^{\alpha}\right)^{-1} \Sigma_{xx}\delta \\
&= \sum_{j=1}^N \frac{\hat{q}_j}{\hat{\lambda}_j^2} \left\langle S_{xx}\delta, \hat{\phi}_j \right\rangle \hat{\phi}_j - \sum_{j=1}^N \frac{q_j}{\lambda_j^2} \left\langle \Sigma_{xx}\delta, \phi_j \right\rangle \phi_j \\
&= \sum_{j=1}^N \hat{q}_j \left\langle \delta, \hat{\phi}_j \right\rangle \hat{\phi}_j - \sum_{j=1}^N q_j \langle \delta, \phi_j \rangle \phi_j.
\end{aligned}$$

Let $I_\alpha = \left\{j : \lambda_j^2 \geq \alpha\right\}$ and $\hat{I}_\alpha = \left\{j : \hat{\lambda}_j^2 \geq \alpha\right\}$. Assume $I_\alpha \subset \hat{I}_\alpha$ (the case $\hat{I}_\alpha \subset I_\alpha$ can be treated similarly). We have

$$\left(S_{xx}^{\alpha}\right)^{-1} S_{xx}\delta - \left(\Sigma_{xx}^{\alpha}\right)^{-1} \Sigma_{xx}\delta = \sum_{j \in I_\alpha} \left\langle \delta, \hat{\phi}_j \right\rangle \hat{\phi}_j - \sum_{j \in I_\alpha} \langle \delta, \phi_j \rangle \phi_j + \sum_{j \in \hat{I}_\alpha - I_\alpha} \left\langle \delta, \hat{\phi}_j \right\rangle \hat{\phi}_j.$$

Eigenvectors associated with multiple eigenvalues are not uniquely defined and hence we can not claim the consistency of $\hat{\phi}_j$ toward ϕ_j. However, the eigenvalues are continuous functions of the elements of the matrix and hence are consistent. Similarly the projection operators are consistent. The projection operator on the space of eigenvectors corresponding to the eigenvalue λ_i^2 is defined by $P_i = \sum_{j\in I_i} \langle \phi_j, .\rangle \phi_j$ where $I_i = \left\{ j : \lambda_j^2 = \lambda_i^2 \right\}$. Similarly define $\hat{P}_i = \sum_{j\in I_i} \left\langle \hat{\phi}_j, . \right\rangle \hat{\phi}_j$. By Propositions 3 and 7 of Dauxois, Pousse and Romain[22] (1982), $\hat{P}_i$ converges to P_i at $\sqrt{T}$-rate. It follows that

$$\begin{aligned}\sum_{j\in I_\alpha} \left(\left\langle \delta, \hat{\phi}_j \right\rangle \hat{\phi}_j - \langle \delta, \phi_j \rangle \phi_j \right) &= O_p\left(\frac{|I_\alpha|}{\sqrt{T}} \right) \\ &= O_p\left(\frac{1}{\alpha\sqrt{T}} \right).\end{aligned}$$

where $|.|$ denotes the cardinal. The second equality follows from

$$\begin{aligned}|I_\alpha| &= \sum_j I\left(\lambda_j^2 \geq \alpha\right) = \sum_j \frac{\lambda_j^2}{\lambda_j^2} I\left(\lambda_j^2 \geq \alpha\right) \\ &\leq \frac{1}{\alpha}\sum_j \lambda_j^2 I\left(\lambda_j^2 \geq \alpha\right) \\ &= O\left(\frac{1}{\alpha}\right).\end{aligned}$$

On the other hand,

$$\begin{aligned}\left\| \sum_{j\in \hat{I}_\alpha - I_\alpha} \left\langle \delta, \hat{\phi}_j \right\rangle \hat{\phi}_j \right\|^2 &= \sum_{j\in \hat{I}_\alpha - I_\alpha} \left\langle \delta, \hat{\phi}_j \right\rangle^2 \\ &\leq \left| \hat{I}_\alpha - I_\alpha \right| \sup_j \left\langle \delta, \hat{\phi}_j \right\rangle^2\end{aligned}$$

where $|.|$ denotes the cardinal. As the eigenvalues are consistent (see Dauxois, Pousse, and Romain (1982)) and $\left|\hat{\lambda}_j^2 - \lambda_j^2\right| = O_p\left(1/\sqrt{T}\right)$, it follows that $\left|\hat{I}_\alpha - I_\alpha\right| = O_p\left(1/\sqrt{T}\right)$ and because $\|\delta\| < \infty$, we obtain $\left\|\sum_{j\in\hat{I}_\alpha - I_\alpha} \left\langle \delta, \hat{\phi}_j \right\rangle \hat{\phi}_j \right\|^2 = O_p\left(1/\sqrt{T}\right)$.

It follows that for R, LF, and SC,

$$\left\| \widehat{\delta}^\alpha - \delta^\alpha \right\| = O_p\left(\frac{1}{\alpha\sqrt{T}} \right).$$

Moreover,

$$\left\| x_t'\widehat{\delta}^\alpha - x_t'\delta^\alpha \right\| \leq \|x_t\| \left\| \widehat{\delta}^\alpha - \delta^\alpha \right\|$$

and $\|x_t\|$ is bounded under Assumption 1. We conclude that $\left\| x_t'\widehat{\delta}^\alpha - x_t'\delta^\alpha \right\| = O_p\left(\frac{1}{\alpha\sqrt{T}}\right)$.

[22] Dauxois, Pousse, and Romain (1982) assume a random sample. For results on weakly dependent data, see Chen and White (1998).

