

The uncertainty of principal components in dynamic factor models

Esther Ruiz

(joint with J. de Vicente)

Department of Statistics
Universidad Carlos III de Madrid

Joint Statistics and Econometrics Workshop

London School of Economics

24 March 2017

Outline of the talk

- Motivation
- Revisiting DFM and factor extraction
- Asymptotic distribution of principal components factors
- Extant bootstrap procedures for principal components factors
- A new procedure
- Empirical application
- Conclusions

Motivation

Dynamic factor model

Dynamic factor models (DFMs) are useful in different contexts:

- Representing [business cycles](#) ; Aruoba et al. (2012, JBES), Camacho et al. (2015, JAE) and Breitung and Eickmeier (2016, Advances in Econometrics).
- [Instrumental variables](#) ; Bai and Ng (2010, Econometric Theory) and Kapetanios and Marcellino (2010, CSDA).
- [Regressors in FAVAR and FECM](#) ; Stock and Watson (2005, 2010), Bernanke et al. (2005, QJE) and Banerjee et al. (2014, IJF).
- [Factor augmented predictive regressions](#) ; Stock and Watson (2006, Handbook of Economic Forecasting), Bai and Ng (2013, JE), Ando and Tsay (2014, ER), Djogbenou et al. (2015, JTSA).

Motivation

Principal Components

Principal Components (PC) is not efficient when compared with factor extraction based on Kalman filter and smoothing algorithms or Dynamic Principal Components (Forni et al. (2000, REandS)).

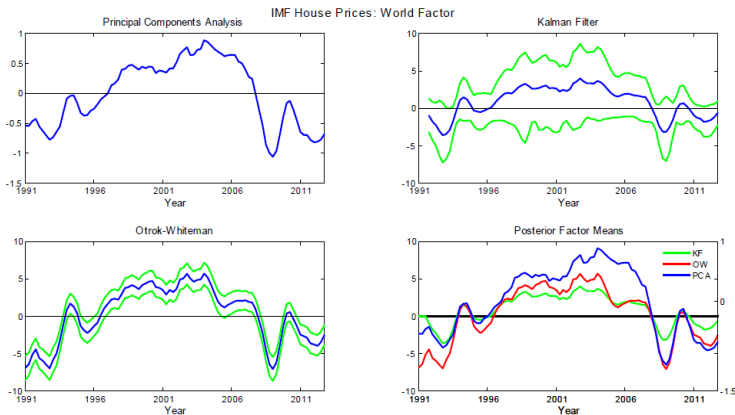
However, it is still among the most popular alternative factor extraction procedures when the dimension of the system is very large: Ludvigson and Ng (2007, 2009, 2011), Ando and Tsay (2014, ER), Goncalves and Perron (2014, JE), Djogbenou et al. (2015, JTSA) and Jackson et al. (2016, Advances in Econometrics).

Measuring the uncertainty associated with factor estimates should be part of interpreting these estimates : Bai (2003, Econometrica and 2004, JE), Bai and Ng (2006, Econometrica) and Jackson et al. (2016, Advances in Econometrics)

Motivation

Principal Components

Jackson et al. (2016, Advances in Econometrics): World factor extracted from IMF real house prices in advanced and emerging economies



Motivation

Contributions in this paper

1. Analyse the performance of the asymptotic and available bootstrap procedures when constructing confidence bands for the PC factors.
2. Propose an alternative bootstrap procedure that overcome some of the limitations of extant procedures.
3. Analyze its (asymptotic and) finite sample properties.
4. Empirical application.

