# Large Vector Autoregressions with Stochastic Volatility and Flexible Priors

Andrea Carriero[1], Todd Clark[2], and Massimiliano Marcellino[3]

[1]Queen Mary, University of London

[2]Federal Reserve Bank of Cleveland

[3]Bocconi University and CEPR

June 2016

# Introduction

**Two VAR features helpful for forecasting and structural analysis:**

- Large variable set
  - Banbura, Giannone, and Reichlin (2010), Carriero, Clark, and Marcellino (2015), Giannone, Lenza, and Primiceri (2015) and Koop (2013)
- Time variation in volatility
  - Clark (2011), Clark and Ravazzolo (2015), Cogley and Sargent (2005), D'Agostino, Gambetti and Giannone (2013), and Primiceri (2005)

# Introduction

**Few papers provide approaches for accommodating both features. Recent exceptions:**

- Koop and Korobilis (2013), Koop, et al. (2016): computational shortcut using exponential smoothing of volatility
- Carriero, Clark, and Marcellino (2016): single volatility factor and specific prior that permits use of N-W steps

# Introduction

**Allowing large VARs with homoskedasticity requires symmetry of likelihood and prior.**

- Homoskedastic VARs: SUR models w/ the same regressors in each equation
- Symmetry across equations → likelihood has a Kronecker structure → OLS estimation equation by equation
- With homoskedasticity, large BVARs require a specific prior structure, of conjugate N-W:
  - The coefficients of each equation feature the same prior variance matrix (up to a constant of proportionality).
  - Priors are correlated across equations, with a correlation structure proportional to $\Sigma$.

# Introduction

## More general priors break symmetry and make large models computationally difficult.

- Priors more general than conjugate N-W break the Kronecker structure and symmetry.
  - Examples: prior with Litterman-style cross-variable shrinkage or Normal-diffuse prior
- Model needs to be vectorized for estimation
- Drawing the VAR coefficients from the conditional posterior involves a variance matrix of dimension $N^2 \times$ lags.

## SV also breaks symmetry and makes large models difficult

- Each equation driven by a different volatility → Model needs to be vectorized
- Drawing the VAR coefficients involves a variance matrix of dimension $N^2 \times$ lags.

# Introduction

**We develop a new estimation approach that makes tractable large models with asymmetric priors or SV**

- Algorithm exploits a simple triangularization of the VAR, which permits drawing VAR coefficients equation by equation
- This reduces the computational complexity for estimating the VAR model from $N^6$ to $N^4$, greatly speeding up estimation.
- The triangularization can easily be inserted in any pre-existing algorithm for estimation of BVARs.
  - Example code to be available on Carriero and Marcellino webpages
- Estimation of large VARs with SV and flexible priors becomes feasible.

# Introduction

### Application 1: Structural analysis of BVAR-SV in 125 monthly variables

- SV estimates: heterogeneity and yet much commonality
- Impulse responses for a policy shock

### Application 2: Out-of-sample forecasts from BVAR-SV in 20 monthly variables

- Larger model forecasts better than smaller model
- SV improves accuracy of both density and point forecasts

# Outline

1. BVAR-SV specification and impediments to large models
2. Our estimation method for large BVARs
3. Application 1: Structural analysis with large BVAR-SV
4. Application 2: Out-of-sample forecasting
5. Conclusions

# BVAR-SV Model

## With $N$ variables:

$$
\begin{aligned}
y_t &= \Pi_0 + \Pi(L) y_{t-1} + v_t \\
v_t &= A^{-1} \Lambda_t^{0.5} \epsilon_t, \ \ \epsilon_t \sim iid \ \ N(0, I_N); \ \ \text{var}(v_t) \equiv \Sigma_t = A^{-1} \Lambda_t A^{-1} \\
\ln \lambda_{j,t} &= \ln \lambda_{j,t-1} + e_{j,t}, \ \ j = 1, \ldots, N \\
e_t &\sim iid \ \ N(0, \Phi)
\end{aligned}
$$

- Let $X_t$ denote the $(Np + 1)$-dimensional vector of regressors in each equation

## Collect parameter blocks and latent states:

- Parameters: $\Theta = \{\Pi, A, \Phi\}$
- Latent states $\ln \lambda_{j,t}$, $t = 1, \ldots, T$, $j = 1, \ldots, N$

