Structural Analysis with Multivariate Autoregressive Index Models

Andrea Carriero [1] *George Kapetanios* [2] *Massimiliano Marcellino* [3]

September 25, 2015

[1] Queen Mary, University of London
[2] Queen Mary, University of London
[3] Universita' Bocconi, IGIER and CEPR

Introduction

- Econometric models for large datasets widely used in applied econometrics literature
- A large information set helps in structural analysis:
  - Large datasets better reflect the information set of central banks and the private sector
  - Large models allow to study the effect of shocks on a wide range of variables
- A large information set helps in improving forecast accuracy
- Two main approaches to deal with overparameterization: factor models and BVARs

Factor models

- Large scale: Forni, Hallin, Lippi, and Reichlin (2000), Stock and Watson (2002)
- Often two step approach (estimate factors, then treat them as known), though full ML possible, e.g. Doz, Giannone, and Reichlin (2006)
- Relies on N diverging for consistent estimation
- Conditions on the idiosyncratic and common component are required
- Complex to identify economically the factors, e.g. Bai and Ng (2006, 2010), though structural FAVAR is a solution, e.g. Forni et al. (2009), Gambetti and Forni (2010)

BVARs

- Large Bayesian VARs offer an alternative to factor models. Feasible with a conjugate prior (Banbura, Giannone, Reichlin (2010))
- BVARs perform well in forecasting
- In a large system it can be difficult to identify some shocks
- A structural shock is modelled as a shock to one particular variable
    - The choice of a specific data series to represent a general economic concept (e.g. "real activity") is often arbitrary to some degree

Multivariate Autoregressive Index (MAI) models

- MAI models proposed by Reinsel (1983) bridge VARs and factor models by imposing a rank reduction on a VAR
- Reduced rank regressions have been considered in Anderson (1951) and Geweke (1996). The proposed way to impose rank reduction in MAI models differs from these approaches in two respects:
  - Makes the MAI similar to a factor model
  - Allows to give the factors an economic interpretation which facilitates structural analysis
- Moreover, MAI models
  - Do not rely on N diverging for consistency
  - Do not require conditions on the idiosyncratic and common component
- We review estimation via ML and study the case of N large, provide an MCMC algorithm for Bayesian estimation, and show how MAI models can be used for structural analysis

Multivariate Autoregressive Index model

- Consider a VAR for a N-dimensional vector $Y_t = (y_{1,t}, y_{2,t}, ..., y_{N,t})'$:

$$Y_t = \Phi(L) Y_t + \epsilon_t, \tag{1}$$

where $\Phi(L) = \Phi_1 L + ... + \Phi_p L^p$ and ϵ_t are i.i.d. $N(0, \Sigma)$

- Assume $\Phi(L) = A(L) B_0$, where $A(L) = A_1 L + ... + A_p L^p$, each A_u is $N \times r$, B_0 is $r \times N$ with rank r. Then:

$$Y_t = \sum_{u=1}^{p} A_u B_0 Y_{t-u} + \epsilon_t \tag{2}$$

- If r much smaller than N, the MAI has much fewer parameters than the VAR. For example, if $N = 20$, $p = 13$, and $r = 3$, there are $N^2 p = 5200$ parameters in the VAR and $Nr(p+1) = 840$ in the MAI
- Reinsel (1983) studied ML estimation of this model

MAI models and factors

- Recall the model:

$$Y_t = A(L)B_0 Y_t = \sum_{u=1}^{p} A_u B_0 Y_{t-u} + \epsilon_t \tag{3}$$

- Defining:

$$F_t = B_0 Y_t \tag{4}$$

  we have:

$$Y_t = A(L)F_t + \epsilon_t = \sum_{u=1}^{p} A_u F_{t-u} + \epsilon_t \tag{5}$$

- As in factor models, the loadings A_u and the factor weights B_0 are not uniquely identified, we set $B_0 = (I_r,\ \widetilde{B}_0)$
- Importantly, restrictions on $\widetilde{B}_0$ can be easily imposed

Data and restrictions on B

Variable	FRED code	F1	F2	F3
Employees on nonfarm payroll	PAYEMS	1	0	0
Average hourly earnings	AHETPI	$b_{1,2}$	0	0
Personal income	A229RX0	$b_{1,3}$	0	0
Real Consumption	PCE÷PCEPI	$b_{1,4}$	0	0
Industrial Production Index	INDPRO	$b_{1,5}$	0	0
Capacity Utilization	TCU	$b_{1,6}$	0	0
Unemployment rate	UNRATE	$b_{1,7}$	0	0
Housing starts	HOUST	$b_{1,8}$	0	0
CPI all items	CPIAUCSL	0	1	0
Producer Price Index (finished goods)	PPIFGS	0	$b_{2,10}$	0
Implicit price deflator for personal consumption expenditures	PCEPI	0	$b_{2,11}$	0
PPI ex food and energy	PPILFE	0	$b_{2,12}$	0
Federal Funds, effective	FEDFUNDS	0	0	1
M1 money stock	M1SL	0	0	$b_{3,14}$
M2 money stock	M2SL	0	0	$b_{3,15}$
Total reserves of depository institutions	TOTRESNS	0	0	$b_{3,16}$
Nonborrowed reserves of depository institutions	NONBORRES	0	0	$b_{3,17}$
S&P's common stock price index	S&P	0	0	$b_{3,18}$
Interest rate on treasury bills, 10 year constant maturity	GS10	0	0	$b_{3,19}$
Effective Echange rate	CCRETT01USM661N	0	0	$b_{3,20}$

Factor dynamics

- The factors $F_t = B_0 Y_t$ have closed form $VAR(p)$ representation, obtained by pre-multiplying (5) by B_0:

$$F_t = B_0 \sum_{u=1}^{p} A_u F_{t-u} + B_0 \epsilon_t = C(L) F_t + u_t \tag{6}$$

where

$$C(L) = B_0 A_1 L + B_0 A_2 L^2 + ... + B_0 A_p L^p, \tag{7}$$

and

$$u_t = B_0 \epsilon_t; \; u_t \sim i.i.d. N(0, \Omega); \; \Omega = B_0 \Sigma B_0'. \tag{8}$$