Revisiting Dynamic Factor Models and factor extraction

Consider the following **stationary single factor** DFM

$$Y_t = PF_t + \varepsilon_t$$
$$F_t = \phi F_{t-1} + \eta_t$$

where $Y_t = (Y_{1t}, \dots, Y_{Nt})'$ is the vector of observations at time $t = 1, \dots, T$, $P = (p_1, \dots, p_N)'$ is the vector of fixed factor loadings, F_t is the unobserved factor with zero mean and variance one (for identification), ε_t , the vector of idiosyncratic noises, is **Gaussian white noise with scalar covariance matrix** $\Sigma_\varepsilon = q^{-1}I$ such that the idiosyncratic noises are homoscedastic and uncorrelated and **q is the signal to noise ratio**, η_t is Gaussian white noise with variance $1 - \phi^2$ independent of ε_t and $|\phi| < 1$.

The marginal distribution of the underlying factor is given by

$$F_t \sim N(0, 1)$$

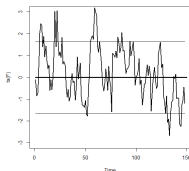
Confidence intervals for F_t can be constructed using this marginal distribution (no information about $\{Y_t\}_{t=1}^T$ is used).

$$\pm z_{\alpha/2}$$

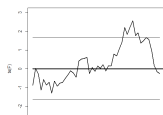
Revisiting Dynamic Factor Models and factor extraction

Plot of "true" underlying factor together with 80% confidence intervals.

DFM with $\phi = 0.7$, $q = 1$, $T = 150$ and $N = 70$. $RMSE = 1$ and coverage 0.77



DFM with $\phi = 0.98$, $q = 1$, $T = 50$ and $N = 50$. $RMSE = 1$ and coverage 0.77



Revisiting Dynamic Factor Models and factor extraction

Reduce uncertainty using the conditional distribution

$$F_t | Y_{t-1} \sim N(f_{t|t-1}, V)$$

$$f_{t|t-1} = \phi (f_{t-1|t-2} + K(Y_{t-1} - Pf_{t-1|t-2}))$$

$$K = VP' (PVP' + q^{-1}I)^{-1}$$

where V is the steady-state variance which is given by

$$V = \frac{(1 - \phi^2)A - 1 + \phi^2 + \sqrt{(1 - \phi^2)A - 1 + \phi^2 + 4(1 - \phi^2)A}}{2A}$$

$$A = qP'P$$

Note that the steady-state variance does not depend on the observations.

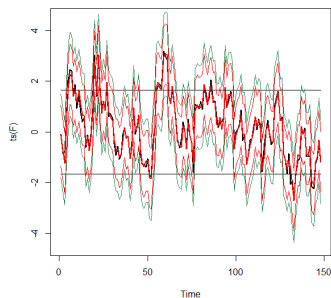
Confidence intervals for the factor are then given by

$$f_{t|t-1} \pm z_{\alpha/2} \sqrt{V}$$

Revisiting Dynamic Factor Models and factor extraction

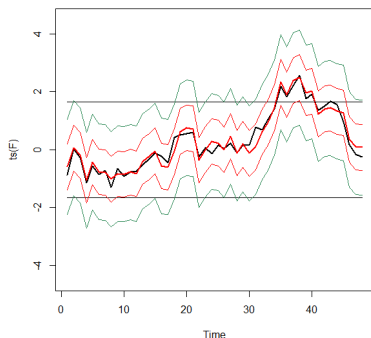
Plot of "true" underlying factor together with 80% confidence intervals based on steady-state one-step-ahead means and RMSE.

DFM with $\phi = 0.7$, $q = 1$, $T = 150$ and $N = 70$. $RMSE = 0.73$ and coverage 0.76.



Revisiting Dynamic Factor Models and factor extraction

DFM with $\phi = 0.98$, $q = 1$, $T = 50$ and $N = 50$. $RMSE = 0.47$ and coverage 1.



Uncertainty can be further reduced using the (smoothed) distribution conditional on the full sample $\{Y_t\}_{t=1}^T$

$$F_t | Y_T \sim N(f_{t|T}, S)$$

$$f_{t|T} = f_{t|t-1} + K(Y_t - Pf_{t|t-1}) + \frac{\phi}{1 + VA} (f_{t+1|T} - \phi (f_{t|t-1} + K(Y_t - Pf_{t|t-1})))$$

$$S = \frac{V(1 + VA - \phi^2)}{(1 + VA)^2 - \phi^2}$$

; see Poncela and Ruiz (2015) for the expressions of the steady-state variance.