Proof of Proposition 3. Let $\delta^\alpha = E\left(\hat{\delta}^\alpha | X\right)$. The expected MSPE can be decomposed as the sum of variance term and a bias term:

$$\frac{1}{T} E\left[\left\| X\hat{\delta}^\alpha - X\delta \right\|^2 |X\right] = \frac{1}{T} E\left[\left\| X\hat{\delta}^\alpha - X\delta^\alpha \right\|^2 |X\right] + \frac{1}{T} \left\| X\delta^\alpha - X\delta \right\|^2.$$

We analyze these two terms separately. Note that, conditional on X, $\hat{\lambda}_j^2$, $\hat{\phi}_j$, and $\hat{\psi}_j$ are deterministic.

$$\begin{aligned} X\widehat{\delta}^\alpha &= M_T^\alpha y = M_T^\alpha X\delta + M_T^\alpha \varepsilon \\ X\delta^\alpha &= M_T^\alpha E\left(y|X\right) = M_T^\alpha X\delta. \\ X\widehat{\delta}^\alpha - X\delta^\alpha &= M_T^\alpha \varepsilon = \sum_{j=1}^{\min(N,T)} q\left(\alpha, \hat{\lambda}_j^2\right) \left\langle \varepsilon, \hat{\psi}_j \right\rangle \hat{\psi}_j. \end{aligned}$$

$$\begin{aligned} \frac{1}{T} \left\| X\widehat{\delta}^\alpha - X\delta^\alpha \right\|^2 &= \sum_j \hat{q}_j^2 \left\langle \varepsilon, \hat{\psi}_j \right\rangle^2 \\ &= \sum_j \hat{q}_j^2 \frac{\left(\varepsilon' \hat{\psi}_j\right)^2}{T^2}. \end{aligned}$$

$$\begin{aligned} \frac{1}{T} E\left[\left\| X\widehat{\delta}^\alpha - X\delta^\alpha \right\|^2 |X\right] &= \frac{1}{T^2} \sum_j \hat{q}_j^2 E\left(\hat{\psi}_j' \varepsilon\varepsilon' \hat{\psi}_j | X\right) \\ &= \frac{\sigma_\varepsilon^2}{T^2} \sum_j \hat{q}_j^2 \hat{\psi}_j' \hat{\psi}_j \\ &= \frac{\sigma_\varepsilon^2}{T} \sum_j \hat{q}_j^2 \\ &= \frac{1}{T} O\left(\frac{1}{\alpha}\right). \end{aligned}$$

For PC, $\sum_j \hat{q}_j^2 = k$.

$$\begin{aligned} X\delta^\alpha - X\delta &= \left(M_T^\alpha - I_T\right) X\delta \\ &= \sum_j \left(1 - \hat{q}_j\right) \left\langle X\delta, \hat{\psi}_j \right\rangle \hat{\psi}_j. \end{aligned}$$

$$\begin{aligned}
\frac{1}{T}\left\|X\delta^{\alpha}-X\delta\right\|^{2} &= \sum_{j}\left(1-\hat{q}_{j}\right)^{2}\left\langle X\delta,\hat{\psi}_{j}\right\rangle^{2} \\
&= \sum_{j/\hat{\lambda}_{j}^{2}>0}\left(1-\hat{q}_{j}\right)^{2}\hat{\lambda}_{j}^{2\beta}\frac{\left\langle X\delta,\hat{\psi}_{j}\right\rangle^{2}}{\hat{\lambda}_{j}^{2\beta}} \\
&\leq \sup_{j}\left(1-\hat{q}_{j}\right)^{2}\hat{\lambda}_{j}^{2\beta}\sum_{j/\hat{\lambda}_{j}^{2}>0}\frac{\left\langle X\delta,\hat{\psi}_{j}\right\rangle^{2}}{\hat{\lambda}_{j}^{2\beta}} \\
&= \begin{cases} O\left(\alpha^{\beta}\right) & \text{for SC and LF,} \\ O\left(\alpha^{\min(\beta,2)}\right) & \text{for Ridge.}\end{cases}
\end{aligned}$$

According to Lemma 1, $\left\|S_{xx}-\Sigma_{xx}\right\|=O_{p}\left(1/\sqrt{T}\right)$, hence the eigenvalues and eigenspaces of S_{xx} are consistent and for T large enough,

$$\sum_{j/\hat{\lambda}_{j}^{2}>0}\frac{\left\langle X\delta,\hat{\psi}_{j}\right\rangle^{2}}{\hat{\lambda}_{j}^{2\beta}}=\sum_{j/\hat{\lambda}_{j}^{2}>0}\frac{\left\langle \delta,\hat{\phi}_{j}\right\rangle^{2}}{\hat{\lambda}_{j}^{2\beta-2}}<\infty$$

by Assumption 2.

