Constrained principal components estimation of large approximate factor models

Rachida Ouysse

University of New South Wales

American Economic Association Meetings
Philadelphia 2018

Introduction

- The growing availability of financial and macroeconomic data sets including a large number of time series have generated interest in predictive models with many possible predictors. Examples.
- Curse of dimensionality. Standard techniques such as OLS, MLE, or Bayesian inference perform poorly, since the proliferation of regressors magnifies estimation uncertainty and produces inaccurate out-of-sample predictions.

  As a consequence, inference methods aimed at dealing with the curse of dimensionality have become increasingly popular.
- **Curse to blessing. Dense modeling techniques** recognize that all possible explanatory variables might be important for prediction, although the impact of some of them can be small.
- Factor Models. Stock & Watson (2002) (Diffusion indexes). Comovements in the predictors are treated as arising from a **small** number of unobserved sources, or common factors. Factor

Introduction: Approximate factor model

- The approximate factor model of Chamberlain and Rothschild [1983] allows (i) the number of observations to be large in both the cross-section (N) and the time (T) dimensions and, (ii) weak serial and cross-sectional correlation in the idiosyncratic dynamics.
- The large-dimensional nature of the panel opens the horizon for consistent estimation of the factors, something that is not possible when the number of cross-section units is small.
- Large literature on estimating the (approximate) factor model, and developping the large-sample inferential theory.
  1. **Principal components (PC) method**: Stock & Watson (1998, 2002a,2002b), Bai & Ng (2002), Bai(2003), Forni, Hallin, Lippi, & Reichlin, 2000, 2005.
  2. Factor Analysis and MLE: Geweke, 1977; Sargent & Sims, 1977; Geweke & Singleton, 1980;Watson & Engle, 1983; Stock & Watson, 1989;; Doz, Giannone& Reichlin, 2012; Bai and Li (2012a, 2012b)
- BUT these methods are not efficient in the presence of heteroskedasticity/dependence in the errors

This paper

- Novel **principal components PC** based estimation method when N is large and errors are cross-sectionally dependent
  - PC are consistent estimators of the common factors for both the cross-sectional dimension N and the sample size T going to infinity
  - PC are a solution of a computational problem since they can be easily computed even if the cross-sectional dimension N is large and possibly larger than the sample size T
  - PC are feasible since consistency can be achieved for any path of N and T
- The estimator is computationally tractable:
  - based on PC method of a modified covariance matrix
  - doesn't require inverting a large covariance matrix

Efficiency Literature

- Heterogeneity and time dependence:
  - Forni,*et. al*(2005, JoE) proposed a two-step dynamic principal components approach in the frequency domain to exploit the cross sectional heterogeneity of the idiosyncratic component;
  - Giannone, *et. al* (2004, 2005 JoE) used a parametric time domain two-step estimator involving dynamic principal components and kalman filter to exploit idiosyncratic heteroscedasticity;
  - Breitung & Tenhofen (2011,JASA) used a Gaussian log-likelihood to estimate the factor structure in the case of heteroscedastic and serially correlated errors.
- Cross sectional dependence:
  - Choi (2012, ET): GPCE (FGPCE) of common components shown to be more effciient than ordinary PCA, and the variance of the forecasting error smaller. However, this estimator requires the inverse of sample covariance matrix. It requires $N < T$.
- This papar is concerned with cross sectional correlation: incorporate the very feature that defines an approximate factor structure in the estimation of the model without imposing any structure on the error covariance matrix.

Literature with cross sectional dependence and large N

- Random matrix theory: Estimation of large (sparse) covariance matrices using threshholding (Bickel & Levina (2008, Ann.Stat), Rothman et al. (2009) and penalized MLE (Lam & Fan (2009), Bien & Tibshirani (2011))

- Bai & Liao (2016): threshholding and penalized MLE to the approximate factor model.
- This paper is related to Bai & Liao (2016) **BUT** is different in two dimensions:
  - this paper uses the PC estimation framework
  - this paper doesn't explicitely put any structure on the covariance matrix

Why care about Efficiency?

- Boivin and Ng [2006]: forecasts are less efficient and the factor estimates are adversely affected when the errors are cross correlated and/or have vast heterogeneity in the variances:
  - "Weighting the data by their properties when constructing the factors also lead to improved forecasts"
  - with cross-correlated errors the estimated factors may be less useful for forecasting when more series are available.
- Cross-sectional dependence is a likely feature of the data in many applications, Fig-sparsity

Approximate factor model: Notation

- In matrix notation, the model is written as

$$\mathbf{X} = \mathbf{F}\mathbf{\Lambda}' + \mathbf{e}, \tag{1}$$

where $\mathbf{X} = [\underline{X}_1, \cdots, \underline{X}_T]'$ is the $T \times N$ matrix of observations,
$\mathbf{e} = [\underline{e}_1, \cdots, \underline{e}_T]'$ is a $T \times N$ matrix of idiosyncratic errors,
$\mathbf{F} = [F_1, \cdots, F_T]'$ is the $T \times r$ matrix of common factors and
$\mathbf{\Lambda} = [\lambda_1, \cdots, \lambda_N]'$ is $N \times r$ matrix of factor loadings.

- Additional assumptions include that the matrices Σ_Λ and Σ_F, where $\frac{1}{N}\sum_{i=1}^{N} \lambda_i \lambda_i' \to \Sigma_\Lambda$ as $N \to \infty$, and $\frac{1}{T}\sum_{i=1}^{T} F_t F_t' \to \Sigma_F$ as $T \to \infty$, are bounded and positive definite.

Approximate factor model: dependence assumption

- The approximate factor structure of Chamberlain & Rothschild (1983) allows for weak cross-section correlation in the error component in the following sense:
  there exist a positive constant M such that for all N, $t = 1, \cdots, T$,

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|E(e_{it}e_{jt})| \leq M \tag{2}$$

- Assume errors are independent across time and time and cross-sectional dynamics are separable, $E(\underline{e}_t\underline{e}_t') = \Omega$. Ω an $N \times N$ matrix

GLS-PC estimation

- The standard PCA estimates minimize

$$V(\lambda_i, F_t) = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \lambda_i' F_t)^2 = (NT)^{-1} \sum_{t=1}^{T} \underline{e}_t' \underline{e}_t.$$

- The objective function $V(\lambda_i, F_t)$ has an ordinary least squares form with spherical errors.
- Assume that the errors are independent across time and that the time and cross sectional dynamics are separable: $E(\underline{e}_t \underline{e}_t') = \Omega$
- If Ω is known, a GLS type PC estimates can be constructed:

$$V^*(\lambda_i, F_t) = (NT)^{-1} \sum_{t=1}^{T} \underline{e}_t' \Omega^{-1} \underline{e}_t$$

- A candidate estimate for Ω is the sample covariance matrix. However, when $N > T$, $\widehat{\Omega}$ is singular and minimizing $V^*(\lambda_i, F_t)$ is unfeasible.