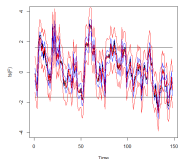
Confidence intervals are then given by

$$f_{t|T} \pm z_{\alpha/2} \sqrt{S}$$

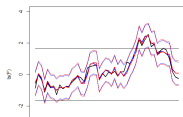
Revisiting Dynamic Factor Models and factor extraction

Plot of "true" underlying factor together with 80% confidence intervals based on smooth means and RMSE.

DFM with $\phi = 0.7$, $q = 1$, $T = 150$ and $N = 70$. $RMSE = 0.28$ and coverage 0.74.



DFM with $\phi = 0.98$, $q = 1$, $T = 50$ and $N = 50$. $RMSE = 0.43$ and coverage 0.97.



Revisiting Dynamic Factor Models and factor extraction

In practice, the parameters are estimated and substituted to estimate the means and MSE: $\hat{f}_{t|T}$ and \hat{S} .

Note that \hat{S} is not an estimate of the MSE of $\hat{f}_{t|T}$ but it estimates the MSE of $f_{t|T}$. The MSE of $\hat{f}_{t|T}$ can be decomposed into the filter MSE and the estimation MSE

$$\begin{aligned} E_T(\hat{f}_{t|T} - F_t)^2 &= \\ E_T(\hat{f}_{t|T} - f_{t|T})^2 + E_T(f_{t|T} - F_t)^2 + 2E_T(\hat{f}_{t|T} - f_{t|T})(f_{t|T} - F_t) &= \\ E_T(\hat{f}_{t|T} - f_{t|T})^2 + E(f_{t|T} - F_t)^2 \end{aligned}$$

If $f_{t|T} = E_T(F_t)$, the last term is equal to zero; see Pfefferman and Tiller (2005, JTSA). The filter uncertainty does not depend on the observed data.

Compute the expectation conditional on the parameter estimates and integrate over all possible values of the parameter estimates; Hamilton (1986, JE)

$$E_{\hat{\theta}} \left(E_T \left((\hat{f}_{t|T} - F_t)^2 | \hat{\theta} \right) \right) = \\ E_{\hat{\theta}} \left(E_T \left((\hat{f}_{t|T} - f_{t|T})^2 | \hat{\theta} \right) \right) + E_{\hat{\theta}} \left(E \left((f_{t|T} - F_t)^2 | \hat{\theta} \right) \right)$$

In this context, we could use Rodríguez and Ruiz (2012, CSDA) to obtain bootstrap confidence intervals for the factors that account for the error and parameter uncertainty.

Revisiting Dynamic Factor Models and factor extraction

Consider now that we use **PC** to extract factors (**filter based on information contained in the full sample**). Intervals based on either marginal or one-step-ahead MSEs are not appropriate.

$$V(r) = \min_{P,F} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \left(Y_{it} - P_i' F_t \right)^2$$

Using the normalization, $F'F/T = I_r$, the estimated factors, \tilde{f} , is \sqrt{T} times eigenvectors corresponding to the r largest eigenvalues of the $T \times T$ matrix YY' and $\tilde{P}' = \frac{1}{T} \tilde{f}' Y$.

$$\hat{f} = \tilde{f} \tilde{V} = \frac{1}{N} Y \tilde{P}$$

where \tilde{V} is the $r \times r$ diagonal matrix consisting of the first r eigenvalues of the matrix $\frac{1}{TN} YY'$ arranged in decreasing order. Note that \hat{f} satisfies that $\frac{1}{T} \hat{f}' \hat{f} = I_r$.