- Note both factors and data follow a VAR. This does not happen in factor models (Dufour and Stevanovic, 2010)

MA representation (1)

- The factors have the following MA representation:

$$F_t = (I - C(L))^{-1} u_t = (I - B_0 A(L))^{-1} B_0 \epsilon_t \tag{9}$$

- Therefore the moving average representation of $Y_t = A(L)F_t + \epsilon_t$ is:

$$Y_t = (A(L)(I - B_0 A(L))^{-1} B_0 + I)\epsilon_t. \tag{10}$$

- Representation (10) is similar to the one used in the BVAR literature. There are as many shocks as variables (N)

MA representation (2)

- Define the matrix $B_{0\perp}$ as the $(N-r)\times N$ full row rank matrix orthogonal to B_0. Then, consider the following decomposition (Centoni and Cubadda 2003):

$$\Sigma B_0'(B_0\Sigma B_0')^{-1}B_0 + B_{0\perp}'(B_{0\perp}\Sigma^{-1}B_{0\perp}')^{-1}B_{0\perp}\Sigma^{-1} = I_N. \tag{11}$$

- This key identity can now be inserted into the Wold representation in (10) to yield:

$$Y_t = (\Sigma B_0'\Omega^{-1} + A(L)(I - B_0A(L))^{-1})u_t + B_{0\perp}'(B_{0\perp}\Sigma^{-1}B_{0\perp}')^{-1}\xi_t, \tag{12}$$

where $u_t = B_0\epsilon_t$, $\xi_t = B_{0\perp}\Sigma^{-1}\epsilon_t$, and $\Omega = B_0\Sigma B_0'$.

- The representation in (12) shows that each element of Y_t is driven by a set of r common errors, the u_t that are the drivers of the factors F_t, and by linear combinations of ξ_t. Since

$$E(u_t\xi_t') = E(B_0\epsilon_t\epsilon_t'\Sigma^{-1}B_{0\perp}') = 0, \tag{13}$$

$$E(u_{t-i}\xi_t') = 0, \quad E(u_t\xi_{t-i}') = 0, \quad i > 0, \tag{14}$$

u_t and ξ_t are uncorrelated at all leads and lags.

Relation with factor models

- In summary, the *MAI* is close to the generalized dynamic factor model of Forni, Hallin, Lippi, and Reichlin (2000) and Stock and Watson (2002a, 2002b), and even more to the parametric versions of these models in the FAVAR literature, e.g. Bernanke et al. (2005) and Kose et al. (2005))
- Can answer questions similar to those considered by Forni et al. (2009), Forni and Gambetti (2010) using structural factor models
- But also possibly relevant differences

Relation with factor models

- Imposing economically meaningful restrictions on the factors F_t, such as equality of one factor to a specific economic variable, or group of variables, can be much simpler in the *MAI* context
- In the factor literature factors are unobservable and can be consistently estimated only when N diverges. Within an *MAI* context it is possible to consistently estimate the factors with N finite
- In the factor literature consistency requires conditions on the common and idiosyncratic components. For the *MAI* standard ML results apply

Relation with multivariate regressions

- Reduced rank regressions have been considered in Anderson (1951), Velu et al. (1986), and Geweke (1996). Consider, again:

$$Y_t = \Phi(L) Y_t + \epsilon_t, \tag{15}$$

- Assume $\Phi(L) = A_1 B(L)$, where $B(L) = B_0 L + B_1 L^2 + ... + B_{p-1} L^p$, A_1 is $N \times r$, each B_v is $r \times N$. Defining $X_t = (Y'_{t-1}, ..., Y'_{t-p})'$, the resulting model can be written as:

$$\underset{N\times 1}{Y_t} = \underset{N\times r}{A_1} \underset{r\times Np}{[B_0, \ ..., B_{p-1}\ ]} \underset{Np\times 1}{X_t} + \underset{N\times 1}{\epsilon_t}, \tag{16}$$

- It is useful to compare (16) with the MAI model:

$$\underset{N\times 1}{Y_t} = \underset{N\times rp}{[A_1, ..., A_p]} \underset{rp\times Np}{(I_p \otimes B'_0)'} \underset{Np\times 1}{X_t} + \underset{N\times 1}{\epsilon_t}. \tag{17}$$

- Estimation of (16) is easier than estimation of the MAI model, but the MAI model allows to derive a finite order VAR representations for a set of r factors.

Estimation

- For estimation, we compactly rewrite the *MAI* as:

$$Y_t = AZ_{t-1} + \epsilon_t, \tag{18}$$

where:

$$\begin{aligned} Z'_{t-1} &= (F'_{t-1}, ..., F'_{t-p}) = (Y'_{t-1}B'_0, ..., Y'_{t-p}B'_0) = (Y'_{t-1}, ..., Y'_{t-p})(I_p \otimes B'_0) \\ B_0 &= (I_r,\ \widetilde{B}_0) \\ A &= (A_1, ..., A_p) \end{aligned}$$

Estimation via Maximum Likelihood

- The likelihood function is:

$$-0.5\,T\log|\Sigma| - 0.5\sum_{t=1}^{T}(Y_t - AZ_{t-1})'\Sigma^{-1}(Y_t - AZ_{t-1}), \tag{19}$$

where $Z'_{t-1} = (Y'_{t-1},..., Y'_{t-p})(I_p \otimes B'_0)$ and $B_0 = (I_r,\ \widetilde{B}_0)$

- Reinsel (1983) studies this model extensively. He provides the FOCs and updating rule for the gradient of the ML estimator for this case
- ML estimates can also be obtained by iterating over the first order conditions of the maximization problem with respect to A , $\tilde{B}_0$, and Σ
- In the paper, we extend the consistency proof to the case of N diverging

Estimation via Markov Chain Monte Carlo

- Recall the model:

$$Y_t = AZ_{t-1} + \epsilon_t, \tag{20}$$

where $Z'_{t-1} = (Y'_{t-1}, ..., Y'_{t-p})(I_p \otimes B'_0)$ and $B_0 = (I_r,\ \widetilde{B}_0)$