Weighted PCA

- Boivin & NG (2006) propose to minimize

$$W(\lambda_i, F_t) = \sum_{i=1}^{N} \sum_{t=1}^{T} w_{it}(X_{it} - \lambda_i' F_t)^2,$$

- Weighting schemes they consider:
  - w_{it} is the inverse of the diagonal element of $\widehat{\Omega}_T$ estimated using data up to time T
  - w_{it} is the inverse of $N^{-1} \sum_{i=1}^{N} |\widehat{\Omega}_T(i,j)|$

Bai & Liao (2016, JoE)

- The first is a two-step estimator that minimizes the negative log-likelihood function,

$$-L_1(\mathbf{\Lambda}, \Omega) = \frac{1}{N}\text{log}|\text{det}(\mathbf{\Lambda}\mathbf{\Lambda}' + \Omega_N)| + \frac{1}{N}\text{tr}\left(S_{\mathbf{X}}(\mathbf{\Lambda}\mathbf{\Lambda}' + \Omega_N)^{-1}\right),$$

  where $S_{\mathbf{X}}$ is the sample covariance matrix of the data. An estimator of Ω_N is obtained in a first step estimation using threshholding.

- The second joint estimator they propose is an l_1-penalized maximum likelihood estimator that minimizes,

$$L_2\mathbf{\Lambda}, \Omega) = -L_1(\mathbf{\Lambda}, \Omega) + \frac{1}{N}\sum_{i \neq j}\mu_T w_{ij}|\Omega_{ij}|$$

This paper: Constrained PC estimation (Cn-PC)

- Let $\tau_{ij,t} = E(e_{it}e_{jt}) = \tau_{ij}$ and $e_{it} = X_{it} - \lambda_i' F_t$.
- This paper proposes an estimator that solves:

$$\underset{\lambda_i, F_t}{\text{minimize}} \; (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} e_{it}^2$$

$$\text{s.t } \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \text{sign}(\tau_{ij}) \tau_{ij} \leq M$$

Constrained PC estimation

- Let $\mathcal{S}$ be $N \times N$ matrix with elements $[\mathcal{S}_{ij}]$ defined as,

$$\mathcal{S}_{i,j} = 0 \text{ for } i = j \tag{3}$$
$$\mathcal{S}_{i,j} = 1 \times \text{sign}(\tau_{ij}) \text{ for } i \neq j. \tag{4}$$

- Let $L_1(F, \Lambda) = \sum_{t=1}^{T} \underline{e}_t' \underline{e}_t$ and $L_2(F, \Lambda) = \frac{1}{NT} \sum_{t=1}^{T} \underline{e}_t' \mathcal{S} \underline{e}_t - M$.
  The Cn-PC optimization can be written as:

$$\underset{\Lambda, F}{\text{minimize}} \{L_1(\Lambda, F, r) \text{ s.t } L_2(F, \Lambda, r) \leq 0\}, \tag{5}$$

  under the normalization of either $T^{-1} \sum_{t=1}^{T} F_t F_t' = I_r$, or $N^{-1} \sum_{i=1}^{N} \lambda_i \lambda_i' = I_r$.

The Cn-PC Estimator

Proposition

- *The constrained principal component estimator (Cn-PC) for F, denoted $\hat{F}_{\mu_{NT}}$ is:*

  $\sqrt{T}$ times the matrix consisting of the eigenvectors corresponding to the r largest eigenvalues of the matrix $\mathbf{X}\mathcal{A}_N\mathbf{X}'$, where $(\mathcal{A}_N = I_N + \mu_{NT}\mathcal{S})/NT$
- *The tuning parameter μ_{NT} is the Lagrange multiplier*
- *The Cn-PC estimator for Λ^0, denoted $\hat{\Lambda}$ is given by $\hat{\Lambda}_{\mu_{NT}} = \frac{1}{T}\mathbf{X}\hat{F}_{\mu_{NT}}$.*

CnPC

Comparison to OLS-PC and GLS-PC

- OLS-PC:

$$\hat{F} \quad : \quad \sqrt{T} \times \text{first } r \text{ principal components of } \mathbf{X}\mathbf{X}'/NT$$

- GLS-PC:

$$\hat{F} \quad : \quad \sqrt{T} \times \text{first } r \text{ principal components of } \mathbf{X}\Omega^{-1}\mathbf{X}'/NT$$

- Cn-PC:

$$\begin{aligned}
\hat{F}_{\hat{\mu}_{NT}} \quad &: \quad \sqrt{T} \times \text{first } r \text{ principal components of } \mathbf{X}\left(I_N + \hat{\mu}_{NT}\mathcal{S}\right)\mathbf{X}'/NT, \\
\hat{\Lambda}_{\hat{\mu}_{NT}} \quad &: \quad \hat{\Lambda}_{\hat{\mu}_{NT}} = \mathbf{X}'\hat{F}_{\hat{\mu}_{NT}}/T, \\
\hat{\mu}_{NT} \quad &: \quad M = (NT)^{-1}\sum_{t=1}^{N}\underline{\hat{e}}_t'\mathcal{S}\underline{\hat{e}}_t
\end{aligned}$$

Choosing M

- Boivin & NG (2006) use $\hat{\tau}^* = \max_i \hat{\tau}_i^*/N$, where $\hat{\tau}_i^* = \sum_{j=1}^{N} |T^{-1} \sum_{t=1}^{T} \hat{e}_{it}\hat{e}_{jt}|$ as indicator for M/N, which should be small and decreasing with N. That is, the bounding quantity M is of order $O_p(N)$.

- Cross validation. Let $M_0 = \sum_{j=1}^{N} \sum_{i=1}^{N} |\hat{\tau}_{ij}|$, $\hat{\tau}_{ij} = \sum_{t=1}^{T} \hat{e}_{it}\hat{e}_{jt}/T$. Calibrate and estimate M by cross-validation. Using a normalized parameter $m = M/M_0$ to index the constrained estimates of F and Λ over a grid of values of m between 0 and 1.

Asymptotic properties of Cn-PC

Theorem (Convergence)

For any fixed (known) $r \geq 1$, there exists a suitable $(r \times r)$ full rank rotation matrix $\mathcal{H}$ such that under Assumption A1-A5

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t - \mathcal{H}' F_t^0\right\|^2 = O_p(\delta_{NT}^{-2}) + O_p(\mu_{NT}^{-2}\delta_{NT}^{-2}),$$

where $\mathcal{H} = \left(\frac{\Lambda' \mathcal{A}_N \Lambda}{N}\right)\left(\frac{F'\hat{F}}{T}\right) V_{NT}^{-1}$. *Or equivalently,*

$$\omega_{NT}^2 \left(\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t - \mathcal{H}' F_t^0\right\|^2\right) = O_p(1),$$

where $\delta_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$ *and* $\omega_{NT} = \min\left\{\delta_{NT}, \delta_{NT}\mu_{NT}\right\}$.

where, V_{NT} be an $r \times r$ matrix consisting of the largest eigenvalues of $\frac{1}{NT}\mathbf{X}\left(I_N + \mu_{NT}\mathcal{S}\right)\mathbf{X}'$.