Revisiting Dynamic Factor Models and factor extraction

True (scaled) factor

$$HF_t = \frac{1}{N} P' Y_t - \frac{1}{N} P' \varepsilon_t, \quad H = \left(\frac{P' P}{N} \right)$$

Factor extracted using **known parameters**

$$f_t = \frac{1}{N} P' Y_t$$

Factor extracted using **estimated parameters**

$$\hat{f}_t = \frac{1}{N} \tilde{P}' Y_t$$

$$E_T \left[(\hat{f}_t - HF_t)(\hat{f}_t - HF_t)' \right] = E_T \left[(\hat{f}_t - f_t)(\hat{f}_t - f_t)' \right] + E_T \left[(f_t - HF_t)(f_t - HF_t)' \right] + 2E_T \left[(\hat{f}_t - f_t)(f_t - HF_t)' \right]$$

Uncertainty due to parameter estimation

$$E_T \left[(\hat{f}_t - f_t)(\hat{f}_t - f_t)' \right] = \frac{1}{N^2} E_T \left[(\tilde{P} - P)' Y_t Y_t' (\tilde{P} - P) \right]$$

Uncertainty due to the filter

$$E_T \left[(f_t - HF_t)(f_t - HF_t)' \right] = \frac{1}{N^2} E_T \left[P' \varepsilon_t \varepsilon_t' P \right]$$

Covariance

$$\begin{aligned} E_T \left[(\hat{f}_t - f_t)(f_t - HF_t)' \right] &= \frac{1}{N^2} E_T \left[(\tilde{P} - P)' Y_t \varepsilon_t' P \right] \\ &= \frac{1}{N^2} E_T \left[\tilde{P}' \varepsilon_t \varepsilon_t' P - P' \varepsilon_t \varepsilon_t' P \right] \end{aligned}$$

The MSE is given by

$$\begin{aligned} E_T \left[(\hat{f}_t - HF_t)(\hat{f}_t - HF_t)' \right] &= \frac{1}{N^2} E_T \left[(\tilde{P} - P)' Y_t Y_t' (\tilde{P} - P) \right] + \\ &\quad \frac{2}{N^2} E_T \left[\tilde{P}' \varepsilon_t \varepsilon_t' P \right] - \frac{1}{N^2} E_T \left[P' \varepsilon_t \varepsilon_t' P \right] \end{aligned}$$

OBJECTIVE: Compute

$$\begin{aligned} E_{\hat{\theta}} \left(E_T (\hat{f}_t - HF_t)^2 | \hat{\theta} \right) = \\ \frac{1}{N^2} E_{\hat{\theta}} \left(E_T \left[(\tilde{P} - P)' Y_t Y_t' (\tilde{P} - P) \right] | \hat{\theta} \right) + \frac{2}{N^2} E_{\hat{\theta}} \left(E_T \left[\tilde{P}' \varepsilon_t \varepsilon_t' P \right] | \hat{\theta} \right) \\ - \frac{1}{N} E_{\hat{\theta}} \left(E_T \left[P' \varepsilon_t \varepsilon_t' P \right] | \hat{\theta} \right) \end{aligned}$$

Asymptotic distribution of principal components factors

The asymptotic distribution of PC factor estimates is derived by Bai and Ng (2013, JE) in the context of the particular model considered today

$$\sqrt{N} \left(\hat{f}_t - \frac{P'P}{N} F_t \right) \xrightarrow{d} N(0, \Gamma_t)$$

Bai and Ng (2006, Econometrica) propose alternative estimators of Γ_t depending on the properties of the idiosyncratic term. They conclude that the estimator based on assuming cross-sectionally uncorrelated but heteroscedastic noises is more convenient in practice.

The asymptotic covariance matrix corresponds to the second term of the MSE: **Does not take into account parameter uncertainty** .