- The model contains three sets of parameters, in the matrices A, $\widetilde{B}_0$, and Σ. The joint posterior distribution $p(A', \widetilde{B}_0, \Sigma | Y)$ has not a known form, but it can be simulated by drawing in turn from:

$$p(A', \Sigma | \widetilde{B}_0, Y) \tag{21}$$

$$p(\widetilde{B}_0 | A', \Sigma, Y) \tag{22}$$

- Draws from (21) can be obtained using $p(\Sigma | \widetilde{B}_0, Y)$ and $p(A' | \Sigma, \widetilde{B}_0, Y)$, which are both available given a suitable choice of the prior (conjugate)
  - Conjugate N-IW prior
- Draws from (22) can be obtained via a RW-Metropolis step
  - Prior based on auxiliary model on pre-sample

Determining the rank of the system - Classical

- Two main approaches: information criteria or sequential testing
- Standard info criteria can be used. An attractive feature is that both the rank r and the number of lags p can be jointly determined
- Sequential testing: starting with the null hypothesis of $r = 1$, a sequence of tests is performed. If the null hypothesis is rejected, r is augmented by one and the test is repeated until the null cannot be rejected

Determining the rank of the system - Bayesian

- Compute the marginal data density $p_r(Y)$ as a function of the chosen r. The optimal rank can be obtained as:

$$r^* = \arg\max_r p_r(Y), \tag{23}$$

  note r^* corresponds to the posterior mode of r under a prior assigning equal probabilities to each candidate rank
- The density $p_r(Y)$ can be efficiently approximated numerically by using Rao-Blackwellization and the harmonic mean estimator, as in Fuentes-Albero and Melosi (2013).
- The lag length can be chosen similarly

Monte Carlo evaluation

- We produce artificial data from two alternative DGPs. Recall:

$$Y_t = \sum_{u=1}^{p} \Phi_u Y_{t-u} + \epsilon_t, \ \epsilon_t \sim i.i.d.N(0, \Sigma). \tag{24}$$

- DGP1 is an unrestricted VAR, so it uses (24) without imposing any further restriction
- DGP2 is the MAI, so it imposes the rank reduction restriction:

$$\Phi_u = A_u B_0, \quad u = 1, ..., p. \tag{25}$$

- For each DGPs we estimate i) the MAI under the Bayesian approach, ii) the MAI under the classical approach, iii) an unrestricted BVAR

Monte Carlo evaluation - results

- We focus on the Root Mean Squared Error ($RMSE$) and Mean Absolute Error (MAE) arising from estimation of the conditional mean parameters $\Phi_1, ..., \Phi_p$
- We evaluate the performance along various dimensions, considering different values for the total number of variables N, the number of observations T, and the system rank r
- Overall, the Monte Carlo experiments suggest that Bayesian estimation of the MAI model is systematically better than classical estimation
- The ranking of the MAI and full rank BVAR models is -as one would expect- dependent on the true DGP

Table 1. MC results under the MAI DGP

PANEL A- r=3, increasing N and T

	N=5			**N=10**		
	T=300	**T=460**	**T=720**	**T=300**	**T=460**	**T=720**
RMSE						
Bayesian MAI	0.76	0.74	0.74	0.74	0.74	0.74
Classical MAI	6.23	4.75	3.64	4.69	3.56	2.70
BVAR (benchmark)	0.009	0.010	0.009	0.011	0.011	0.011
MAE						
Bayesian MAI	0.90	0.89	0.88	0.84	0.84	0.84
Classical MAI	5.94	4.53	3.49	4.25	3.23	2.48
BVAR (benchmark)	0.008	0.008	0.008	0.009	0.009	0.009
	N=15			**N=20**		
	T=300	**T=460**	**T=720**	**T=300**	**T=460**	**T=720**
RMSE						
Bayesian MAI	0.53	0.48	0.43	0.49	0.44	0.39
Classical MAI	4.99	3.64	2.80	4.28	2.89	2.08
BVAR (benchmark)	0.010	0.010	0.010	0.010	0.010	0.010
MAE						
Bayesian MAI	0.52	0.48	0.43	0.48	0.43	0.39
Classical MAI	4.47	3.30	2.55	3.86	2.63	1.88
BVAR (benchmark)	0.008	0.008	0.008	0.008	0.008	0.008