Asymptotic properties of Cn-PC

Theorem (Limiting Distribution)

Suppose that Assumptions A1-A7 hold.

1. *If* $\frac{\sqrt{N}}{T\mu_{NT}} \to 0$,

$$\sqrt{N}\left(\hat{F}_t - \mathcal{H}' F_t^0\right) \xrightarrow{d} N(0, V^{-1/2}\mathcal{Q}\Psi_t\mathcal{Q}'V^{-1/2}).$$

Efficiency of Cn-PC: $r = 1$

Cn-PC:

$$\hat{F}_{t,\text{Cn-PC}} \backsimeq \frac{F_t^0}{\sqrt{\Sigma_F}} + \frac{1}{\sqrt{N}} N\left(0, \frac{1}{\Sigma_F}\Sigma_{\Lambda*}^{-1}\Psi_t\Sigma_{\Lambda*}^{-1}\right)$$

OLS-PC:

$$\hat{F}_{t,\text{OLS-PC}} \backsimeq \frac{F_t^0}{\sqrt{\Sigma_F}} + \frac{1}{\sqrt{N}} N\left(0, \frac{1}{\Sigma_F}\Sigma_{\Lambda}^{-1}\Psi_t\Sigma_{\Lambda}^{-1}\right),$$

Note: $\Sigma_{\Lambda*} = \Sigma_\Lambda + \mu_{NT}$ plim $\frac{\Lambda'\mathcal{S}\Lambda}{N} \geq \Sigma_\Lambda$

Monte Carlo design (Bai& Liao (2016,JoE))

- The errors $\{u_{it}\}_{i\leq N,t\leq T}$ are iid as $N(0,1)$.
- The cross-sectional dynamics are generated as follows.

$$e_{1t} = u_{1t}, \quad e_{2t} = u_{2t} + a_1 u_{1t}, \quad e_{3t} = u_{3t} + a_2 u_{2t} + b_1 u_{1t},$$

$$e_{i+1,t} = u_{i+1,t} + a_i u_{i,t} + b_{i-1} u_{i-1,t} + c_{i-2} u_{i-2,t}$$

  where $\{a_i, b_i, c_i\}$ are independently drawn from $N(0, d^2)$ with $d = 0.7$
- The factors are independently generated from $N(0,1)$, and the loadings are i.i.d uniform on $[0,1]$.
- $\Omega_{1,1} = 1,\ \Omega_{2,2} = 1+a_1^2,\ \Omega_{3,3} = 1+a_2^2+b_1^2,\ \Omega_{i+1,i+1} = 1+a_i^2+b_{i-1}^2+c_{i-2}^2,$ and off diagonal elements,

  $\Omega_{1,2} = a_1,\ \omega_{1,3} = b_1,,\ \Omega_{2,3} = a_2 + a_1 b_1,$
  then for $i = 3, \cdots, N-1, \Omega_{i+1,i} = a_i + a_{i-1} b_{i-1} + c_{i-2} b_{i-2},$
  $\Omega_{i+1,i-1} = b_{i-1} + a_{i-2} c_{i-2},$
  $\Omega_{i+1,i-2} = c_{i-2}, \Omega_{i+1,i-3} = 0.$

Monte Carlo evaluation: Estimation

For $l = 1, \cdots, L(= 2000)$,

(i) Compute the OLS-PC estimators $\hat{F}^{(l)}_{OLS-PC}$, $\hat{\Lambda}^{(l)'}_{OLS-PC}$ and the estimated errors $\hat{e}^{l}_{OLS-PC}$. Using the sample covariance matrix $\hat{\Omega}_{OLS-PC} = \hat{e}'\hat{e}/NT$, construct an estimate for sign matrix, $\hat{\mathcal{S}}^{(l)}$.

(ii) For $M = m \cdot M_0$, where $m \in [0, 1]$, compute $\left(\hat{F}^{(l)}, \hat{\mu}^{l}_{NT}\right)$:

- Start $\mu_{NT} = \mu_0$, $\mu_0 = 0.5\sqrt{tr(\hat{e}'\hat{e})/tr(\hat{e}'\mathcal{A}_N\hat{e})}$, and $\mathcal{A}_\mu = I_N - \mu\mathcal{S}$, solve $\mathcal{L}(\mu)$:

$$\hat{\mu}_{NT} = \text{arg max}_\mu (NT)^{-1}\left[\text{tr } \mathbf{X}\mathcal{A}_\mu\mathbf{X}' - \text{tr } \hat{F}'_\mu\mathbf{X}\mathcal{A}_\mu\mathbf{X}'\hat{F}_\mu\right] - M,$$

where $\hat{F}_\mu$ is $\sqrt{T}$ times eigenvectors corresponding $\mathbf{X}\mathcal{A}_\mu\mathbf{X}'$.

(iii) Estimation accuracy:

- $S^{(l)}_{\hat{F},F^0} = \dfrac{\text{tr}\left(F^{0\prime}\hat{F}^{(l)}\left(\hat{F}^{(l)\prime}\hat{F}^{(l)}\right)^{-1}\hat{F}^{(l)\prime}F^0\right)}{\text{tr}(F^{0\prime}F^0)}$.
- *Small sample bias* of $\tilde{F}_t \equiv \mathcal{H}^{-1}\hat{F}_t$:bias$^{(l)} = \frac{1}{L}\sum_{l=1}^{L}\tilde{F}^{(l)}_{tk} - F^0_{tk}$, for $k = 1$ and $t = 1, [T/2], T$.
- *Empirical mean squared errors (MSEs)*: $MSEs^{(l)} = r^{-1}\left\|\hat{F}^{(l)}_t - F^{0(l)}_t\right\|^2$.