For PC, we have

$$\sup_{j}\left(1-\hat{q}_{j}\right)^{2}\hat{\lambda}_{j}^{2\beta}=\hat{\lambda}_{k+1}^{2\beta}.$$

Proof of Proposition 4. Let δ_{PLS}^{k} be the population version of the regularized estimator as defined in Delaigle and Hall (2012, page 328). It is the solution of

$$\delta_{PLS}^{k}=\arg\min_{\delta^{k}\in\mathcal{K}^{k}}E\left(\left(x_{t}'\delta^{k}-x_{t}'\delta\right)^{2}\right)$$

where $\mathcal{K}^{k}$ is the Krylov subspace defined as $\mathcal{K}^{k}\left(\Sigma_{xx},\Sigma_{xy}\right)=\left\{\Sigma_{xy},\Sigma_{xx}\Sigma_{xy},...,\Sigma_{xx}^{k-1}\Sigma_{xy}\right\}$. Then using arguments similar to those of Blazère et al. (2014a), we are going to show that δ_{PLS}^{k} can be written as

$$\delta_{PLS}^{k}=\sum_{i=1}^{\min(N,T)}q_{ki}\left\langle\delta,\phi_{i}\right\rangle\phi_{i}$$

where q_{kj} is as in (13) with $\hat{\lambda}_{j_l}$ replaced by λ_{j_l} and $\hat{p}_{j_l}$ replaced by $p_{j_l}=\lambda_{j_l}\left\langle\delta,\phi_{j_l}\right\rangle$ so that q_{kj} is not random. For an integer k, let $\mathcal{P}_{k,1}$ denote the set of polynomials of degree less than k with constant term equal to 1. Note that δ_{PLS}^{k} takes the form $\sum_{j}\gamma_{j}\Sigma_{xx}^{j-1}\Sigma_{xy}=\sum_{j}\gamma_{j}\Sigma_{xx}^{j-1}\Sigma_{xx}\delta$. By an argument similar to that of Blazère et al. (2014a, Proposition 3.1), we have

$$\begin{aligned}
E\left(\left(x_{t}'\delta_{PLS}^{k}-x_{t}'\delta\right)^{2}\right) &= \left(\delta_{PLS}^{k}-\delta\right)'\Sigma_{xx}\left(\delta_{PLS}^{k}-\delta\right) \\
&= \left\|\Sigma_{xx}^{1/2}\left(\delta_{PLS}^{k}-\delta\right)\right\|^{2} \\
&= \left\|\Sigma_{xx}^{1/2}Q_{k}\left(\Sigma_{xx}\right)\delta\right\|^{2}
\end{aligned}$$

where $Q_k(t) = 1 - tP_k(t) \in \mathcal{P}_{k,1}$ satisfies

$$Q_k \in \arg\min_{Q \in \mathcal{P}_{k,1}} \left\| \Sigma_{xx}^{1/2} Q(\Sigma_{xx}) \delta \right\|^2$$

Then using a proof similar to that of Blazère et al. (2014a, Proposition 3.2), it can be shown that $Q_0 = 1$, $Q_1, ...$is a sequence of orthonormal polynomials with respect to the measure $d\mu = \sum_j \lambda_j^2 p_j^2 \delta_{\lambda_j}$ where $p_j = \lambda_j \langle \delta, \phi_j \rangle$. The specific expression of the q_{kj} follows from Theorem 4.1. of Blazère[23] et al. (2014a).

Now, we analyze the regularization bias.

$$\begin{aligned}
E\left(\left(x_t'\delta_{PLS}^k - x_t'\delta\right)^2\right) &= \left(\delta_{PLS}^k - \delta\right)' \Sigma_{xx} \left(\delta_{PLS}^k - \delta\right) \\
&= \sum_{i=1}^{\min(N,T)} (1 - q_{ki})^2 \lambda_i^2 \langle \delta, \phi_i \rangle^2 \\
&= \sum_{i=1}^{\min(N,T)} (1 - q_{ki})^2 p_i^2 \\
&= \sum_{i=1}^{\min(N,T)} (1 - q_{ki}) p_i^2
\end{aligned}$$

where the last equality can be proved using an argument similar to Blazère et al. (2014a, Lemma 3.6). Again using Blazère et al. (2014a, Propositions 6.1 and 6.2), we have

$$\begin{aligned}
&\sum_{i=1}^{\min(N,T)} (1 - q_{ki}) p_i^2 \\
&= \sum_{i=1}^{\min(N,T)} \left[\sum_{(j_1,...,j_k) \in I_k^+} w_{j_1...j_k} \prod_{l=1}^{k} \left(1 - \frac{\lambda_i^2}{\lambda_{j_l}^2}\right) \right]^2 p_i^2 \\
&= \sum_{\min(N,T) > j_1 > ... > j_k \geq 1} w_{j_1...j_k} \sum_{i=j_1+1}^{\min(N,T)} \prod_{l=1}^{k} \left(1 - \frac{\lambda_i^2}{\lambda_{j_l}^2}\right) p_i^2 \\
&\leq \sum_{i=k+1}^{\min(N,T)} \prod_{l=1}^{k} \left(1 - \frac{\lambda_i^2}{\lambda_l^2}\right) p_i^2.
\end{aligned}$$

For any $i \geq k+1$, $0 \leq \prod_{l=1}^{k} \left(1 - \frac{\lambda_i^2}{\lambda_l^2}\right) \leq 1$. Hence,

$$\sum_{i=1}^{\min(N,T)} (1 - q_{ki}) p_i^2 \leq \sum_{i=k+1}^{\min(N,T)} p_i^2 = E\left(\left(x_t'\delta_{PC}^k - x_t'\delta\right)^2\right).$$

[23] In Blazère et al (2014a, b), q_{kj} is denoted $1 - \widehat{Q}_k(\lambda_j)$.

Moreover,

$$
\begin{aligned}
& E\left(\left(x_t'\delta_{PLS}^k - x_t'\delta\right)^2\right) \\
& \leq \sum_{i\geq k+1} \lambda_i^{2\beta} \prod_{l=1}^{k}\left(1-\frac{\lambda_i^2}{\lambda_l^2}\right)^2 \frac{\lambda_i^2 \langle \delta, \phi_i\rangle^2}{\lambda_i^{2\beta}} \\
& \leq \sup_{i\geq k+1} \lambda_i^{2\beta} \prod_{l=1}^{k}\left(1-\frac{\lambda_i^2}{\lambda_l^2}\right)^2 \sum_i \frac{\langle \delta, \phi_i\rangle^2}{\lambda_i^{2\beta-2}}.
\end{aligned}
$$

So an upper bound is given by $C\lambda_{k+1}^{2\beta}$.