Table 2. MC results under the VAR DGP

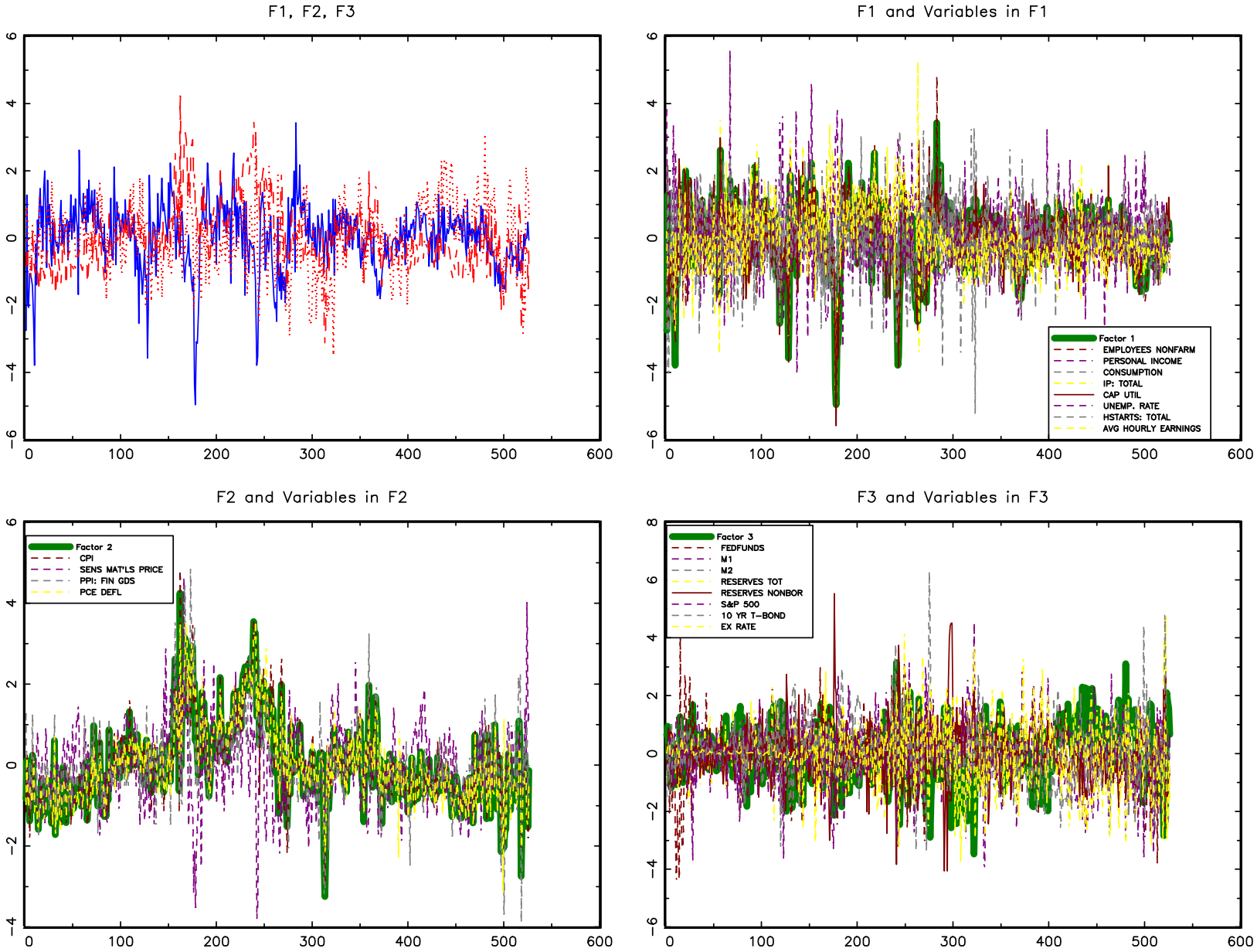
PANEL A- r=3, increasing N and T

	N=5			N=10		
	T=300	T=460	T=720	T=300	T=460	T=720
RMSE						
Bayesian MAI	1.45	1.43	1.51	1.33	1.38	1.37
Classical MAI	4.84	3.86	3.22	4.57	3.51	2.82
BVAR (benchmark)	0.012	0.011	0.011	0.011	0.010	0.010
MAE						
Bayesian MAI	1.52	1.57	1.65	1.38	1.44	1.48
Classical MAI	4.48	3.65	3.14	4.22	3.33	2.74
BVAR (benchmark)	0.010	0.010	0.009	0.009	0.009	0.008

	N=15			N=20		
	T=300	T=460	T=720	T=300	T=460	T=720
RMSE						
Bayesian MAI	1.21	1.22	1.19	1.19	1.17	1.16
Classical MAI	5.58	3.88	2.87	4.91	3.35	2.53
BVAR (benchmark)	0.010	0.010	0.010	0.010	0.010	0.009
MAE						
Bayesian MAI	1.25	1.30	1.29	1.24	1.26	1.27
Classical MAI	5.04	3.61	2.77	4.49	3.17	2.48
BVAR (benchmark)	0.009	0.008	0.008	0.008	0.008	0.008

Empirical application

- Dataset of 20 U.S. macroeconomic variables
- Monthly data from January 1974 to December 2013 (first 7 years used as pre-sample)
- By searching over 455 specifications, we set the system rank to 3 and the lag length to 13
- We identify an output factor, a price factor, and a financial/monetary factor by imposing restrictions on the matrix B_0


F1, F2, F3
F1 and Variables in F1
Factor 1
EMPLOYEES NONFARM
PERSONAL INCOME
CONSUMPTION
IP: TOTAL
CAP UTIL
UNEMP. RATE
HSTARTS: TOTAL
AVG HOURLY EARNINGS
F2 and Variables in F2
Factor 2
CPI
SENS MAT'LS PRICE
PPI: FIN GDS
PCE DEFL
F3 and Variables in F3
Factor 3
FEDFUNDS
M1
M2
RESERVES TOT
RESERVES NONBOR
S&P 500
10 YR T-BOND
EX RATE

Responses to monetary policy shock

- The impulse responses are based on the representation:

$$Y_t = \{A(L)[I - B_0 A(L)]^{-1} B_0 + I\}\Lambda^{-1}\epsilon_t^*, \tag{26}$$

  where $\epsilon_t^* = \Lambda\epsilon_t$ are structural shocks and Λ^{-1} is the Cholesky factor of the variance of the reduced form shocks ϵ_t (Σ)
- We shock the element of ϵ_t^* corresponding to the Fed Funds rate
- We also compute the impulse responses using ML point estimates