Results: Estimation accuracy

1. Sample Bias: overall, the proposed Cn-PC estimator have smaller bias compared to (OLS-PC). Table1.
2. % explained variation $S_{\hat{F},F^0}$
   - The GLS-PC performs better in case of T large and N small.
   - When N is large, GLS-PC performs poorly with $S_{\hat{F},F^0}$ considerably lower than the ones for the OLS-PC and Cn-PC estimators. Table2.
3. Sample covariances $\hat{\Omega}_{ij}$
   - The Cn-PC estimator's sample covariances are shrunk relative to the OLS-PC.
   - This shrinkage is less signifficant for the case of $N = T = 150$, although the spread is still small for the Cn-PC. Figure6
4. The sampling distribution of maximum average cross-correlation $\hat{\tau}^*$ show that the Cn-PC estimates are shrunk to zero relative to the OLS-PC. Figure7
5. Choosing M... Figure5 SffFig

Empirical comparison: Data

- Dataset consists of monthly observations on 125 U.S. macroeconomic time series from 1959:01 to 2009:01, and is sourced from the replication files of Stock & Watson (2012).
- Predictors include real variables such as sectoral industrial production, employment and hours worked; nominal variables such as consumer and price indexes, wages, money aggregates; in addition to stock prices and exchange rates.
  - Real variables: sectoral industrial production, employment and hours worked;
  - Nominal variables: consumer and price indices, wages, money aggregates;
  - Asset prices: stock prices and exchange rates
- Data transformed to achieve stationarity: for real variables, take the monthly growth rate (industrial production, sales $\cdots$) and first differences for variables already expressed in rates (unemployment rate, capacity utilization, $\cdots$)

Empirical comparison: data

- The sample has a monthly frequency from 1959:01 to 2009:01.
- The sample is divided into an in-sample portion of size $T = 120$ starting from $1959:01 - 1969:12$, and an out-of-sample evaluation portion with first date December 1970 and last date January 2009 with $R = 458$ out-of-sample evaluation points split into pre- and post-1985 periods with cut-off date December 1984.
- I consider rolling estimates with a window of 10 years, i.e. parameters are estimated at each time T using the most recent 10 years of data.

Forecasts Accuracy: relative MSFEs

- It is observed that the gains in forecast accuracy depend on the sample period and on the target series. Generally, the gains are not significant and range from 0% to 6% decrease in the pseudo-out-of-sample mean-squared forecast errors.
- Consumer price Index forecasts appear to benefit the most from incorporating dependence features using the Cn-PC estimators of the predictors $\hat{F}_t$. These benefits are more appreciable during the period of post moderation of 1985-2002. Table3

Concluding remarks

- I presented a simple and novel PC-based method for estimating for large approximate factor models with cross sectional dynamics in the errors.
- The estimator is computationally more tractable than ML-based alternatives. It doesn't require estimating large covariance matrices and is obtained by performing PC of a regularised data matrix.
- Monte carlo results suggest the estimator outperforms the OLS-PC and the GLS-PC in estimating the space of the true factors.
- Applied to forecasting, small improvements oer OLS-PC. Perhaps more gains in the context of APT and pricing of returns.

THANK YOU...

To summarize,

$$\hat{F}_{\hat{\mu}_{NT}} \quad : \quad \sqrt{T} \times \text{first } r \text{ principal components of } \mathbf{X}\left(I_N + \hat{\mu}_{NT}\mathcal{S})\right)\mathbf{X}', \tag{6}$$

$$\hat{\Lambda}_{\hat{\mu}_{NT}} \quad : \quad \hat{\Lambda}_{\hat{\mu}_{NT}} = \mathbf{X}'\hat{F}_{\hat{\mu}_{NT}}/T, \tag{7}$$

$$\mu \quad : \quad M = (NT)^{-1}\sum_{t=1}^{N} \underline{\hat{e}}_t'\mathcal{S}\underline{\hat{e}}_t \tag{8}$$

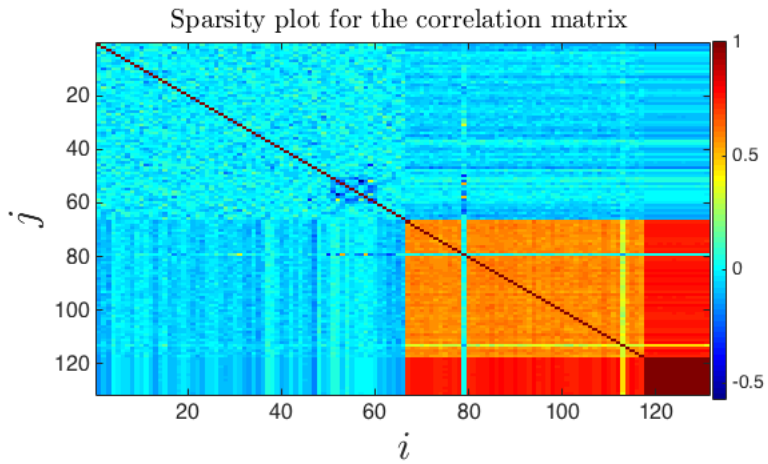
I solve for $(\hat{F}_{\hat{\mu}}, \hat{\mu})$ which minimizes the reduced Lagrangian $\mathcal{L}(F, \mu)$ in (**??**) subject to the constraint $F'F/T = I_r$. The problem can be solve as in the standard primal-dual procedure, whereby the Lagrangian is further concentrated to a reduced function of μ, after replacing F by $\hat{F}(\mu)$. The dual problem solves the maximum of the concentrated objective function, $\mathcal{L}(\mu)$, which is equal to:

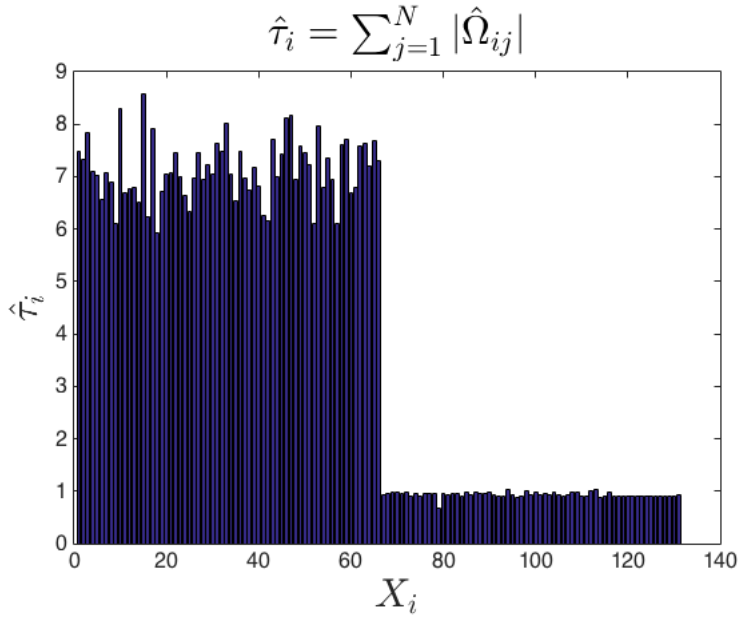
$$(NT)^{-1}\left[\text{tr } X\left(I_N + \hat{\mu}_{NT}\mathcal{S}\right)X' - \text{tr } \hat{F}'_{\hat{\mu}_{NT}}\left(\mathbf{X}\left(I_N + \hat{\mu}_{NT}\mathcal{S}\right)\mathbf{X}'\right)\hat{F}_{\hat{\mu}_{NT}}\right] - \hat{\mu}_{NT}M \tag{9}$$

Back to The Cn-PC Estimator.

Example: Crosse section of macroeconomic variables used in forecasting

Figure: Sparsity of the sample covariance matrix and total dependence, $\hat{\tau}_i$, for series X_i, $i = 1, \cdots, N$





Back to Efficiency.