# BVAR-SV Model: standard system estimation

Priors:

$$\begin{aligned} \text{vec}(\Pi) &\sim N(\text{vec}(\underline{\mu}_{\Pi}), \underline{\Omega}_{\Pi}) \\ A &\sim N(\underline{\mu}_{A}, \underline{\Omega}_{A}) \\ \Phi &\sim IW(\underline{d}_{\Phi} \cdot \underline{\Phi}, \underline{d}_{\Phi}) \\ \ln \lambda_{j,0} &\sim \text{uninformative Gaussian} \end{aligned}$$

# BVAR-SV Model: standard system estimation

## Posteriors:

$$\begin{aligned} \text{vec}(\Pi)|A, \Lambda_T, y_T &\sim N(\text{vec}(\bar{\mu}_\Pi), \overline{\Omega}_\Pi) \\ A|\Pi, \Lambda_T, y_T &\sim N(\bar{\mu}_A, \overline{\Omega}_A) \\ \Phi|\Lambda_T, y_T &\sim IW((\underline{d}_\Phi + T) \cdot \bar{\Phi}, \underline{d}_\Phi + T), \end{aligned}$$

- Means and variances of conditional normal distributions take GLS-based form, combining prior moments and likelihood moments

## Gibbs sampler for $p(\Theta, \Lambda_T|y_T)$:

- Draw from $p(\Theta|\Lambda_T, y_T)$ using conditional posteriors above
- Draw from $p(\Lambda_T|\Theta, y_T)$ using the mixture of normals approximation and multi-move algorithm of Kim, Shepard and Chib (1998)

# BVAR-SV Model: impediments to standard system estimation with a large model

- Sampling the VAR coefficients involves drawing a $N(Np+1)-$dimensional vector `rand`, and computing

$$\text{vec}(\Pi^m) = \bar{\Omega}_\Pi \left\{ \text{vec}\left( \sum_{t=1}^{T} X_t y_t' \Sigma_t^{-1} \right) + \underline{\Omega}_\Pi^{-1} \text{vec}(\underline{\mu}_\Pi) \right\} + chol(\bar{\Omega}_\Pi) \times \texttt{rand} \quad (1)$$

- This calculation requires: i) computing $\bar{\Omega}_\Pi$ by inverting

$$\bar{\Omega}_\Pi^{-1} = \underline{\Omega}_\Pi^{-1} + \sum_{t=1}^{T} (\Sigma_t^{-1} \otimes X_t X_t');$$

  ii) computing its Cholesky factor $chol(\bar{\Omega}_\Pi)$; iii) multiplying the matrices obtained in i) and ii) by the vector in the curly brackets of (1) and the vector `rand`, respectively.
- Each operation requires $O(N^6)$ elementary operations, making the total computational complexity to draw $\Pi^m$ equal $4 \times O(N^6)$.

# Homoskedastic BVARs: similar impediments with flexible priors

**Model:**

$$y_t = \Pi_0 + \Pi(L) y_{t-1} + v_t, \;\; v_t \sim iid \;\; N(0, \Sigma)$$

**Consider a general N-W prior:**

$$\text{vec}(\Pi) \sim N(\text{vec}(\underline{\mu}_\Pi), \underline{\Omega}_\Pi); \;\; \Sigma \sim IW(\underline{d}_\Sigma \cdot \underline{\Sigma}, \underline{d}_\Sigma)$$

**Posterior:**

$$\text{vec}(\Pi)|\Sigma, y \;\; \sim \;\; N(\text{vec}(\bar{\mu}_\Pi), \overline{\Omega}_\Pi); \;\; \Sigma|\Pi, y \sim IW((\underline{d}_\Sigma + T) \cdot \bar{\Sigma}, \underline{d}_\Sigma + T)$$

$$\bar{\Omega}_\Pi^{-1} \;\; = \;\; \underline{\Omega}_\Pi^{-1} + \sum_{t=1}^{T} (\Sigma^{-1} \otimes X_t X_t')$$

**Impediment to large models: Computational requirements with system variance $\bar{\Omega}_\Pi$ that also exist with SV formulation**

# Homoskedastic BVARs: standard approach to making large models tractable

Following literature on large VARs, make the prior conjugate (and symmetric) N-W.

$$\text{vec}(\Pi)|\Sigma \sim N(\text{vec}(\underline{\mu}_\Pi), \Sigma \otimes \Omega_0)$$