To be more precise, we consider the special case where there is a constant $C>0$ such that

$$\frac{1}{C}j^{-2} \leq \lambda_j^2 \leq Cj^{-2}.$$

Without loss of generality, we focus on the case $\lambda_j^2 = j^{-2}$. Then

$$\sup_{\lambda} \lambda^{2\beta} \prod_{l=1}^{k}\left(1-\frac{\lambda^2}{\lambda_l^2}\right)^2 = \sup_{j\geq k+1} j^{-4\beta} \prod_{l=1}^{k}\left(1-\left(\frac{l}{j}\right)^2\right)^2.$$

Let us take the log

$$-4\beta \ln(j) + 2\sum_{l=1}^{k} \ln\left(1-\left(\frac{l}{j}\right)^2\right). \tag{27}$$

We see that the first term decreases with j whereas the second term increases. As the function $f(x) = \ln\left(1-\left(\frac{x}{j}\right)^2\right)$ is decreasing and continuous in x, we can apply the comparison theorem

$$\int_1^{k+1} f(s)\,ds \leq \sum_{l=1}^{k} \ln\left(1-\left(\frac{l}{j}\right)^2\right) \leq \int_1^{k} f(s)\,ds.$$

When $k+1=j$, the lower bound goes to -∞ but for $j>k+1$, the lower bound is finite and both bounds can be used to derive an equivalent of the series.

Using a change of variables $x=s/j$, we obtain

$$
\begin{aligned}
\int_1^k f(s)\,ds &= \int_1^k \ln\left(1-\left(\frac{s}{j}\right)^2\right) ds \\
&= j\int_{1/j}^{k/j} \ln\left(1-x^2\right) dx.
\end{aligned}
$$

Now we use an integration by parts with $u=\ln\left(1-x^2\right)$, $u' = -2x/\left(1-x^2\right)$, $v'=1$, $v=x$ to obtain

$$
\begin{aligned}
& \int_a^b \ln\left(1-x^2\right) dx \\
& = b\ln\left(1-b^2\right) - a\ln\left(1-a^2\right) + 2\int_a^b \frac{x^2}{1-x^2}dx \\
& = b\ln\left(1-b^2\right) - a\ln\left(1-a^2\right) + 2\left[a-b+\frac{1}{2}\ln\left(\frac{1+b}{1-b}\right) - \frac{1}{2}\ln\left(\frac{1+a}{1-a}\right)\right]
\end{aligned}
$$

where the formula for $\int_a^b \frac{x^2}{1-x^2}dx$ was found in Spiegel (1993, Formules et tables mathématiques). Hence,

$$\int_1^k f(s)\,ds = k\ln\left(1-\left(\frac{k}{j}\right)^2\right) - \ln\left(1-\frac{1}{j^2}\right) + 2 - 2k + j\ln\left(\frac{1+k/j}{1-k/j}\right) - j\ln\left(\frac{1+1/j}{1-1/j}\right).$$

This expression can be used to derive an equivalent of $\sum_{l=1}^{k}\ln\left(1-\left(\frac{l}{j}\right)^2\right)$ as j and k go to infinity. Recall that $j \geq k+1$.

We consider two possible scenarios:

1) $k/j \to c^*$ with $0 < c^* < 1$,

2) $k/j \to 0$

and see which of these possibilities is compatible with the maximization of the criterion given in (27).

1) When $k/j \to c^* > 0$, we have

$$\int_1^k f(s)\,ds = k\ln(1-c^*) + j\ln\left(\frac{1+c^*}{1-c^*}\right) + o(1).$$

To maximize (27), we need to equate the rates for $k\ln(1-c^*) + j\ln\left(\frac{1+c^*}{1-c^*}\right)$ and $-4\beta\ln(j)$ or by dividing by j, we need to equate the rates of $c^*\ln(1-c^*) + \ln\left(\frac{1+c^*}{1-c^*}\right)$ and $-4\beta\ln(j)/j$. This is possible only if c^* is close to zero which yields the second case.

2) When $k/j \to 0$, we obtain

$$\int_1^k f(s)\,ds = -\frac{k^3}{j^2} + o(1)$$

using $\ln(1+x) \sim x$ as x goes to zero.

Now using the first order condition of the maximization of $-\frac{2k^3}{j^2} - 4\beta\ln(j)$, we see that the maximum is reached for $j = Ck^{3/2}$ for some constant $C > 0$. This concludes the proof.

The result for the estimation error follows from Theorem 5.3 and Equation (5.11) of Delaigle and Hall (2012).

Proof of Proposition 5. The results can be proved using a proof similar to that of Proposition 1 of De Mol. et al. (2008) exploiting some results already used in the proof of Proposition 2 combined with the shrinkage properties of our estimators, namely $\|\delta^\alpha\| \leq \|\delta\|$.