- **Macro 1: Macroeconomic forecasting using many predictors.** Macro 1.
- **Macro 2: The determinants of economic growth in a cross-section of countries.** The database includes data for 90 countries and 60 potential predictors, Barro and Lee (1994).
- **Finance 1: Equity premium prediction.** study the predictability of US aggregate stock returns
- **Finance 2: Explaining the cross section of expected returns.**
- **Micro 1: Understanding the decline in crime rates in US states in the 1990s.**
- **Micro 2: The determinants of government takings of private property in US judicial circuits**

Back to Introduction.

Macroeconomic forecasting

Aim. Forecasting monthly growth rate of US industrial production, Consumer Price index $y_{t+h|\mathcal{I}_t}$

- Dataset consists of monthly observations on 130 U.S. macroeconomic time series from 1959:01 to 2009:01, and is sourced from the replication files of Stock & Watson (2012).
- Predictors include:
  - Real variables: sectoral industrial production, employment and hours worked;
  - Nominal variables: consumer and price indices, wages, money aggregates;
  - Asset prices: stock prices and exchange rates
- Forecasting with many predictors provides opportunity to exploit a much richer base of information than is conventionally used for time series forecasting.
- Using many predictors may provide some robustness against structural instability that plagues low-dimensional forecasting.
- However, many predictors forecasting brings substantial challenges: many parameters which overwhelms the information in the data with estimation error.

Back to Examples.

The curse of dimensionality

- Let Z_t be the $N-$vector of covariance stationary processes. We are interested in estimating the linear projection, $y_{t+h} = proj\{y_{t+h}|Z_{t-s}, s = 0, 1, 2, \cdots\}$.
- Traditional time series methods approximate the projection using a finite number, p, of lags of Z_t,

$$y_{t+h} = Z_t'\beta_0 + \cdots + Z_{t-p}'\beta_p + u_{t+h} = X_t'\beta + u_{t+h},$$

- Given a sample of size T, let $X = (X_{p+1}, \cdots, X_{T-h})'$ be the $(T-h-p) \times N(p+1)$ matrix of observations for the predictors and $y = (y_{p+h+1}, \cdots, y_T)'$. The traditional forecast is given by $\widehat{y}^{LS}_{T+h|T} = X'\widehat{\beta}^{LS}$, where $\widehat{\beta}^{LS} = (X'X)^{-1}X'y$.

Back to Introduction.

Factor Models and factors based forecasts

Let X_{it} be the **observed** data for the i^{th} cross-section unit at time t $(i = 1, \cdots, N, t = 1, \cdots, T)$. Consider

$$X_{it} = \lambda_i' F_t + e_{it}, \tag{10}$$

where $F_t = \{F_{kt}\}_{1 \leq k \leq r}$, is an $r \times 1$ vector of common factors, $\lambda_i = \{\lambda_{ik}\}_{1 \leq k \leq r}$ is the corresponding vector of factor loading for cross-section unit i, and e_{it} is an idiosyncratic component.

- Let us assume that there are r common fcators driving the co-movements in the the data matrix $X = \{X_1, \cdots, X_i, \cdots, X_N\}'$, $X_i = [X_{i1}, \cdots, X_{it}, \cdots, X_{iT}]'$:

$$X_{it} = \lambda_i' F_t + e_{it}$$

  F_t are called common factors, λ_i are called factor loadings

- Let $\widehat{F}$ be the $T \times r$ matrix of factors estimates, and define $\mathfrak{I}_t^f = span\{\widehat{f}_{1t}, \cdots, \widehat{f}_{rt}\}$, with $r \ll m \times (p+1)$ is a parsimonious representation of the information set $\mathfrak{I}_t$.

  The principal component forecast is defined as:

$$y_{T+h|T}^{PC} = proj\{y_{T+h} | \mathfrak{I}_T^f\}. \tag{11}$$

  The projection is computed by OLS of y_t on $\widehat{F}_t$ for $t = 1 : T$:

$$y_{T+h|T}^{PC} = \widehat{\theta}' \widehat{F}_T, \tag{12}$$

$$\widehat{\theta} = (\widehat{F}\widehat{F}')^{-1} \widehat{F}_T' y, \quad \widehat{F} = (\widehat{f}_{1T}, \cdots, \widehat{f}_{rT})'. \tag{13}$$

Back to Introduction.

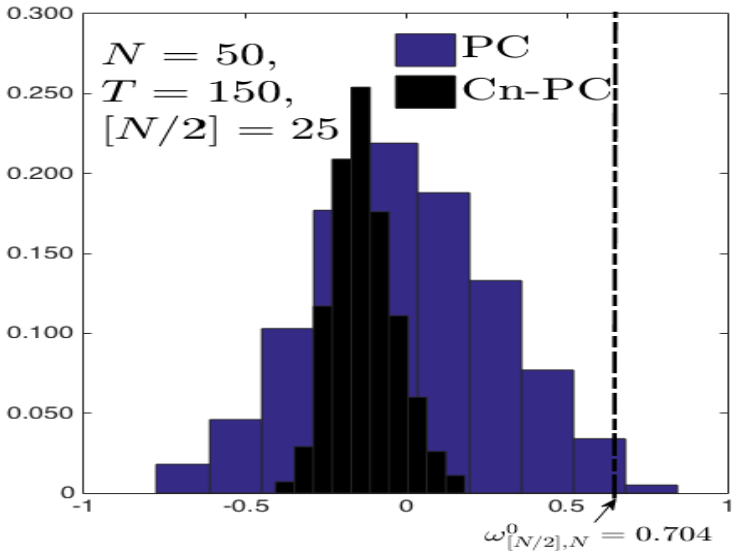
Table 1: Small sample bias and standard errors for the estimated factors $\tilde{F}_{t,1}$, for $t = [T/2], T$ and $r = 1$