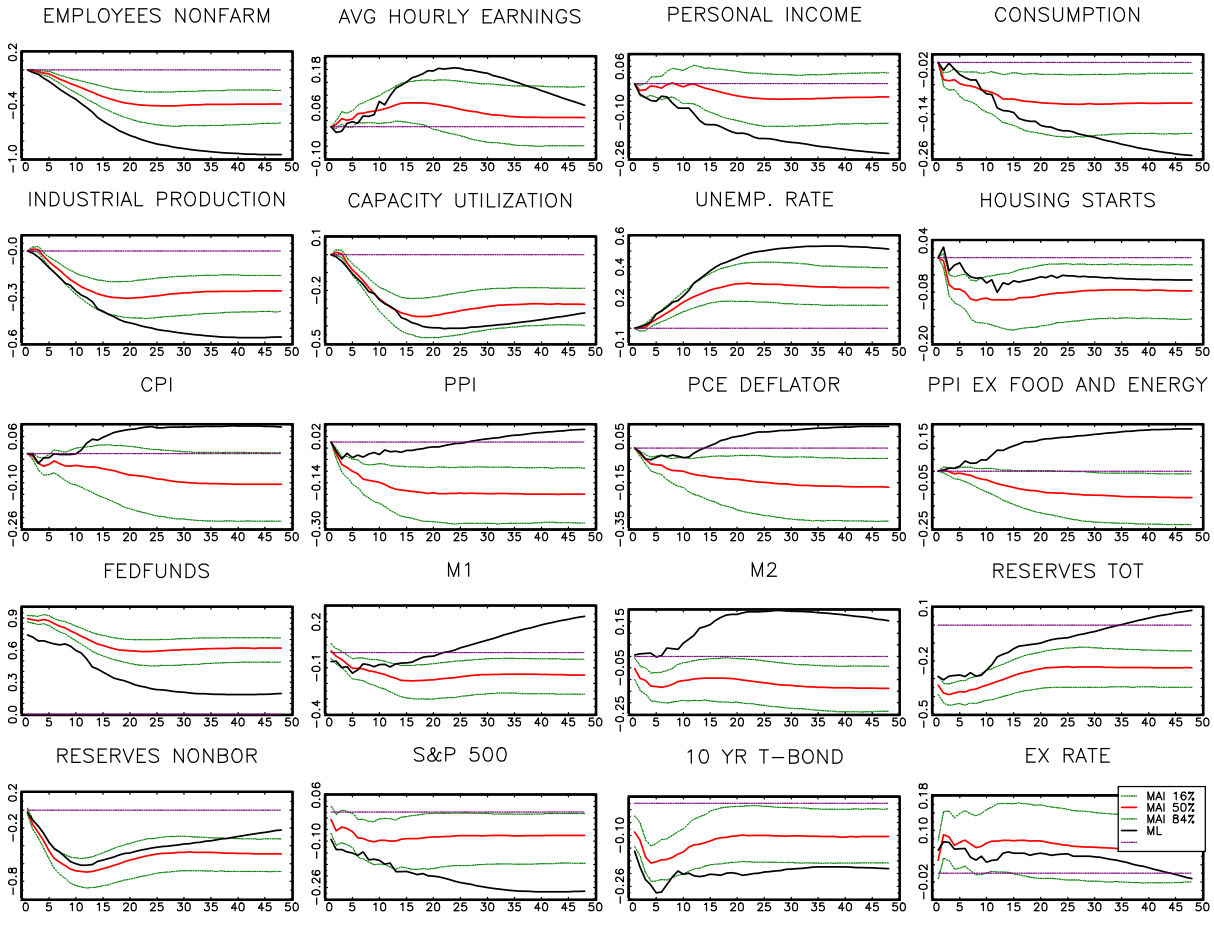


Figure 2: Baysian vs Classical MAI. Responses to a permanent shock to the Federal Funds rate. Red solid line and green dashed lines are the median and 16%-84% quantiles of the Bayesian MAI impulse responses. The solid black line represents the responses computed using maximum likelihood estimation.

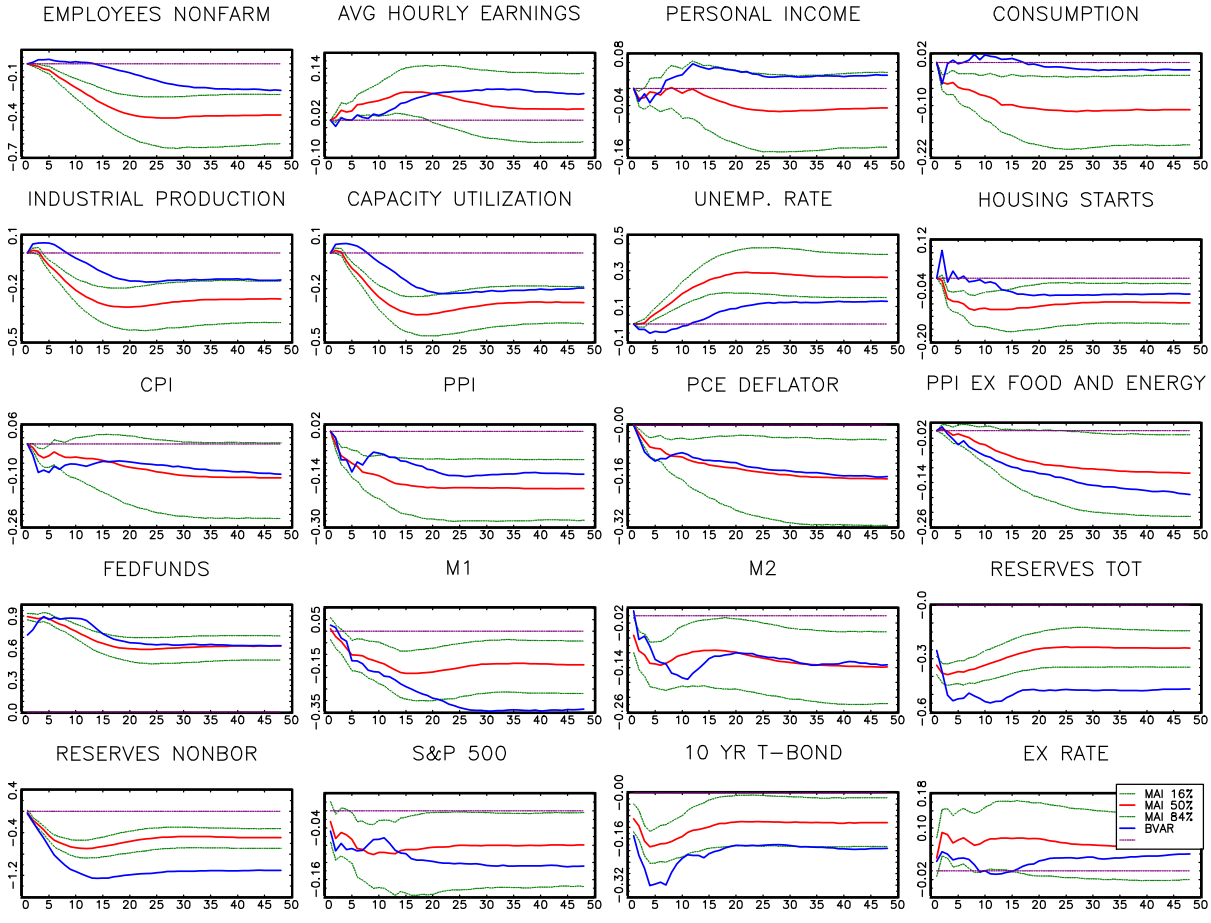


Figure 3: Bayesian MAI vs BVAR. Responses to a permanent shock to the Federal Funds rate. Red solid line and green dashed lines are the median and 16%-84% quantiles of the Bayesian MAI impulse responses. The solid blue line represents the responses computed using the unrestricted BVAR.