We have

$$\begin{aligned}\hat{\delta}^\alpha - \delta^\alpha &= \left(S_{xx}^\alpha\right)^{-1} S_{xy} - \left(\Sigma_{xx}^\alpha\right)^{-1}\Sigma_{xy} \\ &= \left(S_{xx}^\alpha\right)^{-1}\left(S_{xy} - \Sigma_{xy}\right) \qquad (28) \\ &+ \left(\left(S_{xx}^\alpha\right)^{-1} - \left(\Sigma_{xx}^\alpha\right)^{-1}\right)\Sigma_{xy}. \qquad (29)\end{aligned}$$

First consider (28),

$$\begin{aligned}\left\|\left(S_{xx}^{\alpha}\right)^{-1}\left(S_{xy}-\Sigma_{xy}\right)\right\| &\leq \left\|\left(S_{xx}^{\alpha}\right)^{-1}\right\|\left\|S_{xy}-\Sigma_{xy}\right\| \\ &= O_p\left(\frac{1}{\alpha}\right)O_p\left(\frac{\sqrt{N}}{\sqrt{T}}\right)\end{aligned}$$

where $\left\|S_{xy}-\Sigma_{xy}\right\| = O_p\left(\frac{\sqrt{N}}{\sqrt{T}}\right)$ comes from Lemma 2 of De Mol et al. (2008). (29) can be rewritten as

$$\begin{aligned}&\left(\left(S_{xx}^{\alpha}\right)^{-1}-\left(\Sigma_{xx}^{\alpha}\right)^{-1}\right)\Sigma_{xy} \\ &= \left(\left(S_{xx}^{\alpha}\right)^{-1}\Sigma_{xx}-\left(\Sigma_{xx}^{\alpha}\right)^{-1}\Sigma_{xx}\right)\Sigma_{xx}^{-1}\Sigma_{xy} && (30) \\ &= \left(\left(S_{xx}^{\alpha}\right)^{-1}S_{xx}-\left(\Sigma_{xx}^{\alpha}\right)^{-1}\Sigma_{xx}\right)\delta && (31) \\ &\left(\left(S_{xx}^{\alpha}\right)^{-1}\Sigma_{xx}-\left(S_{xx}^{\alpha}\right)^{-1}S_{xx}\right)\delta. && (32)\end{aligned}$$

The term (31) was analyzed in Section 3.1 and is negligible with respect to (32). We have

$$\begin{aligned}&\left\|\left(\left(S_{xx}^{\alpha}\right)^{-1}\Sigma_{xx}-\left(S_{xx}^{\alpha}\right)^{-1}S_{xx}\right)\delta\right\| \\ &\leq \left\|\left(S_{xx}^{\alpha}\right)^{-1}\right\|\left\|S_{xx}-\Sigma_{xx}\right\|\left\|\delta\right\| \\ &= O_p\left(\frac{N}{\alpha\sqrt{T}}\left\|\delta\right\|\right)\end{aligned}$$

where $\left\|S_{xx}-\Sigma_{xx}\right\| = O_p\left(\frac{N}{\sqrt{T}}\right)$ comes from Lemma 2 of De Mol et al. (2008).

Regarding the estimation bias, the only major difference with De Mol et al. (2008) is the presence of the term α^{β} and d. It comes from

$$\begin{aligned}\left\|\delta^{\alpha}-\delta\right\|^2 &= \sum_{j=1}^{\min(N,T)}\left(q_j-1\right)^2\left\langle\delta,\phi_j\right\rangle^2 \\ &= \sum_{j=1}^{r}\left(q_j-1\right)^2\left\langle\delta,\phi_j\right\rangle^2\end{aligned}$$

because $\delta = \left(\Lambda\Lambda' + \Psi\right)^{-1}\Lambda\theta$ belongs to the range of Λ. Then, we have

$$\begin{aligned}\left\|\delta^{\alpha}-\delta\right\|^2 &= \sum_{j=1}^{r}\lambda_j^{2\beta}\left(q_j-1\right)^2\frac{\left\langle\delta,\phi_j\right\rangle^2}{\lambda_j^{2\beta}} \\ &\leq \sup_{j=1,..,r}\lambda_j^{2\beta}\left(q_j-1\right)^2\sum_{j=1}^{r}\frac{\left\langle\delta,\phi_j\right\rangle^2}{\lambda_j^{2\beta}}.\end{aligned}$$

For SC, $\sup_j \lambda_j^{2\beta}\left(q_j-1\right)^2 = \alpha^{\beta}$ if $\alpha \leq \lambda_{\min}\left(\Lambda'\Lambda\right)$ and 0 otherwise. By Weyl's theorem, $\lambda_j^2 \geq \lambda_{\min}\left(\Lambda'\Lambda\right)$, we obtain for SC:

$$\left\|\delta^{\alpha}-\delta\right\|^2 \leq \frac{\alpha^{\beta}}{\lambda_{\min}\left(\Lambda'\Lambda\right)^{\beta}}\left\|\delta\right\|^2.$$

For Landweber-Fridman, $\sup_j \lambda_j^{2\beta} (q_j - 1)^2 \leq (\alpha/d)^\beta$ where d is such that $0 < d < 1/\lambda_{\max}(S_{xx})$ so that d goes to zero with $1/N$. Therefore,

$$\|\delta^\alpha - \delta\|^2 \leq \left(\frac{\alpha}{d}\right)^\beta \frac{\|\delta\|^2}{\lambda_{\min}(\Lambda'\Lambda)^\beta}.$$

The variance term differs for LF from the other two regularizations because $\frac{\hat{q}_j^2}{\hat{\lambda}_j^4} \leq \frac{C}{\alpha^2}$ for some constant C for SC and R, however $\frac{\hat{q}_j^2}{\hat{\lambda}_j^4} \leq \frac{d}{\alpha^2}$ for LF. Given d depends on N, it can not be ignored.

The fact that $\|\delta\| = O\left(\frac{\sqrt{\lambda_{\max}(\Lambda'\Lambda)}}{\lambda_{\min}(\Lambda'\Lambda)}\right)$ has been established by De Mol et al. (2008).

Proof of Proposition 7. Let $\lambda(M_T(\alpha))$ be the largest eigenvalue of $M_T(\alpha)$. Let $R_T(\alpha) \equiv E[L_T(\alpha)|X]$.

First we consider the case with discrete index set. We recall the assumptions of Li (1987).

(A.1) $\lim_{n\to\infty} \sup_{\alpha\in A_T} \lambda(M_T(\alpha)) < \infty$,

(A.2) $E\varepsilon_i^{4m} < \infty$ for some m,

(A.3) $\sum_{\alpha\in A_T} (TR_T(\alpha))^{-m} \to \infty$.

For completeness, we recall here the theorems 2.1 and 3.2 of Li (1987).

Theorem 2.1 (Li, 1987) Assume that (A.1)-(A.3) hold. Then C_L is asymptotically optimal.

Theorem 3.2 (Li, 1987) Assume that (A.1)-(A.3) and the following conditions hold:

(A.4) $\inf_{\alpha\in A_T} L_T(\alpha) \to 0$,

(A.5) for any sequence $\{\alpha_T \in A_T\}$ such that $T^{-1}trM_T(\alpha_T) \to 0$, we have

$$\left(T^{-1}trM_T(\alpha_T)\right)^2 / \left(T^{-1}tr\left(M_T(\alpha_T)^2\right)\right) \to 0,$$

(A.6) $\sup_{\alpha\in A_T} T^{-1}trM_T(\alpha) \leq \gamma_1$, for some $1 > \gamma_1 > 0$,

(A.7) $\sup_{\alpha\in A_T} \left(T^{-1}trM_T(\alpha)\right)^2 / \left(T^{-1}tr\left(M_T(\alpha)^2\right)\right) \leq \gamma_2$, for some $1 > \gamma_2 > 0$.

Then GCV is asymptotically optimal.

First, we are going to check the conditions for SC and LF.

Note that the eigenvalues of $M_T(\alpha)$ are all bounded by 1, so that (A.1) is satisfied. For PC, $M_T(\alpha)$ is a projection matrix and hence Corollary 2.1 of Li (1987) implies that (A.3) can be replaced by

(A.3') $\inf_{\alpha\in A_T} TR_T(\alpha) \to \infty$.

(A.3') implies (A.3) with $m = 2$.

For LF, $M_T(\alpha)$ is not a projection matrix, however we can still establish inequality (2.5) in the proof of Corollary 2.1 by using the fact that $tr\left(M_T(\alpha)^2\right) \geq C/\alpha$ for some constant C. To see this use the mean value theorem on $q_j(\lambda)$ and $q_j(0) = 0$. Hence for LF also, we can replace (A.3) by (A.3').

According to Proposition 3, $TR_T(\alpha) = O\left(\frac{1}{\alpha}\right) + O\left(T\alpha^{\beta}\right)$. The α which minimizes $TR_T(\alpha)$ is of order $T^{-1/(\beta+1)}$. Hence $\inf_{\alpha \in A_T} TR_T(\alpha) = O\left(T^{1/(\beta+1)}\right) \to \infty$. Therefore (A.3') holds.

The justification for replacing σ_ε^2 by $\hat{\sigma}_\varepsilon^2$ is provided by Corollary 2.2. of Li (1987). The optimality of C_L for SC and LF follows from Theorem 2.1. of Li (1987).

For GCV, we need to check Assumptions (A.4) to (A.6).

(A.4) We have $E \inf_{\alpha \in A_T} L_T(\alpha) \leq \inf_{\alpha \in A_T} EL_T(\alpha) = O\left(T^{-\beta/(\beta+1)}\right)$. Hence by Markov's inequality, $\inf_{\alpha \in A_T} L_T(\alpha)$ converges to zero.

$M_T(\alpha)$ is idempotent so (A.5) is automatically satisfied and (A.7) is equivalent to (A.6). $T^{-1} tr M_T(\alpha) = 1/(\alpha T) \leq \gamma_1$. So that $\alpha \geq 1/(T\gamma_1)$. This is satisfied for the set A_T we have selected. Thus, Assumptions (A.4) to (A.6) hold. The optimality of GCV for SC and LF follows from Theorem 3.2. of Li (1987).

Now, we turn our attention to Ridge. Theorem 1 of Li (1986) establishes the optimality of C_L for R under Assumption (A.3') which can be checked using the same argument as above. The optimality of GCV for R follows from Theorem 2 of Li (1986) under the extra assumption 3(iii) which corresponds to Condition (A.2) in Li (1986). Note that the condition (A.2) in Li (1986) is expressed in terms of the eigenvalues of XX' instead of XX'/T but the ratio is invariant to a rescaling of the eigenvalues.