Poncela and Ruiz (2016, Advances in Econometrics) show that **asymptotic confidence bands underestimate the uncertainty associated with PC factors**

Asymptotic distribution of principal components factors

Finite sample performance

- Number of replicates: $R = 5000$
- Two DGPs:
 - DFM with $\phi = 0.5$, $q = 1$, $T = 50$ and $N = 20$
 - DFM with $\phi = 0.7$, $q = 1$, $T = 150$ and $N = 70$
- True MSE are calculated at each moment of time, $t = 1, \dots, T$, as

$$\frac{1}{R} \sum_{i=1}^R \left(\widehat{f}_t^{(i)} - HF_t \right)^2$$

- The asymptotic MSE is calculated at each moment of time as $\frac{1}{N} \Gamma_t$.

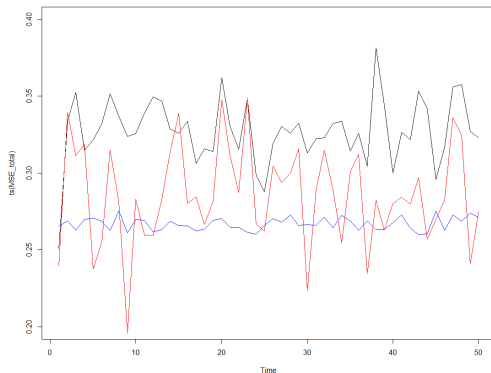
Then, we compute $\frac{1}{RN} \sum_{i=1}^R \Gamma_t$

- We compute the coverage over the Monte Replicates at each moment of time

Asymptotic distribution of principal components factors

Finite sample performance

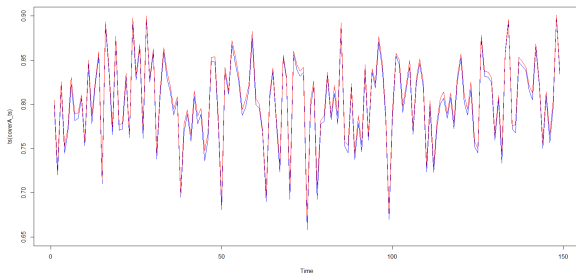
DFM with $\phi = 0.5$, $q = 1$, $T = 50$ and $N = 20$



Asymptotic distribution of principal components factors

Finite sample performance

DFM with $\phi = 0.7$, $q = 1$, $T = 150$ and $N = 70$



Extant bootstrap procedures

Bootstrap procedures used with different goals in the context of PC factor extraction:

1. Confidence intervals for impulse response functions of FAVAR models: Yamamoto (2016, manuscript)
2. Testing hypothesis about the parameters of factor augmented predictive regressions: Gospodinov and Ng (2013, RES), Goncalves and Perron (2014, JE and 2016, manuscript), Neely et al. (2014, Management Science), Djogbenou et al. (2015, JTSA). Constructing forecast intervals: Goncalves et al. (2017, JBES)
3. Testing about the autoregressive parameter of the factor equation: Shintani and Guo (2015, ER)

The bootstrap procedures already proposed in the literature can be classified into two main groups:

1. Procedures based on [moving block bootstrap](#) : Gospodinov and Ng (2013, RES).
2. [Residual bootstrap procedures](#) : Goncalves and Perron (2014, JE), Yamamoto (2016, manuscript) and Shintani and Guo (in press, ER) (who also obtains confidence intervals for diffusion index)

Both the block bootstrap procedure of Gospodinov and Ng (2013, RES) and the residual bootstrap procedure of Shintani and Guo (2015, ER) obtain replicates of the marginal distribution of the factors: They are not very informative.

Extant bootstrap procedures for PC factors

Residual bootstrap: Bootstrapping idiosyncratic residuals

Goncalves and Perron (2014, JE) propose bootstrapping the **idiosyncratic** residuals (similar to Yamamoto (2016, manuscript)):

1 Estimate \tilde{P} and \tilde{f}_t using PC. Obtain the residuals $\tilde{\varepsilon}_t = Y_t - \tilde{P}\tilde{F}_t$ and their empirical distribution, \tilde{G}_ε .