Shocks to factors

- The impulse responses are based on the representation:

$$Y_t = (\Sigma B_0' \Omega^{-1} + A(L)(I - B_0 A(L))^{-1}) P^{-1} v_t + B_{0\perp}' (B_{0\perp} \Sigma^{-1} B_{0\perp}')^{-1} \xi_t, \quad (27)$$

  where $v_t = P u_t$ are structural shocks and P^{-1} is the Cholesky factor of the variance of the reduced form shocks u_t (Ω)

- We shock the element of v_t corresponding to the real activity or prices factors

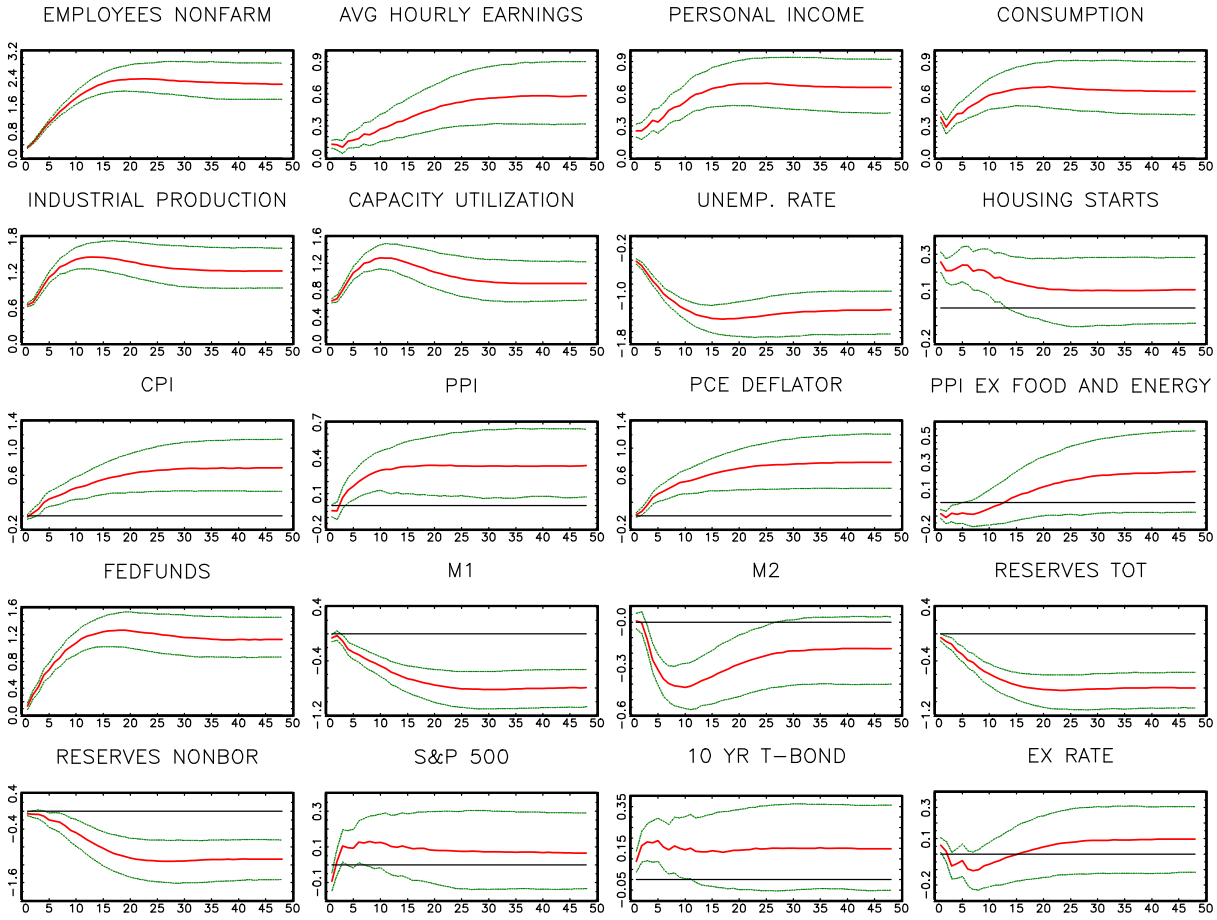


Figure 4: Demand Shock. Responses to a permanent shock to factor 1. Red solid line and green dashed lines are the median and 16%-84% quantiles of the Bayesian MAI impulse responses.