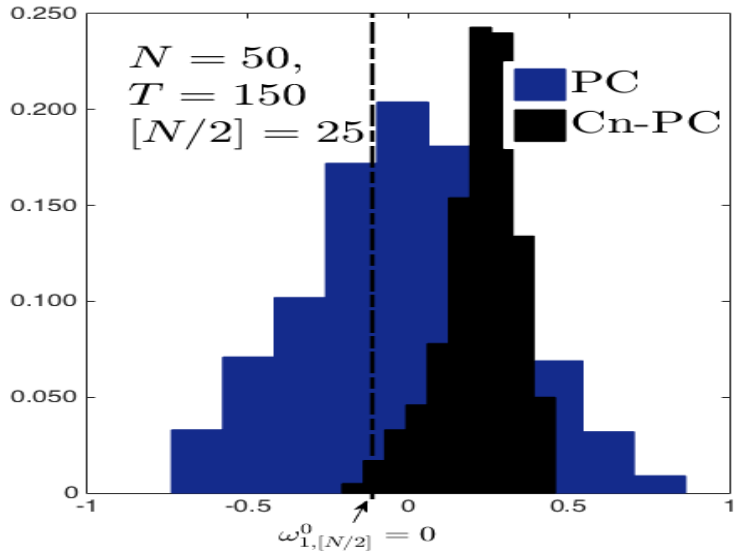
		Cn-PC				PCE			
		$\tilde{F}_{[T/2],1}$		$\tilde{F}_{T,1}$		$\tilde{F}_{[T/2],1}$		$\tilde{F}_{T,1}$	
T	N	bias	std	bias	std	bias	std	bias	std
50	50	-0.018	0.190	-0.092	0.163	-0.024	0.037	-0.139	0.097
	100	0.001	0.179	-0.207	0.092	0.033	0.031	-0.025	0.008
	150	-0.155	0.136	0.293	0.137	-0.211	0.103	0.353	0.166
100	50	-0.004	0.118	0.008	0.092	-0.046	0.025	0.121	0.015
	100	-0.128	0.102	-0.109	0.106	-0.120	0.079	-0.114	0.054
	150	-0.168	0.105	-0.049	0.115	-0.126	0.063	-0.013	0.053
150	50	-0.007	0.089	0.026	0.078	-0.032	0.053	0.070	0.081
	100	0.018	0.097	-0.113	0.086	0.054	0.022	-0.180	0.032
	150	0.093	0.062	0.031	0.065	-0.000	0.020	-0.065	0.021

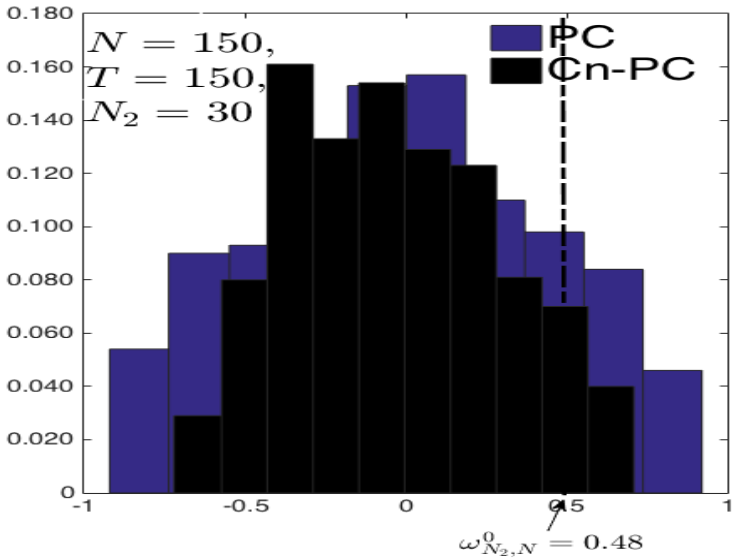
The results are for the sampling distribution of $\tilde{F}_t = \mathcal{J}^{-1}\hat{F}_t$, $\mathcal{J} = \mathcal{H}$ for Cn-PC and $\mathcal{J} = H$ for PCE. The shrinkage factor M is chosen by a 10−fold cross-validation.

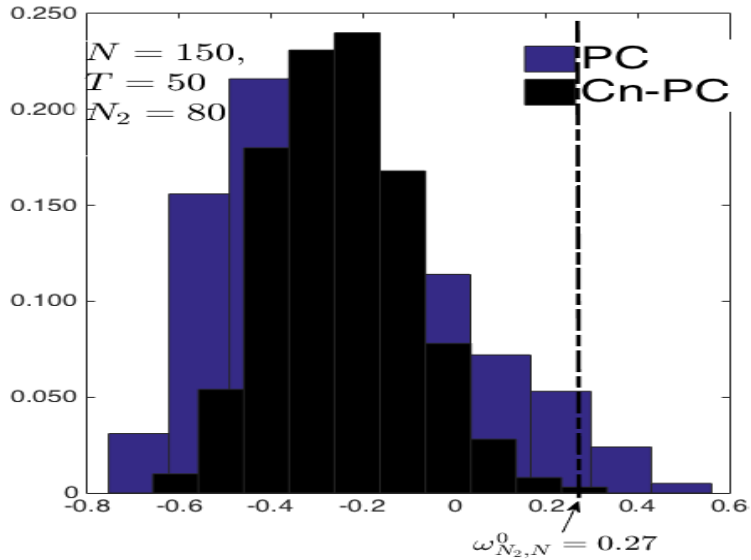
Back to Results1.