2 Bootstrap replicates

$$Y_t^{*(b)} = \tilde{P}\tilde{F}_t + \varepsilon_t^{*(b)}$$

where $\varepsilon_t^{*(b)}$ are random extractions with replacement from \tilde{G}_ε .

3 Using $Y_t^{*(b)}$ obtain PC estimates of the factors: $\tilde{f}_t^{*(b)}$.

4 Repeat steps 2 and 3 for $b = 1, \dots, B$

Very popular procedure implemented by Ludvigson and Ng (2007, 2009, 2011), Djogbenou et al. (2015, JTSA), Goncalves et al. (2016, JBES) and Breigtun and Eickmeier (2016).

Extant bootstrap procedures for PC factors

Residual bootstrap of idiosyncratic residuals

- The factor estimates are not conditional on the observed sample as they are based on $Y_t^{*(b)}$.
- It is possible to reduce the uncertainty around the factor estimates.

New bootstrap procedure for PC factors

The new proposed bootstrap algorithms aim to incorporate parameter uncertainty and to compute the expectation conditional on $\{Y_t\}_{t=1}^T$.

- 1 Estimate \tilde{P} and \hat{f}_t using PC. Obtain the residuals $\hat{\varepsilon}_t = Y_t - \hat{P}\hat{f}_t$ and their empirical distribution, \hat{G}_ε . Regress \hat{f}_t on \hat{f}_{t-1} and estimate $\hat{\Phi}$ by OLS. Obtain the residuals $\hat{u}_t = \hat{f}_t - \hat{\Phi}\hat{f}_{t-1}$ and their empirical distribution function \hat{G}_u .
- 2 Bootstrap replicates (obtain de marginal distribution of the parameter estimates). As in Yamamoto (2016, manuscript)

$$F_t^{*(b)} = \hat{\Phi}F_{t-1}^{*(b)} + u_t^{*(b)}$$

$$Y_t^{*(b)} = \hat{P}F_t^{*(b)} + \varepsilon_t^{*(b)}$$

where $F_1^{*(b)} = \hat{f}_1$ and $u_t^{*(b)}$ and $\varepsilon_t^{*(b)}$ are random extractions with replacement from \hat{G}_u and \hat{G}_ε , respectively. Obtain $\tilde{P}^{*(b)}$ using $Y_t^{*(b)}$ as usual.

New bootstrap procedure for PC factors

3. Compute $\hat{f}_t^{*(b)} = \frac{1}{N} \tilde{P}^{*(b)'} Y_t$ and the corresponding residuals $\hat{\varepsilon}_t^{*(b)} = Y_t - \hat{P}^{*(b)} \hat{f}_t^{*(b)}$

Both $\hat{f}_t^{*(b)}$ and $\hat{\varepsilon}_t^{*(b)}$ are conditional on Y_t and incorporate the parameter uncertainty through $\tilde{P}^{*(b)}$

4. Repeat steps 2 and 3 for $b = 1, \dots, B$

New bootstrap procedure for PC factors

Obtain the bootstrap analog of the MSE as follows:

1. The bootstrap analog of $E_T \left[\left(\hat{f}_t - f_t \right) \left(\hat{f}_t - f_t \right)' \right]$ is given by

$$\frac{1}{B} \sum_{b=1}^B \left(\hat{f}_t^{*(b)} - \hat{f}_t \right) \left(\hat{f}_t^{*(b)} - \hat{f}_t \right)' = \frac{1}{BN^2} \sum_{b=1}^B \left(\tilde{P}^{*(b)} - \tilde{P} \right)' Y_t' Y_t \left(\tilde{P}^{*(b)} - \tilde{P} \right)$$

2. The bootstrap analog of $\frac{2}{N^2} E_T \left[\tilde{P}' \varepsilon_t \varepsilon_t' P \right]$ is given by

$$\frac{2}{BN^2} \sum_{i=1}^B \tilde{P}^{*(b)'} \hat{\varepsilon}_t^{*(b)} \hat{\varepsilon}_t^{*(b)'} \tilde{P} = B$$

3. The bootstrap analog of $\frac{1}{N} E_{\hat{\theta}} \left(E_T \left[P' \varepsilon_t \varepsilon_t' P \right] \mid \hat{\theta} \right)$ is given by

$$\frac{2}{N} \tilde{\Gamma} - \frac{1}{BN} \sum_{b=1}^B \tilde{\Gamma}^{*(b)} = C$$

New bootstrap procedure for PC factors

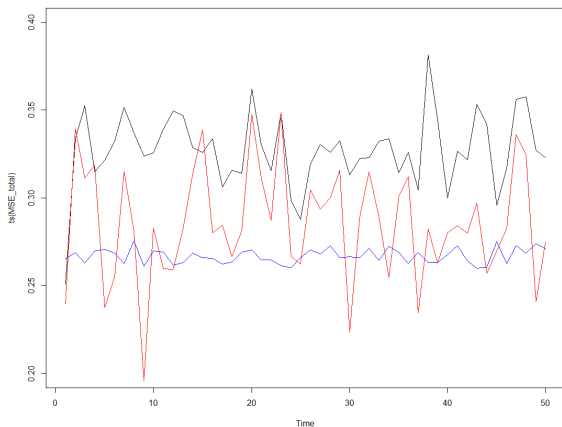
Assuming normality, the corresponding bootstrap $(1 - \alpha)$ % confidence interval for the true factors, F_t , is given by

$$\left(\frac{\tilde{P}'\tilde{P}}{N}\right)^{-1} \hat{f}_t \pm z_{\alpha/2} \left(\frac{\tilde{P}'\tilde{P}}{N}\right)^{-1} (A + B - C) \left(\frac{\tilde{P}'\tilde{P}}{N}\right)^{-1}$$

where $z_{\alpha/2}$ is the $\alpha/2$ quantile of the normal distribution.

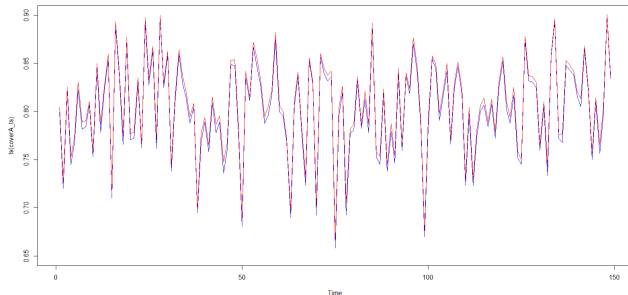
New bootstrap procedure for PC factors: Finite sample properties

DFM with $\phi = 0.5$, $q = 1$, $T = 50$ and $N = 20$



New bootstrap procedure for PC factors: Finite sample properties

DFM with $\phi = 0.7$, $q = 1$, $T = 70$ and $N = 150$



New bootstrap procedure for PC factors: Asymptotic validity

Following Goncalves and Perron (2014, JE) assume:

- The idiosyncratic errors are weakly dependent across time and cross-section dimensions.
- $\frac{\sqrt{T}}{N} \rightarrow c$
- Moment restrictions on the idiosyncratic components and factors

Then

$$\frac{1}{T} \sum_{t=1}^T \|\tilde{f}_t^* - H^* \tilde{f}\|^2 = O_p^*(\delta_{NT}^{-2})$$

where $\delta_{NT} = \min(\sqrt{N}, \sqrt{T})$.

Therefore, it will be also possible to prove that \tilde{P}^* estimates a rotation of \tilde{P} which is consistent for P .

The bootstrap replicates proposed are linear combinations of \tilde{P}^* so, it should be possible to prove its asymptotic validity

Empirical application 1

We consider quarterly observations of a Spanish macroeconomic system with $N=73$ variables observed from 1980Q1 to 2014Q4, $T=140$.