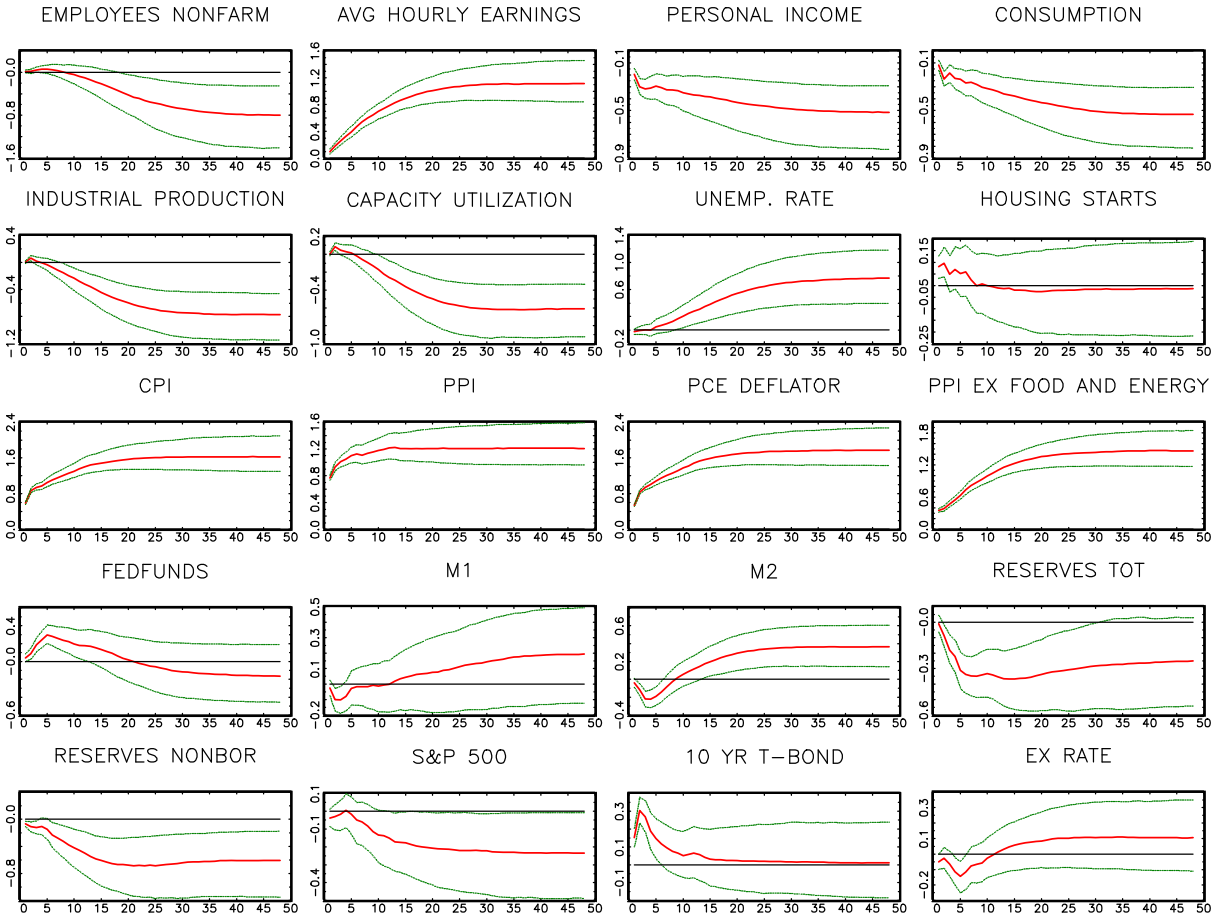


Figure 5: Supply shock. Responses to a permanent shock to factor 2. Red solid line and green dashed lines are the median and 16%-84% quantiles of the Bayesian MAI impulse responses.

Conclusions

- We have proposed a way to impose reduced rank reduction on a VAR which considerably helps in structural analysis
- We have discussed classical and Bayesian estimation and rank determination
- We have illustrated the model trough a MC
- We have implemented an empirical application on the effects of a demand, supply, and monetary policy shocks
- Overall the method looks general, simple, and flexible. Promising for empirical analyses with large datasets

Estimation via Maximum Likelihood - details

- Given A and $\tilde{B}_0$ the maximization with respect to Σ yields:

$$\hat{\Sigma} = \sum_{t=1}^{T} (Y_t - AZ_{t-1})(Y_t - AZ_{t-1})' / T \tag{28}$$

- The FOC with respect to A (given $\tilde{B}_0$ and Σ) is:

$$\frac{\partial l}{\partial vec(A')} = \sum_{t=1}^{T} (I_N \otimes Z'_{t-1}) \Sigma^{-1} \{ Y_t - (I_N \otimes Z'_{t-1}) vec(A') \} = 0 \tag{29}$$

- The FOC with respect to $\tilde{B}_0$ (given A and Ω) is:

$$\frac{\partial l}{\partial vec(\tilde{B}_0)} = \sum_{t=1}^{T} U_{t-1} A' \Sigma^{-1} \{ Y_t - (I_N \otimes Z'_{t-1}) vec(A') \} = 0 \tag{30}$$

  where $U_{t-1} = (I_r \otimes Y_{2,t-1}, ..., I_r \otimes Y_{2,t-p})$ and $Y'_{2,t}$ comes from partitioning Y'_t in the first r and last $N - r$ components: $Y'_t = (Y'_{1,t}, Y'_{2,t})$

- Reinsel (1983) shows that an iterating scheme solving in turn equations (28), (29) and (30) provides the ML estimates

Priors

- Assume a Normal-Inverse Wishart prior for A and Σ:

$$A'|\Sigma \sim N(A_0, \Sigma \otimes V_0),\ \Sigma \sim IW(S_0, v_0). \tag{31}$$

with:

$$A_0 = 0,\ V_0 = \tau D, \tag{32}$$

$$S_0 = S_{AR},\ v_0 = N + 2, \tag{33}$$

where S_{AR} is a diagonal matrix of residual sum of squares from univariate regressions on a pre-sample and where $\sqrt{\tau}$ is selected via maximization of the marginal data density

- The prior variance features a Kronecker structure with D reflecting a Minnesota-style prior
- We use a moderately informative prior on $\widetilde{B}_0$ based on an auxiliary model estimated on a pre-sample

Estimation via Markov Chain Monte Carlo - drawing A

- Under the knowledge of $\widetilde{B}_0$ and Y the variable Z_{t-1} is known, and (20) is a simple multivariate regression model as in Zellner (1973). Then the conditional posterior distributions are:

$$A'|\Sigma, \widetilde{B}_0, Y \sim N(\bar{A}, \Sigma \otimes V_1), \ \Sigma|\widetilde{B}_0, Y \sim IW(\bar{S}, \bar{v}). \tag{34}$$

with:

$$\begin{aligned} V_1 &= (V_0^{-1} + Z'Z)^{-1} \\ \bar{A} &= V_1(V_0^{-1}A_0 + Z'Y) \\ \bar{S} &= S_0 + Y'Y + A_0'V_0^{-1}A_0 - \bar{A}'V_1^{-1}\bar{A} \\ \bar{v} &= v_0 + T \end{aligned}$$

- Draws from $p(A', \Sigma|\widetilde{B}_0, Y)$ can be easily obtained by MC integration by generating a sequence of M draws from $\Sigma|\widetilde{B}_0, Y$ and then from $A'|\widetilde{B}_0, \Sigma, Y$