Additional Empirical Results

Table A1. Empirical Analysis (Forecast horizon = 1 quarter)
Forecast Rationality Regressions

Panel A. GDP Growth				
	Coefficients		Wald Test	
	Constant	Slope	Statistic	P-value
Factor Models:				
Bai and Ng	1.56	-0.61	58.2	0.00
Cross V.	1.35	-0.52	25.4	0.00
Ridge	-1.09	0.25	1.81	0.40
PLS	1.57	-0.60	44.4	0.00
LF	3.86	-1.35	4.51	0.10
Comb. CV	-0.67	0.12	1.25	0.53
Comb. Mall.	0.13	-0.15	2.67	0.26
BMA	0.34	-0.26	6.00	0.05
Comb.	0.04	-0.07	0.61	0.74

Panel B. Inflation				
	Coefficients		Wald Test	
	Constant	Slope	Statistic	P-value
Factor Models:				
Bai and Ng	-0.03	-0.92	141.4	0.00
Cross V.	-0.02	-1.01	28.2	0.00
Ridge	-0.02	-1.01	62.3	0.00
PLS	-0.02	-1.51	51.3	0.00
LF	-0.02	-3.41	9.01	0.01
Comb. CV	-0.02	-1.06	11.4	0.00
Comb. Mall.	-0.02	-0.82	17.3	0.00
BMA	-0.01	-0.46	3.23	0.20
Comb.	-0.04	-0.58	8.60	0.01

Table A2. Empirical Analysis (Forecast horizon = 4 quarter)
Forecast Rationality Regressions

Panel A. GDP Growth				
	Coefficients		Wald Test	
	Constant	Slope	Statistic	P-value
Factor Models:				
Bai and Ng	1.04	-0.43	42.9	0.00
Cross V.	0.82	-0.34	14.7	0.00
Ridge	-1.04	0.25	1.63	0.44
PLS	1.28	-0.50	54.5	0.00
LF	2.87	-1.03	3.56	0.17
Comb. CV	-1.73	0.47	5.78	0.06
Comb. Mall.	-0.41	0.04	1.82	0.40
BMA	-0.08	-0.07	1.31	0.52
Comb.	-0.55	0.14	0.77	0.68
Panel B. Inflation				
	Coefficients		Wald Test	
	Constant	Slope	Statistic	P-value
Factor Models:				
Bai and Ng	-0.12	-0.56	32.4	0.00
Cross V.	-0.09	-0.56	17.4	0.00
Ridge	-0.11	-0.45	13.2	0.00
PLS	-0.11	-0.80	113.4	0.00
LF	-0.07	-1.13	2.08	0.35
Comb. CV	-0.11	-0.12	1.15	0.56
Comb. Mall.	-0.13	-0.20	2.08	0.35
BMA	-0.01	-0.12	0.52	0.77
Comb.	-0.10	-0.31	4.68	0.10

Notes to Tables A1 and A2. The pool of regressors contains three lags of the predictors (x_{t-h} includes Z_{t-h}, Z_{t-h-1}, and Z_{t-h-2}).

Figure A1. Forecasting GDP, h=1. Forecast Rationality.

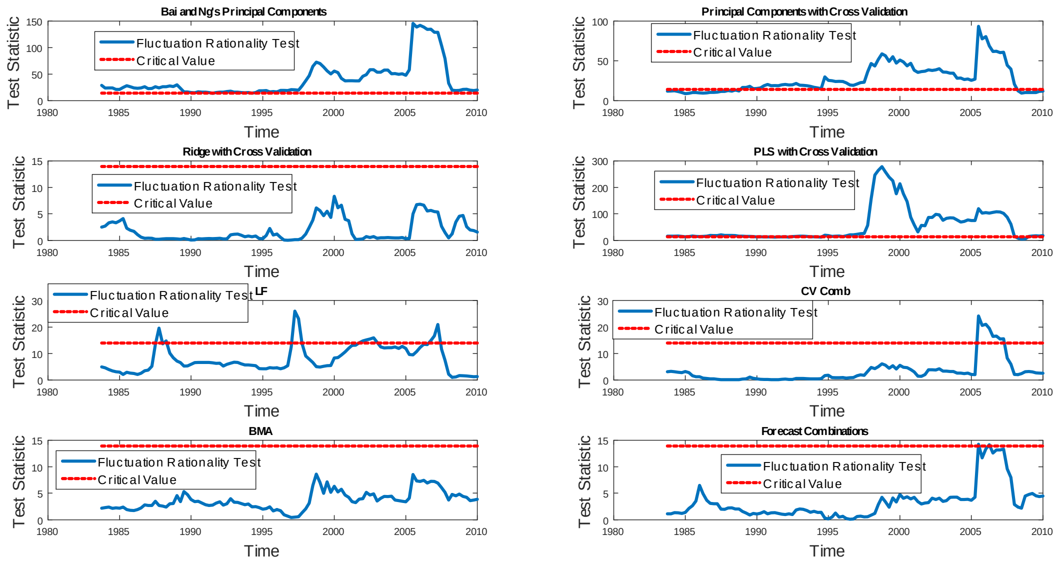


Figure A2. Forecasting Inflation, h=1. Forecast Rationality.

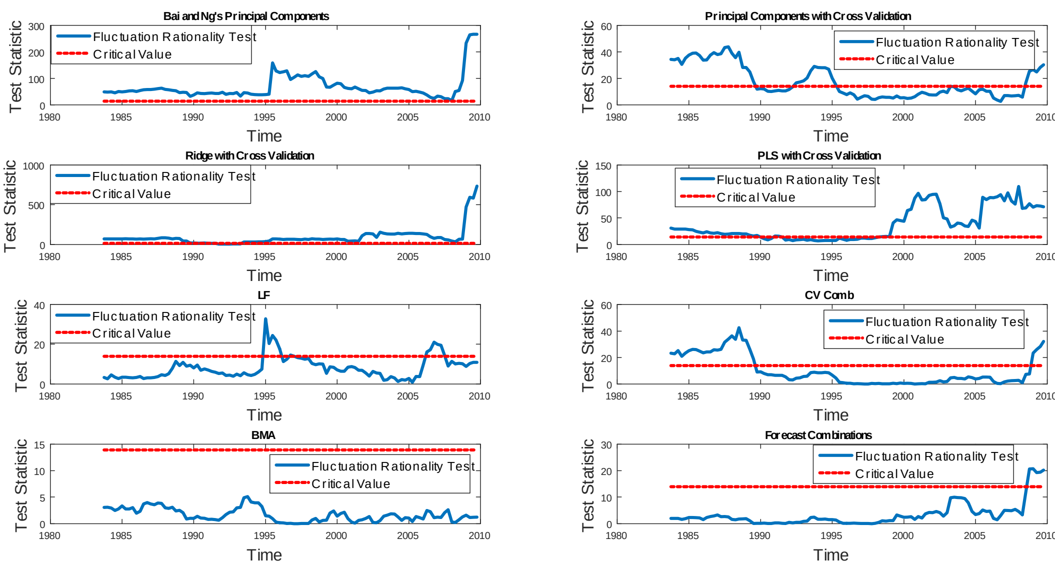


Figure A3. Forecasting GDP, h=4. Forecast Rationality.

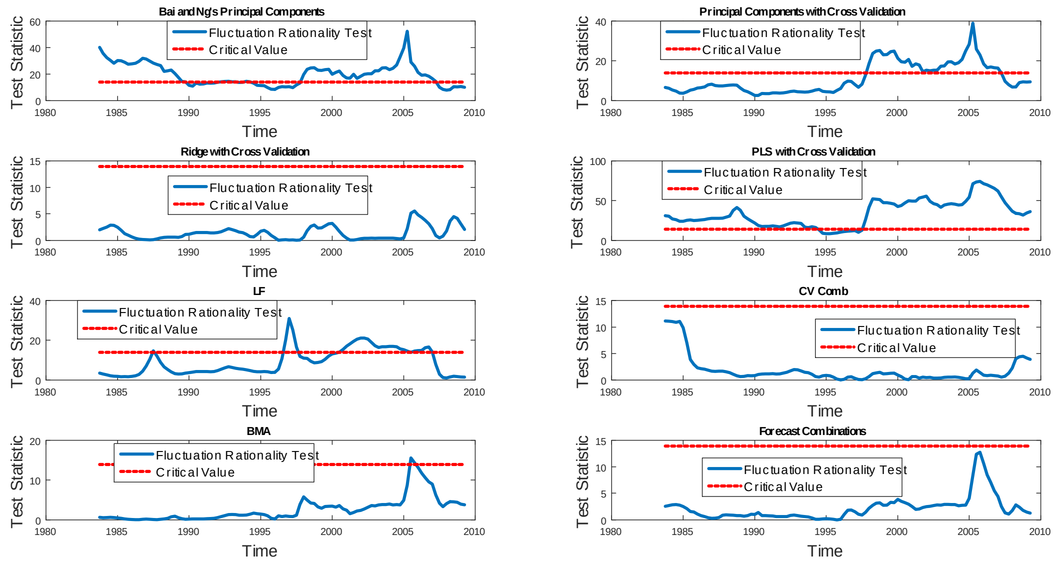
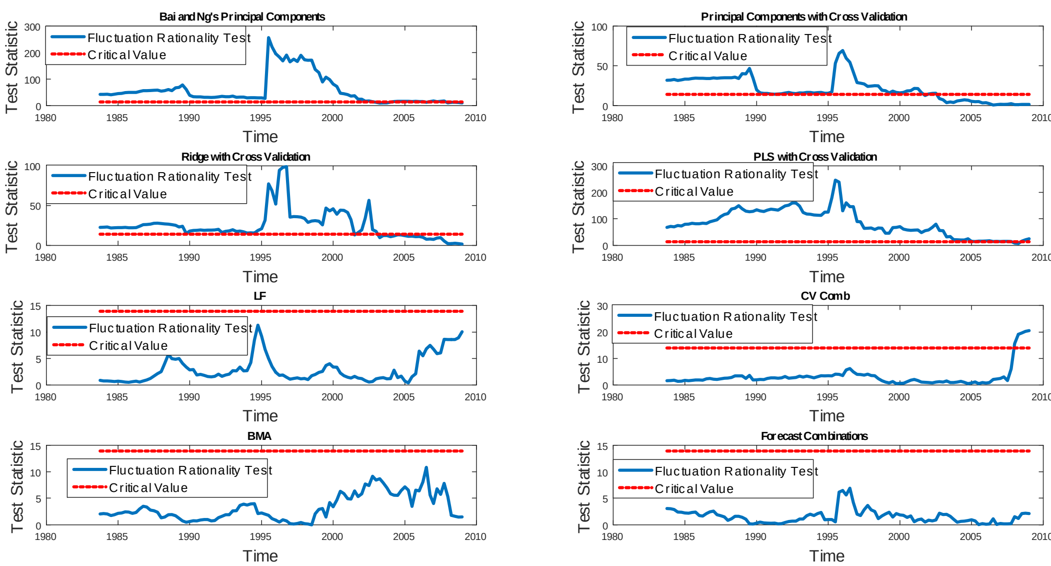


Figure A4. Forecasting Inflation, h=4. Forecast Rationality



Notes to Figures A1-A4. The figures report Rossi and Sekhposyan's (2015) Fluctuation Rationality test (solid line) and critical values (dotted lines) for the forecasting models listed in the title. Figures A1-A2 focus on forecasting output growth and inflation in the short-run, while figures A3-A4 focus on the long-run. The pool of regressors contains three lag of the predictors (x_{t-h} includes Z_{t-h}, Z_{t-h-1}, and Z_{t-h-2}).