- Prior for $\Pi$ is conditional on $\Sigma$

Posterior variance simplifies and speeds up calculations:

$$\bar{\Omega}_\Pi^{-1} = \Sigma^{-1} \otimes \left\{ \Omega_0^{-1} + \sum_{t=1}^{T} X_t X_t' \right\}$$

- Kronecker structure permits manipulating the two terms in the Kronecker product separately, reducing the computational complexity to $N^3$

# Homoskedastic BVARs: standard approach to making large models tractable

**The conjugate (and symmetric) N-W form comes with some unappealing restrictions.**

- Issues discussed by Rothenberg (1963), Zellner (1973), Kadiyala and Karlsson (1993, 1997), and Sims and Zha (1998)
- Rules out asymmetry in the prior across equations; coefficients of each equation feature the same prior variance matrix $\Omega_0$
- Rules out one aspect of the Litterman (1986) prior: extra shrinkage on "other" lags vs. "own" lags
- $\Sigma \otimes \Omega_0$ implies prior beliefs correlated across the equations of the reduced form VAR
  - Sims and Zha (1998) specify a prior featuring independence among the *structural* equations, but does not achieve computational gains for an asymmetric prior on the *reduced form*.

# Our estimation method for large VARs

Key to approach: In the Gibbs sampler, the posterior of the VAR coefficients $\Pi$ is conditional on $A$ and $\Lambda_T$.

- $\pi^{(i)}$ = the vector of coefficients for equation $i$ contained in row $i$ of $\Pi$, for the intercept and coefficients on lagged $y_t$
- Consider the decomposition $v_t = A^{-1}\Lambda_t^{0.5}\epsilon_t$:

$$\begin{bmatrix} v_{1,t} \\ v_{2,t} \\ \dots \\ v_{N,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ a^*_{2,1} & 1 & & \dots \\ \dots & & 1 & 0 \\ a^*_{N,1} & \dots & a^*_{N,N-1} & 1 \end{bmatrix} \begin{bmatrix} \lambda_{1,t}^{0.5} & 0 & \dots & 0 \\ 0 & \lambda_{2,t}^{0.5} & & \dots \\ \dots & & \dots & 0 \\ 0 & \dots & 0 & \lambda_{N,t}^{0.5} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \dots \\ \epsilon_{N,t} \end{bmatrix}$$

# Our estimation method for large VARs

## Rewrite the VAR:

$$y_{1,t} = \pi_1^{(0)} + \sum_{i=1}^{N}\sum_{l=1}^{p} \pi_{1,l}^{(i)} y_{i,t-l} + \lambda_{1,t}^{0.5}\epsilon_{1,t}$$

$$y_{2,t} = \pi_2^{(0)} + \sum_{i=1}^{N}\sum_{l=1}^{p} \pi_{2,l}^{(i)} y_{i,t-l} + a_{2,1}^{*}\lambda_{1,t}^{0.5}\epsilon_{1,t} + \lambda_{2,t}^{0.5}\epsilon_{2,t}$$

$$\ldots$$

with the generic equation $(*)$ for variable $j$:

$$y_{j,t} - (a_{j,1}^{*}\lambda_{1,t}^{0.5}\epsilon_{1,t} + \ldots + a_{j,,j-1}^{*}\lambda_{j-1,t}^{0.5}\epsilon_{j-1,t}) = \pi_j^{(0)} + \sum_{i=1}^{N}\sum_{l=1}^{p} \pi_{j,l}^{(i)} y_{i,t-l} + \lambda_{j,t}\epsilon_{j,t}$$

## Consider estimating these equations in order from $j = 1$ to $j = N$

- In the conditional posterior, the dependent variable of $(*)$ is known.
- Dependent variable$_j$ = $y_j - a\times$ the estimated residuals of all the previous $j-1$ equations.

# Our estimation method for large VARs

- Let $y^*_{j,t} \equiv y_{j,t} - (a^*_{j,1}\lambda^{0.5}_{1,t}\epsilon_{1,t} + ... + a^*_{j,,j-1}\lambda^{0.5}_{j-1,t}\epsilon_{j-1,t})$
- The model is a system of standard generalized linear regression models with indep. Gaussian disturbances with mean 0 and variance $\lambda_{j,t}$:

$$y^*_{j,t} = \pi^{(0)}_j + \sum_{i=1}^{N}\sum_{l=1}^{p}\pi^{(i)}_{j,l}y_{i,t-l} + \lambda_{j,t}\epsilon_{j,t},$$

**Factorize the full conditional posterior distribution of $\Pi$:**

$$\begin{aligned} p(\Pi|A,\Lambda_T,y) \quad = \quad & p(\pi^{(N)}|\pi^{(N-1)},\pi^{(N-2)},\ldots,\pi^{(1)},A,\Lambda_T,y) \\ & \times p(\pi^{(N-1)}|\pi^{(N-2)},\ldots,\pi^{(1)},A,\Lambda_T,y) \\ & \vdots \\ & \times p(\pi^{(1)}|A,\Lambda_T,y), \end{aligned}$$

# Our estimation method for large VARs

**Our conditional posterior for the VAR coefficients:**

$$p(\Pi^{\{j\}}|\Pi^{\{1:j-1\}}, A, \Lambda_T, y) \propto p(y|\Pi^{\{j\}}, \Pi^{\{1:j-1\}}, A, \Lambda_T) p(\Pi^{\{j\}}|\Pi^{\{1:j-1\}})$$

- $p(y|\Pi^{\{j\}}, \Pi^{\{1:j-1\}}, A, \Lambda_T)$ = the likelihood of equation $j$
- $p(\Pi^{\{j\}}|\Pi^{\{1:j-1\}})$ = prior on the $j$-th equation, conditional on the previous equations
- With typical priors, the equation priors are independent: $p(\Pi^{\{j\}}|\Pi^{\{1:j-1\}}) = p(\Pi^{\{j\}})$
- W/o independence, the moments of $p(\Pi^{\{j\}}|\Pi^{\{1:j-1\}})$ can be obtained from the joint prior.

# Our estimation method for large VARs

## Our conditional posterior for the VAR coefficients:

- Draw the coefficient matrix $\Pi$ in separate blocks $\Pi^{\{j\}}$ obtained from:

$$\Pi^{\{j\}}|\Pi^{\{1:j-1\}}, A, \Lambda_T, y \sim N(\bar{\mu}_{\Pi^{\{j\}}}, \overline{\Omega}_{\Pi^{\{j\}}})$$

$$\begin{aligned} \bar{\mu}_{\Pi^{\{j\}}} &= \overline{\Omega}_{\Pi^{\{j\}}} \left\{ \underline{\Omega}^{-1}_{\Pi^{\{j\}}} \underline{\mu}_{\Pi^{\{j\}}} + \sum_{t=1}^{T} X_{j,t} \lambda^{-1}_{j,t} y^{*\prime}_{j,t} \right\} \\ \overline{\Omega}^{-1}_{\Pi^{\{j\}}} &= \underline{\Omega}^{-1}_{\Pi^{\{j\}}} + \sum_{t=1}^{T} X_{j,t} \lambda^{-1}_{j,t} X'_{j,t}, \end{aligned}$$

  where $\underline{\Omega}^{-1}_{\Pi^{\{j\}}}$ and $\underline{\mu}_{\Pi^{\{j\}}}$ = the prior moments on the $j$-th equation, given by the $j$-th column of $\underline{\mu}_{\Pi}$ and the $j$-th block on the diagonal of $\overline{\Omega}^{-1}_{\Pi}$

- Here $\underline{\Omega}^{-1}_{\Pi}$ is block diagonal, as typical; this can be relaxed

# Our estimation method for large VARs

## Computational costs (not much):

- Although we break the conditional posterior for $\Pi$ into pieces, we are still drawing from the conditional posterior for $\Pi$.
- Our triangularization approach produces draws numerically identical to those that would be obtained using system-wide estimation.
- For the VAR coefficients, the ordering of variables does not matter.
- Existing BVAR and BVAR-SV code can easily be modified to draw $\Pi$ with the triangularized system.

# Our estimation method for large VARs

## Computational benefits (significant):

- $\overline{\Omega}^{-1}_{\Pi^{\{j\}}}$ is of dimension $(Np+1)$ square $\rightarrow$ its manipulation only involves operations of order $O(N^3)$
- With $N$ equations, obtaining a draw for $\Pi$ makes the total computational complexity of order $O(N^4)$
- Compared to a standard algorithm, the complexity savings is $N^2$
- CPU savings rise quickly (more than quadratic rate) with the number of variables.
- With 20 variables and 13 lags of monthly data, the estimation of the model using the traditional system-wide algorithm was about **261** times slower.

# Our estimation method for large VARs

## Convergence and mixing

- **In a given unit of time**, our triangular algorithm will always produce many more draws than the traditional system-wide algorithm.
- This speed advantage will improve the precision of MCMC estimates:
  - Many more draws to use in averages
  - Or increased skip-sampling (preferable with large models) to reduce correlation across retained draws

# Application 1: large structural VAR with SV

**Specification: BVAR-SV(13) in 125 monthly variables from the dataset of McCracken and Ng (2015)**

- Extending constant volatility analyses of (FAVAR) Bernanke, Boivin and Eliasz (2005) and (large BVAR) Banbura, Giannone, and Reichlin (2010)
- VAR coefficient prior (asymmetric): independent Normal-Wishart prior, Minnesota form, with cross-variable shrinkage

**Assessments:**

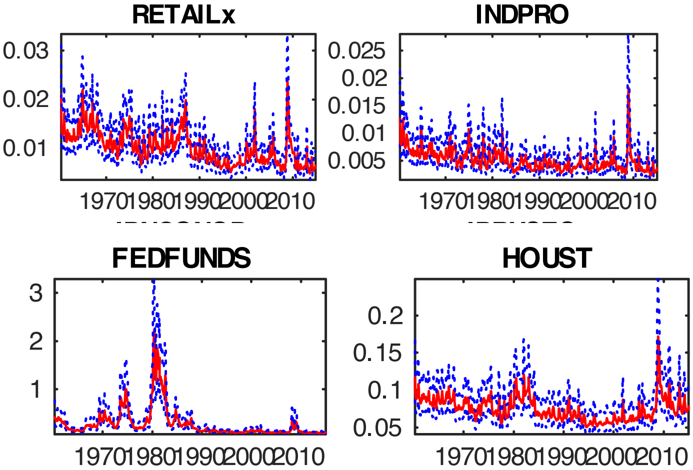
- Estimates of volatilities and comovement
- Responses to monetary policy shock
  - For identification, the federal funds rate is ordered after slow-moving and before fast-moving variables.

# Application 1: large structural VAR with SV

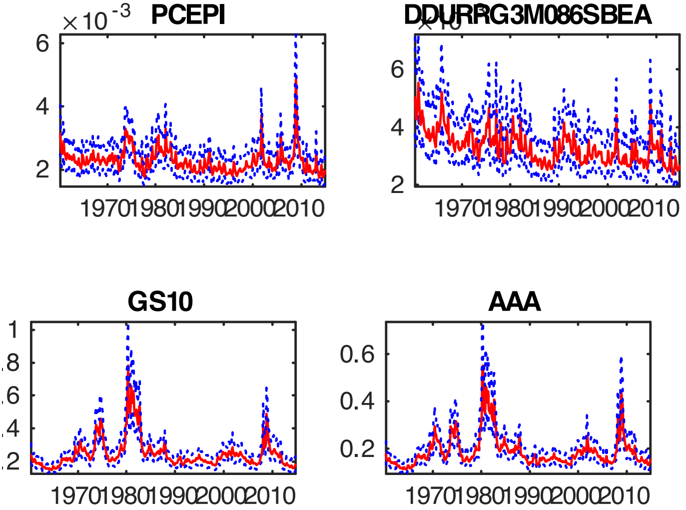
## Computation:

- Model includes 203,250 VAR coefficients
- On a 3.5 GHz Intel Core i7 processor, our algorithm produces 5000 draws (after discarding 500 burning in) in just above 7 hours
- The traditional system-based algorithm would be extremely difficult, just for memory requirements: the covariance matrix of the 203,250 coefficients would require about 330 GB of RAM

# Application 1: large structural VAR with SV



# Application 1: large structural VAR with SV

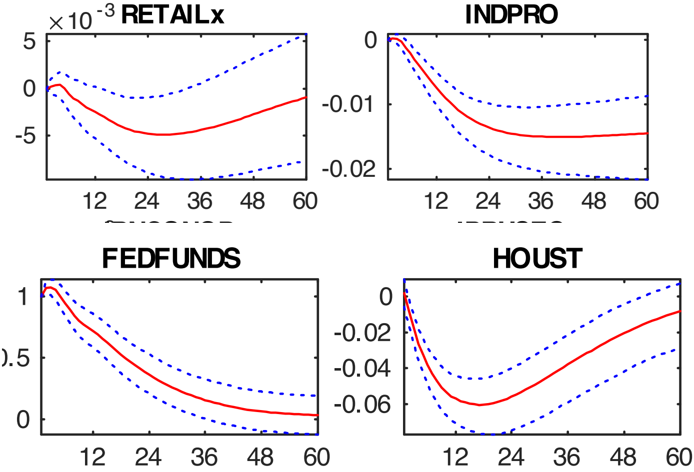


# Application 1: large structural VAR with SV

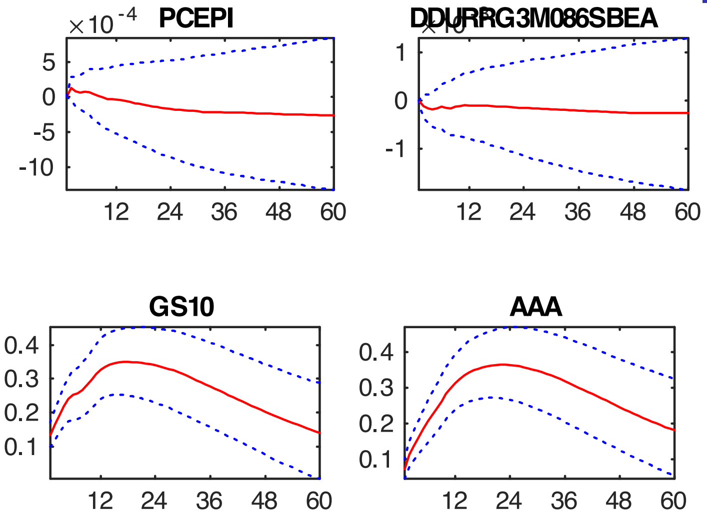
## Results on volatilities:

- Substantial homogeneity in the volatility patterns of variables belonging to the same group, such as IP components
- Heterogeneity across groups of variables
- Principal component analysis on the posterior mean of $\Phi$ indicates macroeconomic volatility is primarily driven by two shocks
- The Great Moderation is evident in most series; the effects of the recent crisis are more heterogeneous.
- Volatilities of real variables and financial variables go back to lower levels after the peak associated with the crisis.
- Volatilities of inflation measures have tended to remain elevated following the crisis.

# Application 1: large structural VAR with SV



# Application 1: large structural VAR with SV



# Application 1: large structural VAR with SV

## Results on impulse responses to FFR shock:

- The patterns of impulse responses align with typical structural models: significant deterioration in real activity, very limited price puzzle, a significant deterioration in stock prices, and a less than proportional increase in the entire term structure
- Inclusion of SV does not affect substantially the VAR coefficient estimates with respect to Banbura, Giannone and Reichlin (2010)
- But it matters for inference and time variation in variance contributions and shares

# Application 2: forecasts of 20 monthly variables

| Variables in baseline specification | |
|---|---|
| Real Personal Income | PPI: Commodities |
| Real PCE | PCE Price Index |
| Real M&T Sales | Federal Funds Rate |
| IP Index | Housing Starts |
| Capacity Utilization: Manufacturing | S&P 500 |
| Unemployment Rate | U.S.-U.K. exchange rate |
| All Employees: Total nonfarm | Spread, 1y Treasury-Fed funds |
| Hours: Manufacturing | Spread, 10y Treasury-Fed funds |
| Avg. Hourly Earnings: Goods | Spread, Baa-Fed funds |
| PPI: Finished Goods | ISM: New Orders Index |

**Samples:**

- Estimation sample begins with 1960:3
- Forecast evaluation sample is 1970:3 to 2014:5.

# Application 2: forecasts of 20 monthly variables

Four models:

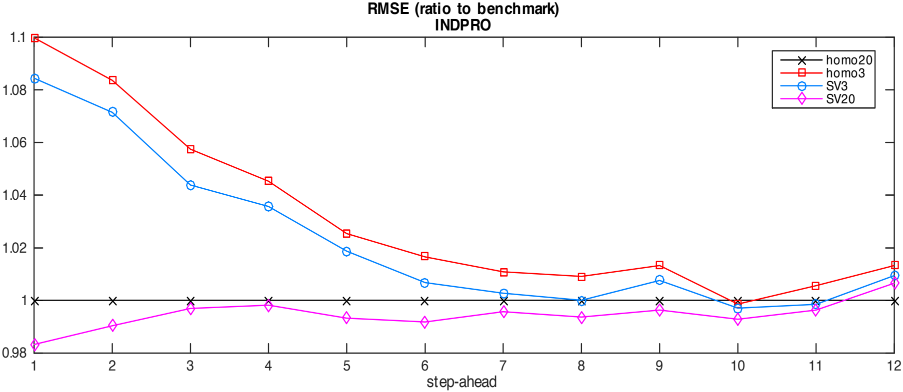
- 3-variable BVAR, homoskedastic: growth rate of IP ($\Delta \ln IP$), PCE inflation ($\Delta \ln PECEPI$), fed funds rate (FFR)
- 3-variable BVAR-SV
- 20-variable BVAR, homoskedastic
- 20-variable BVAR-SV

# Application 2: forecasts of 20 monthly variables

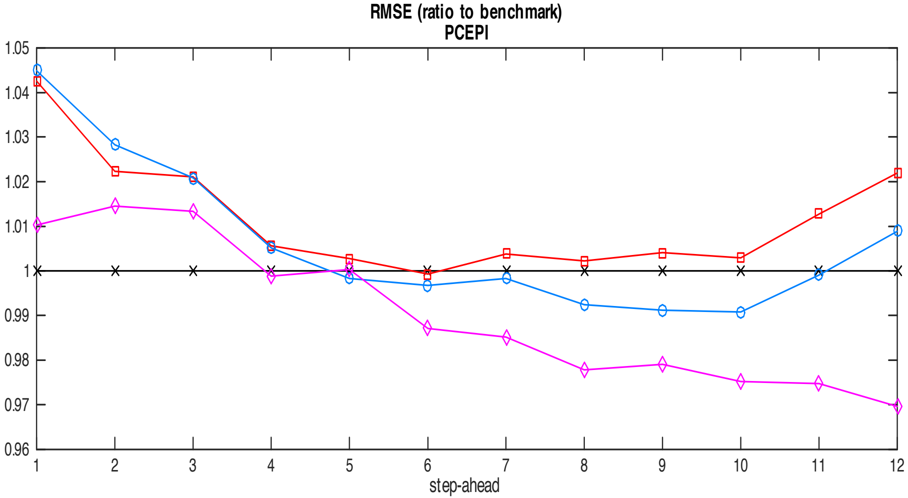
## Drivers of forecast gains:

- Direct effects:
    - SV improves density forecasts by capturing time variation in error variances.
    - Use of a larger dataset should improve point forecasts by improving the conditional means.
- Interactions:
    - A better point forecast improves the density forecast by better centering the predictive density.
    - SV improves the point forecasts by making parameter estimates more efficient (GLS).
    - This efficiency also helps the predictive densities.

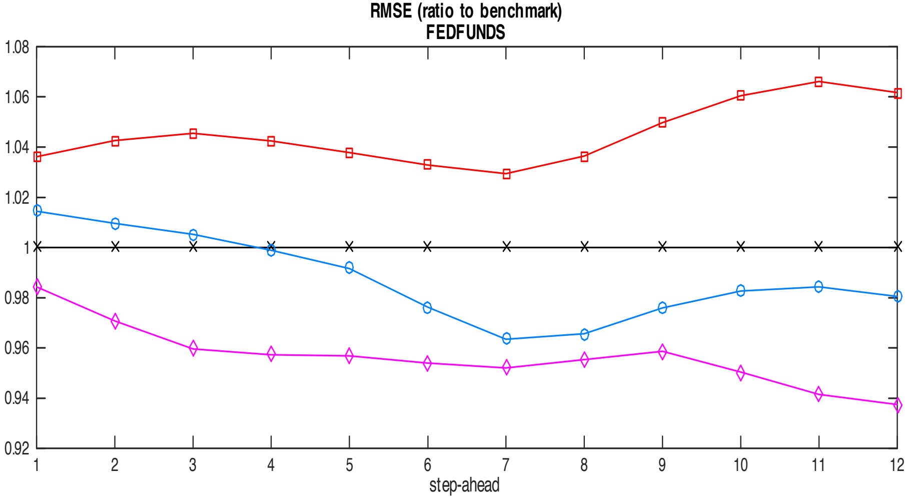
# Application 2: forecasts of 20 monthly variables



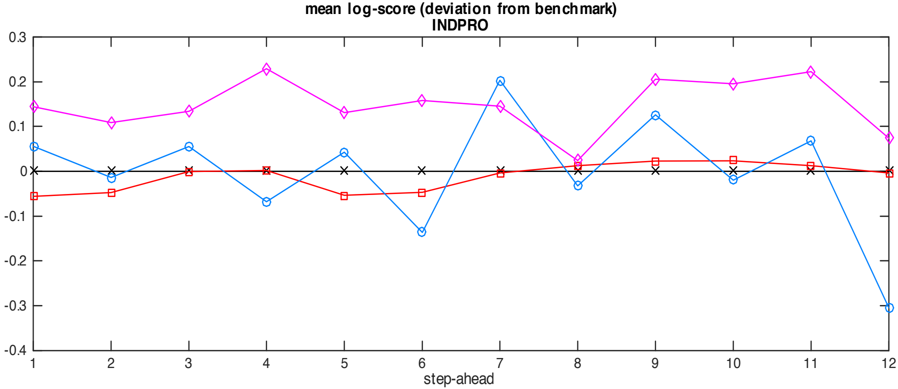
# Application 2: forecasts of 20 monthly variables



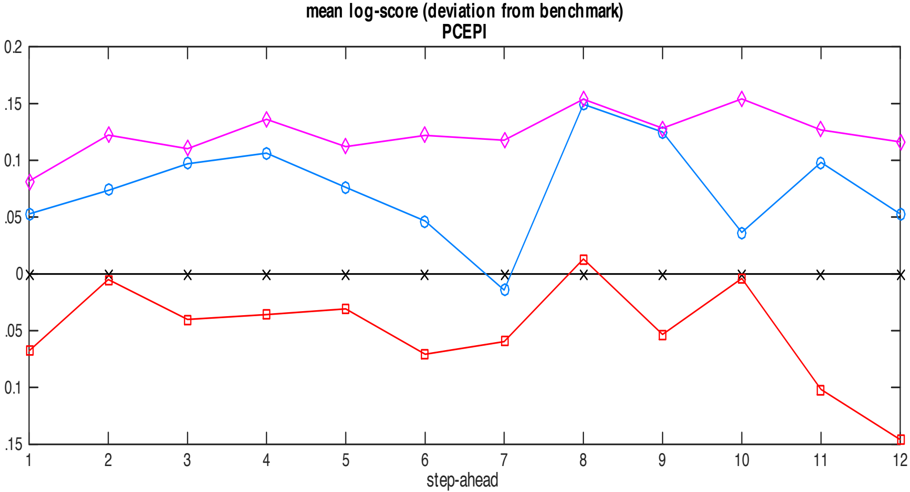
# Application 2: forecasts of 20 monthly variables



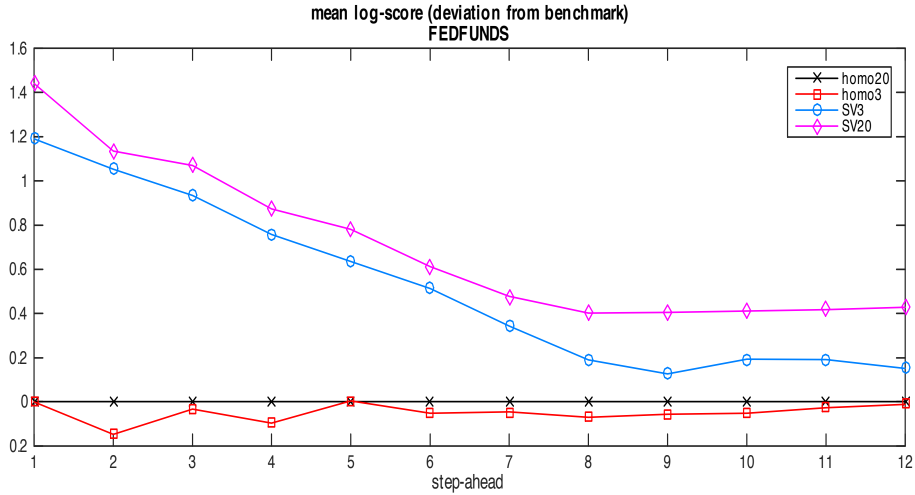
# Application 2: forecasts of 20 monthly variables



# Application 2: forecasts of 20 monthly variables



# Application 2: forecasts of 20 monthly variables



# Conclusions

- We develop a new approach that makes feasible fully Bayesian inference of large BVARs with SV.
  - Also makes feasible the use of asymmetric priors (independent N-W priors) with SV or constant volatility, in large models
- The method is based on a straightforward triangularization of the system, and it is very simple to implement by modifying existing code for drawing VAR coefficients.
- The algorithm ensures computational gains of order $N^2$ and yields better mixing and convergence properties.