Estimation via Markov Chain Monte Carlo - drawing B

- Drawing from $p(\widetilde{B}_0|A, \Sigma, Y)$ is less simple, as $\widetilde{B}_0$ does not have a known conditional posterior. We use a sequence of RW Metropolis steps
- Let $\widetilde{B}_{0ji}$ denote the element in row j and column i in the matrix $\widetilde{B}_0$, and let $\widetilde{B}_{0ji^-}$ denote the set of all the remaining elements of $\widetilde{B}_0$
- At iteration m, a candidate $\widetilde{B}^*_{0ji}$ is drawn, conditional on A', Σ, and the remaining elements $\widetilde{B}_{0ji^-}$, using a random walk proposal:

$$\widetilde{B}^*_{0ji} = \widetilde{B}^{m-1}_{0ji} + c\eta_t, \tag{35}$$

  where η_t is a standard Gaussian i.i.d. process and c is a scaling factor calibrated in order to have a rejection rate of about 65%-70%.
- The candidate draw is accepted with probability

$$\alpha_k = \min\left\{1, \frac{p(\widetilde{B}^*_{0ji}|\widetilde{B}_{0ji^-}, A', \Sigma, Y)}{p(\widetilde{B}^{m-1}_{0ji}|\widetilde{B}_{0ji^-}, A', \Sigma, Y)}\right\}. \tag{36}$$

General reduced rank VAR

- Assume $\Phi(L) = A(L)B(L)$, where $A(L) = A_1 L + ... + A_{p_1} L^{p_1}$, each A_i is $N \times r$, $B(L) = B_0 + B_1 L + ... + B_{p_2} L^{p_2}$ and each B_i is $r \times N$, with $p_1 + p_2 = p$, $p_1 \geq 1$, $p_2 \geq 0$. Then

$$Y_t = A(L)B(L)Y_t + \epsilon_t = \sum_{u=1}^{p_1} \sum_{v=0}^{p_2} A_u B_v Y_{t-u-v} + \epsilon_t \tag{37}$$

- Here we set $p_1 = p$ and $p_2 = 0$ which gives:

$$Y_t = \sum_{u=1}^{p} A_u B_0 Y_{t-u} + \epsilon_t \tag{38}$$

- Geweke (1996) sets $p_1 = 1$ and $p_2 = p - p_1$ which gives:

$$Y_t = \sum_{v=0}^{p-1} A_1 B_v Y_{t-1-v} + \epsilon_t \tag{39}$$

- If $p = 1$ then the two models coincide

Comparison with Geweke (1996)

- Define $X_t = (Y'_{t-1}, ..., Y'_{t-p})'$, of dimension $np \times 1$. Geweke (1996) model:

$$\underset{n\times 1}{Y_t} - \epsilon_t = \underset{n\times r}{A}\underset{r\times 1}{Z_{t-1}} = \underset{n\times r}{A}\underset{r\times np}{B}\underset{np\times 1}{X_t} = \underset{n\times r}{A_1}\underset{r\times np}{[B_0|\ ...|\ B_{p-1}\ ]}X_t. \tag{40}$$

  which is a multivariate reduced rank regression model

- This model:

$$\underset{n\times 1}{Y_t} - \epsilon_t = \underset{n\times rp}{A}\underset{rp\times 1}{Z_{t-1}} = \underset{n\times rp}{A}\underset{rp\times np}{B}\underset{np\times 1}{X_t} = \underset{n\times rp}{[A_1|...|A_p]}\underset{rp\times np}{(I_p \otimes B'_0)'}X_t. \tag{41}$$

- Geweke's derivation of the conditional posterior of $B_0, ..., B_{p-1}$ hinges on the use of the (left) generalised inverse of the matrix A_1. The generalised inverse can be defined in this case as A_1 has full column rank r which gives $A^+ = (A'_1 A_1)^{-1} A'_1$
- Here the matrix A in (41) is of dimension $n \times rp$ with (at most) rank n, so $A'A$ is singular and the left generalised inverse is not defined
- Note Geweke (1996) does not allow to get a VAR for the factors via pre-multiplication by B

Marginal data density

- The density $p_r(Y)$ can be efficiently approximated numerically by using Rao-Blackwellization and the harmonic mean estimator proposed by Gelfand and Dey (1994), as suggested in Fuentes-Albero and Melosi (2013).
- In particular, given M simulated posterior draws $\{\widetilde{B}_0\}_{m=1}^{M}$, we have:

$$\hat{p}_r(Y) = \left[\frac{1}{M} \sum_{m=1}^{M} \frac{1}{p(Y|\widetilde{B}_0^m)p(\widetilde{B}_0^m)} f(\widetilde{B}_0^m) \right]^{-1}, \tag{42}$$

where $f(\cdot)$ is a truncated multivariate normal distribution calibrated using the moments of the simulated posterior draws (see Geweke 1999) and $p(\widetilde{B}_0^m)$ is the prior distribution of $\tilde{B}_0$ evaluated at the posterior draw $\widetilde{B}_0^m$.

- The term $p(Y|\widetilde{B}_0^m)$ is the integrating constant of the conditional posterior distribution $p(A, \Sigma|Y, \widetilde{B}_0)$. Since conditionally on $\widetilde{B}_0^m$ the model is a multivariate regression with a naturally conjugate prior, $p(Y|\widetilde{B}_0^m)$ is available in closed form.

Convergence and mixing

40000 draws obtained with 2 parallel chains of 25000 draws each, removing 5000 for burn-in.

Table 7: Inefficiency Factors and Potential Scale Reduction Factors

	B		A		A-B	
	IF	PSRF	IF	PSRF	IF	PSRF
mean	7.614	1.000	0.898	1.000	1.128	1.000
median	5.649	1.000	0.826	1.000	0.918	1.000
10% quan:	2.810	1.000	0.460	1.000	0.491	1.000
90% quan:	11.235	1.001	1.436	1.000	1.851	1.000
min	2.336	1.000	0.177	1.000	0.151	1.000
max	31.694	1.001	2.334	1.000	13.397	1.000