Figure 6: Distribution of $\hat{\Omega}_{i,j}$



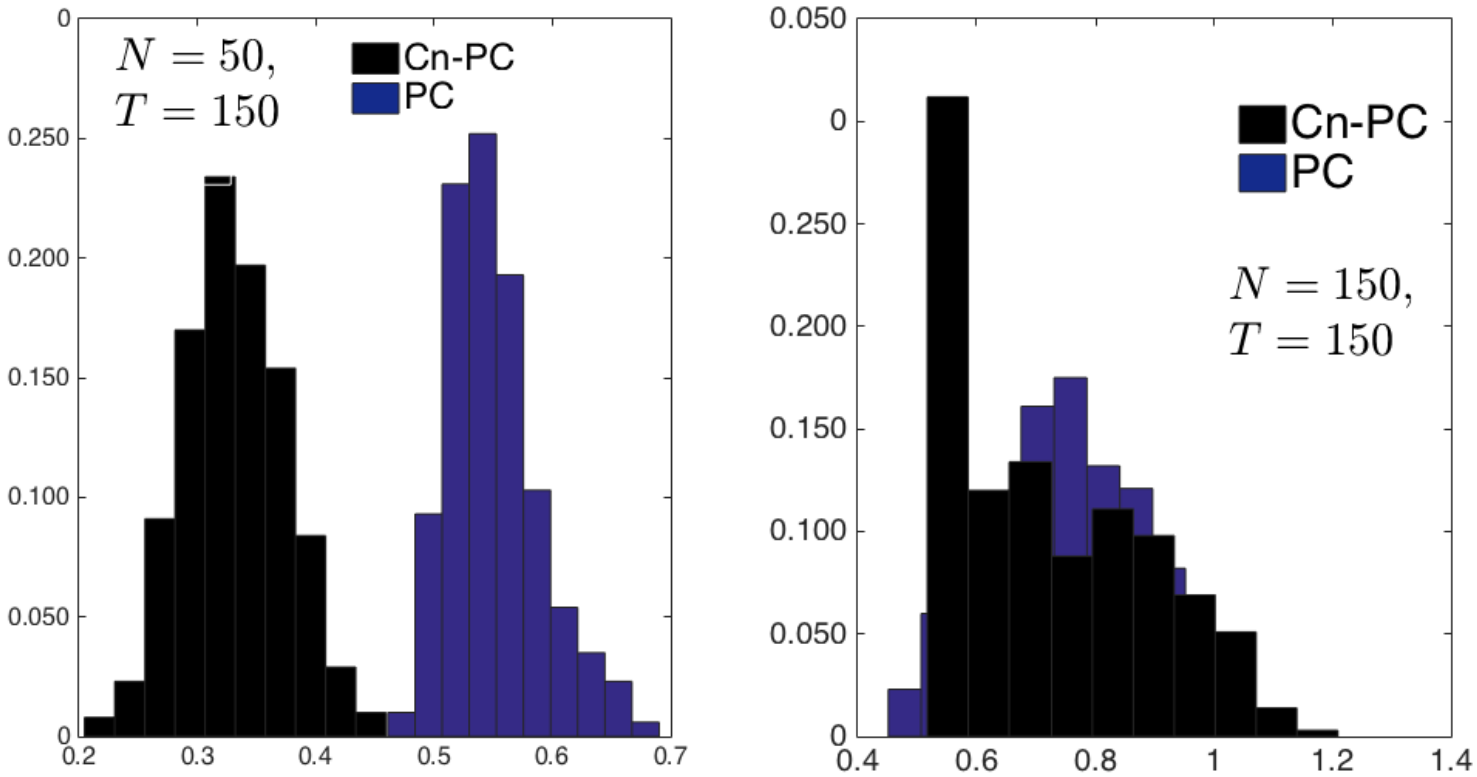






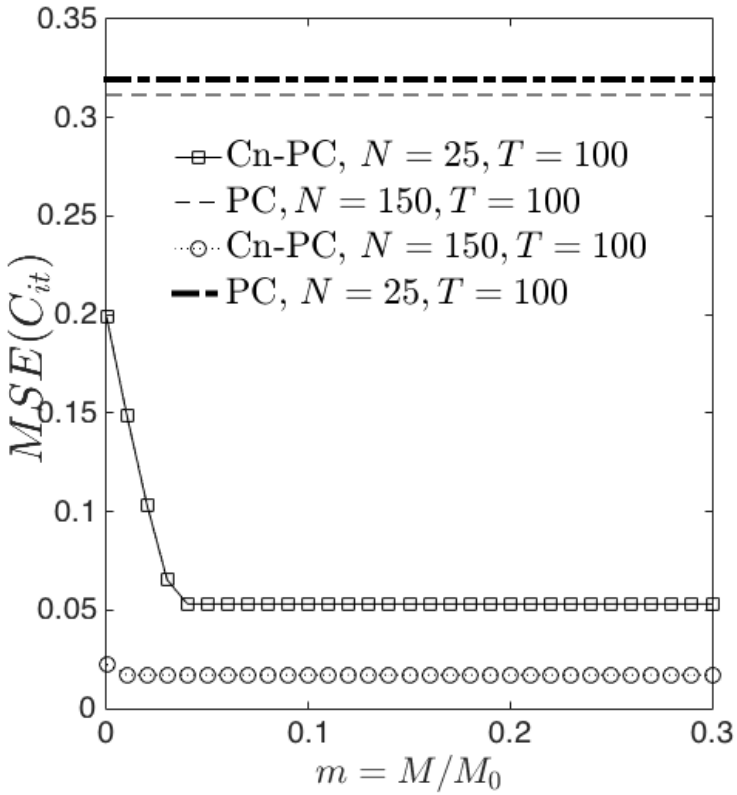
Back to Results1.

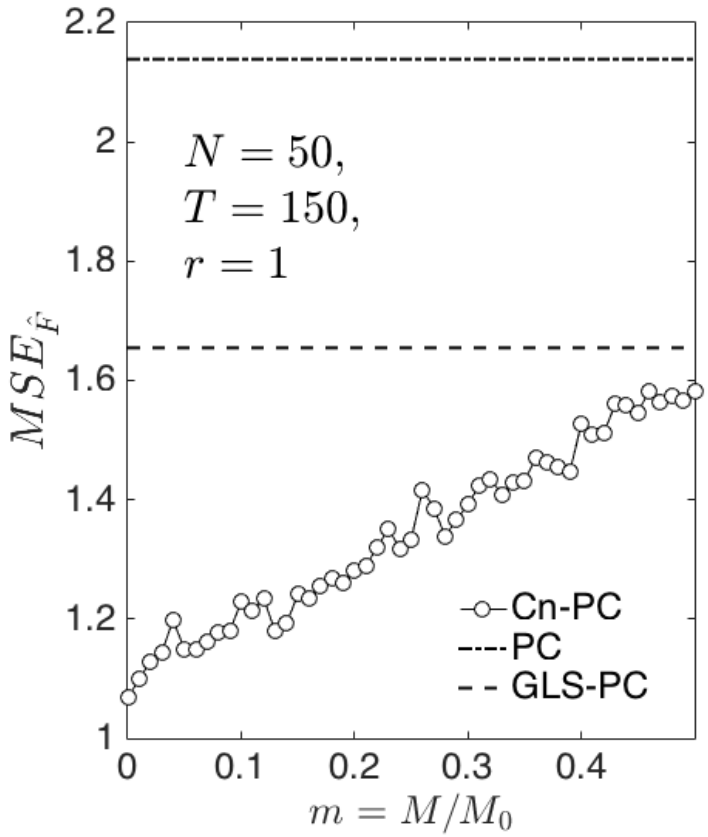
Figure 7: Sampling distribution of $\hat{\tau}^*$