Assume that $r = 1$.

Transform to stationarity.

Estimation results: $\sum \hat{p}_i^2 = 12.21$. Estimated weights larger than 0.8 in absolute value correspond to: Gross capital formation, capital stock, imports, unemployment rate, rest of the world clients' GDP, total resources of public administrations.

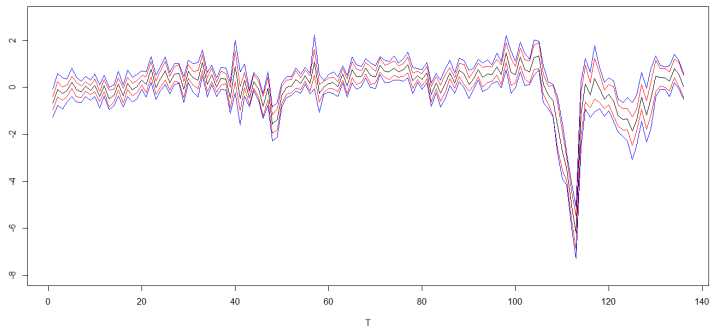
$$\hat{\phi} = 0.74$$

$$\hat{\sigma}_a^2 = [0.28, 0.99] \text{ with the mode around } 0.97$$

Cross-correlations: $[-0.93, 0.86]$ with the mode around 0

$$\hat{\gamma} = [-0.74, 0.97], \text{ distributed uniformly in this interval}$$

Empirical application 1



Empirical application 2

Dataset in house prices in advanced and emerging markets from Cesa-Bianchi et al. (2015, JMCB), studied by Jacks et al. (Advances in Econometrics) observed quarterly from 1998 to 2011 with $N=45$, $T=58$

Implement Bai and Ng (2002), Onatski (2010) or Ahn and Horenstein (2013) to original data: $\hat{r}_{IC} = \hat{r}_{ON} = 9$, $\hat{r}_{AH} = 1$ or 2 ; differentiated data: $\hat{r}_{IC} = 9$, $\hat{r}_{ON} = \hat{r}_{AH} = 1$ or 2 . Assume that $r = 1$.

Estimation results: $\sum \hat{p}_i^2 = 25.87$. Estimated weights are largest for Australia, Austria, Belgium, Canada, United Kingdom, Ukraine, Denmark and Estonia.

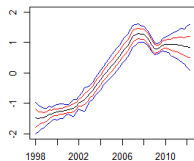
$$\hat{\phi} = 0.98$$

$$\hat{\sigma}_a^2 = [0.008, 0.99] \text{ with mean } 0.94$$

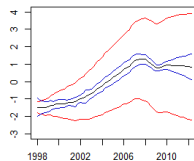
Cross-correlations: $[-0.9, 1]$ with mean 0.006

$$\hat{\gamma} = [0.56, 1.06]$$

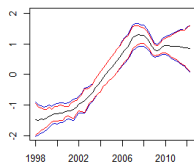
Empirical application 2



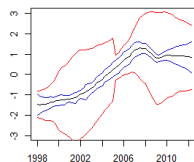
Bai (2003)



Yamamoto (2012)



Gonçalves and Perron (2014)



Gospodinov and Ng (2013)

- Analyze the performance of available bootstrap procedures when implemented to obtain intervals for the factors: they are uninformative, too wide (when based on the marginal distribution).
- Propose (tentatively) a new bootstrap procedure able to incorporate parameter uncertainty computing the conditional expectations: Promising results.

- Formal proof of asymptotic validity
- Consider the case of $r > 1$ number of factors incorporating uncertainty on the number of factors. The asymptotic distribution is not affected when the number of factors is unknown and is estimated; Bai (2003, Econometrica)
- Consider non-Gaussian models.
- Extension to factor augmented predictive regressions. Can we improve over Goncalves and Perron (2016, manuscript)?