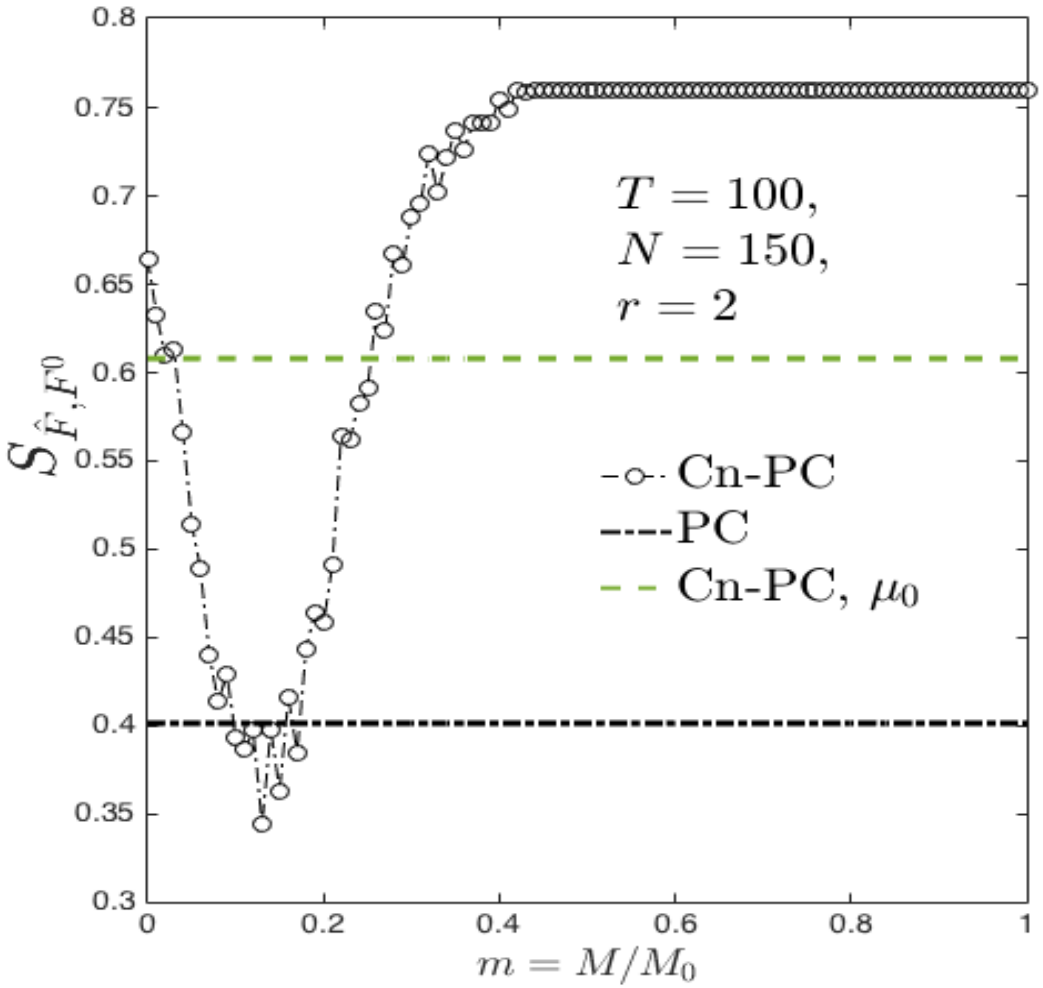
Back to Results1.

Figure 5: Empirical mean squared errors (MSEs) of Cn-PC estimators of common factors and common components $\hat{C}$





Back to Results1.



Back to Results1.

Table 2: Efficiency of estimated common factors: $S_{\hat{F},F^0}$ and $MSE_{\hat{F}}$

		$S_{\hat{F},F^0}$			$MSE_{\hat{F}}$		
T	N	PC	Cn-PC	PC-GLS	PC	Cn-PC	PC-GLS
100	25	0.13	0.319	0.434	2.17	2.16	2.13
	50	0.12	0.382	0.157	1.84	1.76	1.77
150	50	0.10	0.337	0.185	1.78	1.95	1.97
	100	0.10	0.580	0.078	1.83	1.89	1.94
55	50	0.24	0.505	0.072	1.94	1.95	1.75
50	25	0.26	0.341	0.252	1.96	1.36	2.02

Back to Results1.

Table 3: Pseudo-out-of-sample mean squared forecasts errors for US inflation and industrial production

		IPS10				PUNEW			
		$r = 10$		$r = 5$		$r = 10$		$r = 5$	
h=12		PC	Cn-PC	PC	Cn-PC	PC	Cn-PC	PC	Cn-PC
1970-2002	$MSFE$	0.51	0.51	0.52	0.50	0.64	0.62	0.57	0.57
	Var	0.85	0.85	0.66	0.66	0.53	0.53	0.60	0.60
1970-1985	$MSFE$	0.32	0.31	0.31	0.31	0.43	0.40	0.38	0.38
	Var	0.95	0.94	0.75	0.75	0.45	0.45	0.56	0.56
1985-2002	$MSFE$	1.09	1.08	1.13	1.11	1.65	1.63	1.46	1.40
	Var	0.53	0.50	0.39	0.43	0.87	0.85	0.77	0.75

		IPS10				PUNEW			
		$h = 1$		$h = 4$		$h = 1$		$h = 4$	
r=7		PC	Cn-PC	PC	Cn-PC	PC	Cn-PC	PC	Cn-PC
1970-2002	$MSFE$	0.72	0.70	0.57	0.57	0.78	0.75	0.67	0.67
	Var	0.42	0.38	0.56	0.56	0.27	0.27	0.37	0.37
1970-1985	$MSFE$	0.66	0.61	0.49	0.49	0.75	0.71	0.56	0.55
	Var	0.46	0.43	0.56	0.56	0.26	0.25	0.42	0.41
1985-2002	$MSFE$	0.86	0.86	0.86	0.86	0.82	0.82	0.97	0.97
	Var	0.28	0.28	0.54	0.54	0.28	0.28	0.25	0.25

Back to Empirics.

Appendix A1: Cn-PC Estimators

The critical points of the function (3.10) are found by solving the first order conditions on the feasible set:

$$(I): \frac{\partial \mathcal{L}(\Lambda, F)}{\partial \Lambda}|_{\hat{\Lambda},\hat{F}} = 0 \tag{7.1}$$

$$(II): \frac{\partial \mathcal{L}(\Lambda, F)}{\partial F}|_{\hat{\Lambda},\hat{F}} = 0 \tag{7.2}$$

$$M \geq (NT)^{-1} \sum_{t=1}^{N} \underline{\hat{e}}_t' \mathcal{S} \underline{\hat{e}}_t, \quad \hat{\mu}_{NT} \geq 0, \quad \hat{\mu}_{NT} \left(M - (NT)^{-1} \sum_{t=1}^{N} \underline{\hat{e}}_t' \mathcal{S} \underline{\hat{e}}_t \right) = 0 \tag{7.3}$$

The conditions in (7.3) are known as the complementary slackness. The first two sets of conditions in (7.4) and (7.6), lead to the following:

$$(I) \quad : \quad \sum_{t=1}^{T} \left(I_N + \hat{\mu}_{NT} \mathcal{S} \right) \underline{\hat{e}}_t \hat{F}_t' = 0 \tag{7.4}$$

$$\hat{\Lambda} = \left(\sum_{t=1}^{T} \underline{X}_t F_t' \right) \left(\sum_{t=1}^{T} F_t F_t' \right)^{-1} \tag{7.5}$$

$$(II) \quad : \quad \sum_{t=1}^{T} \hat{\Lambda}' \left(I_N + \hat{\mu}_{NT} \mathcal{S} \right) \underline{\hat{e}}_t = 0 \tag{7.6}$$

$$\hat{F}_t = \left(\hat{\Lambda}' \left(I_N + \hat{\mu}_{NT} \mathcal{S} \right) \hat{\Lambda} \right)^{-1} \hat{\Lambda}' \left(I_N + \hat{\mu}_{NT} \mathcal{S} \right) \underline{X}_t \tag{7.7}$$

Substituting (7.5) into the Lagrangian and imposing the identification restriction $F'F/T = I_r$, this concentrates out Λ to obtain a reduced Lagrangian that is a function of F and μ: