# Appendix to "Corrigendum to: Large Bayesian Vector Autoregressions with Stochastic Volatility and Non-Conjugate Priors"

Andrea Carriero* Joshua Chan† Todd E. Clark‡ Massimiliano Marcellino§

This draft: October 2021

**Abstract**

This appendix provides additional application results, including figures from the published paper updated to use the corrected algorithm and other results mentioned in the Corrigendum. It also provides a corrected algorithm that can be used for triangular estimation of homoskedastic VARs.

## 1 Updates of results in the paper

The application results provided in this appendix are the same as in the original paper. They are based on monthly data taken from the FRED-MD data set, for the period January 1960 to December 2014. The model includes the 20 variables indicated in Table 1 and 13 lags.

Figures 1 through 5 provide updates of the results in the paper, using the corrected algorithm. As these figures indicate, correcting the algorithm for estimating the VAR's coefficients does not change the conclusions indicated in the paper, also provided below. The results in Figure 1 and Figure 2 are obtained by (1) running the SWA for a total of 22000 draws, discarding the first 2000 and retaining the remaining 20000 draws; and (2) running the CTA for the same amount of clock time as the SWA, discarding the first 2000 draws, and then adjusting the skip-sampling of the CTA to reduce the sample to 20000 retained draws. The results in Figure 3, Figure 4, and Figure 5 are based on a recursive pseudo-out-of-sample forecasting exercise. In each estimation sample of the exercise, the

*Queen Mary University of London, a.carriero@qmul.ac.uk

†Purdue University, joshuacc.chan@gmail.com

‡Federal Reserve Bank of Cleveland, todd.clark@clev.frb.org

§Bocconi University, IGIER and CEPR, massimiliano.marcellino@unibocconi.it

results are obtained by running the CTA for a total of 5500 draws, discarding the first 500 and retaining 1 in 5 of the remaining 5000 draws, providing a final sample of 1000 draws.

Updating the paper's Figure 2, Figure 1 illustrates the recursive means for some selected coefficients and shows that the corrected triangular algorithm with split-sampling reaches convergence much faster than the system-wide algorithm. This pattern is particularly marked for the volatility component of the model. Figure 2 illustrates the entire posterior distribution of some coefficients, and specifically those on the main diagonal of the matrix $\Pi_1$. As these estimates indicate, the corrected triangular algorithm yields posterior densities that match those obtained with the system-wide algorithm.

**Table 1: Variables in the 20-variable forecasting models**

| **Variable** | **Mnemonic** |
|---|---|
| Real personal income | RPI ($\Delta \ln$) |
| Real PCE | DPCERA3M086SBEA ($\Delta \ln$) |
| Real manufacturing and trade sales | CMRMTSPLx ($\Delta \ln$) |
| Industrial production | INDPRO ($\Delta \ln$) |
| Capacity utilization in manufacturing | CUMFNS |
| Civilian unemployment rate | UNRATE |
| Total nonfarm employment | PAYEMS ($\Delta \ln$) |
| Hours worked: goods-producing | CES0600000007 (ln) |
| Average hourly earnings: goods-producing | CES0600000008 ($\Delta \ln$) |
| PPI for finished goods | PPIFGS ($\Delta \ln$) |
| PPI for commodities | PPICMM ($\Delta \ln$) |
| PCE price index | PCEPI ($\Delta \ln$) |
| Federal funds rate | FEDFUNDS |
| Total housing starts | HOUST (ln) |
| S&P 500 price index | S&P 500 ($\Delta \ln$) |
| U.S.-U.K. exchange rate | EXUSUKx ($\Delta \ln$) |
| 1 yr. Treasury - FEDFUNDS spread | T1YFFM |
| 10 yr. Treasury - FEDFUNDS spread | T10YFFM |
| BAA - FEDFUNDS spread | BAAFFM |
| ISM: new orders index | NAPMNOI |

Updating the paper's Figure 3, the left panels of Figure 3 display the root mean squared forecast error (RMSFE) relative (ratio) to the benchmark (the 20-variable homoskedastic VAR), so that a value below 1 denotes a model outperforming the benchmark. The large

homoskedastic model outperforms the small homoskedastic model for all variables at all horizons, suggesting that the inclusion of more data does improve the specification of the conditional means and therefore the point forecasts. The inclusion of time variation in volatilities consistently improves the performance of the small model, and for FEDFUNDS it also outperforms the benchmark at long horizons. However, the small heteroskedastic model is still largely dominated by the benchmark at short forecast horizons. The model with both time-varying volatilities and a large cross-section instead provides systematically better point forecasts than the benchmark (and than the other models).

The right-hand panels of Figure 3 present results for density forecasts, based on the average log scores. The figure displays the average log scores relative (difference) to the benchmark (the 20-variable homoskedastic VAR), so that a value above 0 denotes a model outperforming the benchmark. Both homoskedastic specifications perform quite poorly in density forecasting, while the heteroskedastic ones can achieve very high gains. Moreover, the large heteroskedastic system consistently outperforms the small heteroskedastic system. In combination with the findings for point forecasts, this result suggests that while both heteroskedastic models provide a better assessment of the overall uncertainty around the forecasts, the model based on the large cross-section centers such uncertainty around a more reliable mean, thereby obtaining further gains in predictive accuracy.

Updating the paper's Figure 4 and 5, Figures 4 and 5 compare forecasts for all the variables included in the cross-section of the models with 20 variables, with point forecasts in Figure 4 and density forecasts in Figure 5. In all of the subplots in Figure 4, the $x$ axes measure the RMSFE obtained by the large VAR when we allow for stochastic volatility, while the $y$ axes measure the same loss function (RMSFE) obtained by the homoskedastic specification. Each point corresponds to a different forecast horizon, and when a point is *above* the 45-degree line, this shows that the RMSFE of the heteroskedastic specification is smaller, indicating that the inclusion of variation in the volatility improved point forecasting performance. As is clear in the graph, in several instances the models produce point forecasts of similar accuracy. However, as the forecast horizon increases (which can be indirectly inferred from the graph as in general higher RMSFEs correspond to longer forecast horizons), the specification with variation in the volatilities tends to outperform the homoskedastic version of the model. The mechanism at play is as follows: the heteroskedastic model provides more efficient estimates and therefore a better characterization of the predictive densities, while the homoskedastic model is misspecified and therefore provides an inferior characterization of the predictive densities. At short forecast horizons, this does not have much effect on point forecasts, but as the forecast horizon increases, the predic-

tive densities cumulate nonlinearly and therefore the misspecification of the homoskedastic model increasingly reduces the relative accuracy.

In Figure 5, the $x$ axes measure the (log) density score obtained by the large VAR when we allow for stochastic volatility, while the $y$ axes measure the same gain function (score) obtained by the homoskedastic specification. Each point corresponds to a different forecast horizon, and when a point is *below* the 45-degree line, this shows that the score of the heteroskedastic specification is larger, indicating that the inclusion of variation in the volatility improved density forecasting performance. The improvement coming from the introduction of time variation in the volatilities is striking, and it is common to nearly all variables. Clearly, stochastic volatility improves the overall assessment of uncertainty with respect to the homoskedastic model, and it does so both directly, by simply using a better variance around the point estimates, and indirectly, by centering the densities toward improved point forecasts (as documented in Figure 4).

## 2 Additional empirical results: algorithm comparison

Finally, to further compare results across algorithms, Figure 6 shows the impulse response functions for a federal funds rate shock (recursively identified) delivered by the various algorithms. This figure extends and updates results in the online appendix to the original paper. In these new results, response estimates from the system-wide algorithm (SWA) and corrected triangular algorithm (CTA) coincide perfectly, while the original triangular algorithm (TA) does show some occasional deviations.

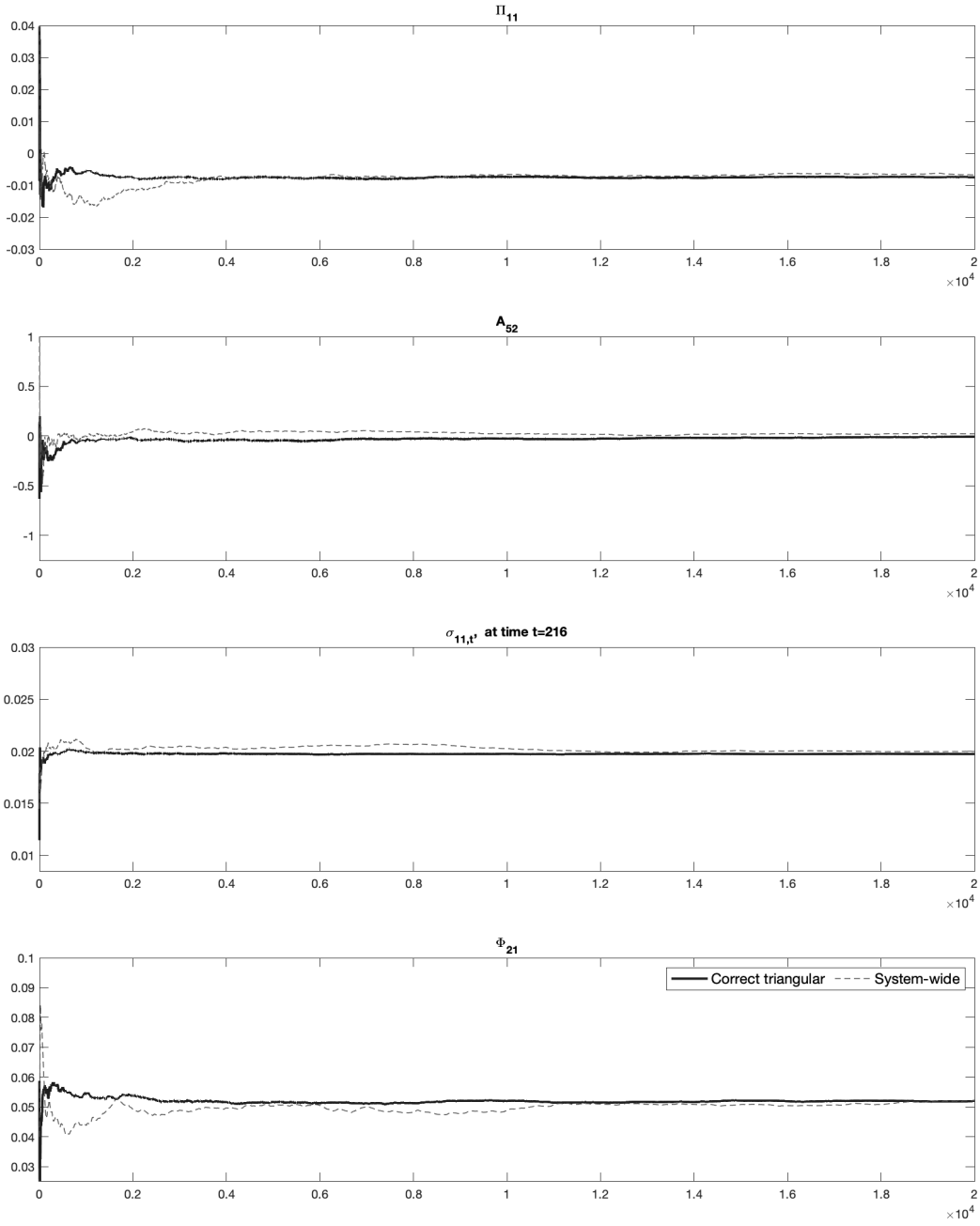


Figure 1: Update of Figure 2 of the paper. Recursive means of selected coefficients. Comparison between the system-wide and correct triangular algorithm. Chains are initialised at the same value (set equal to the priors).

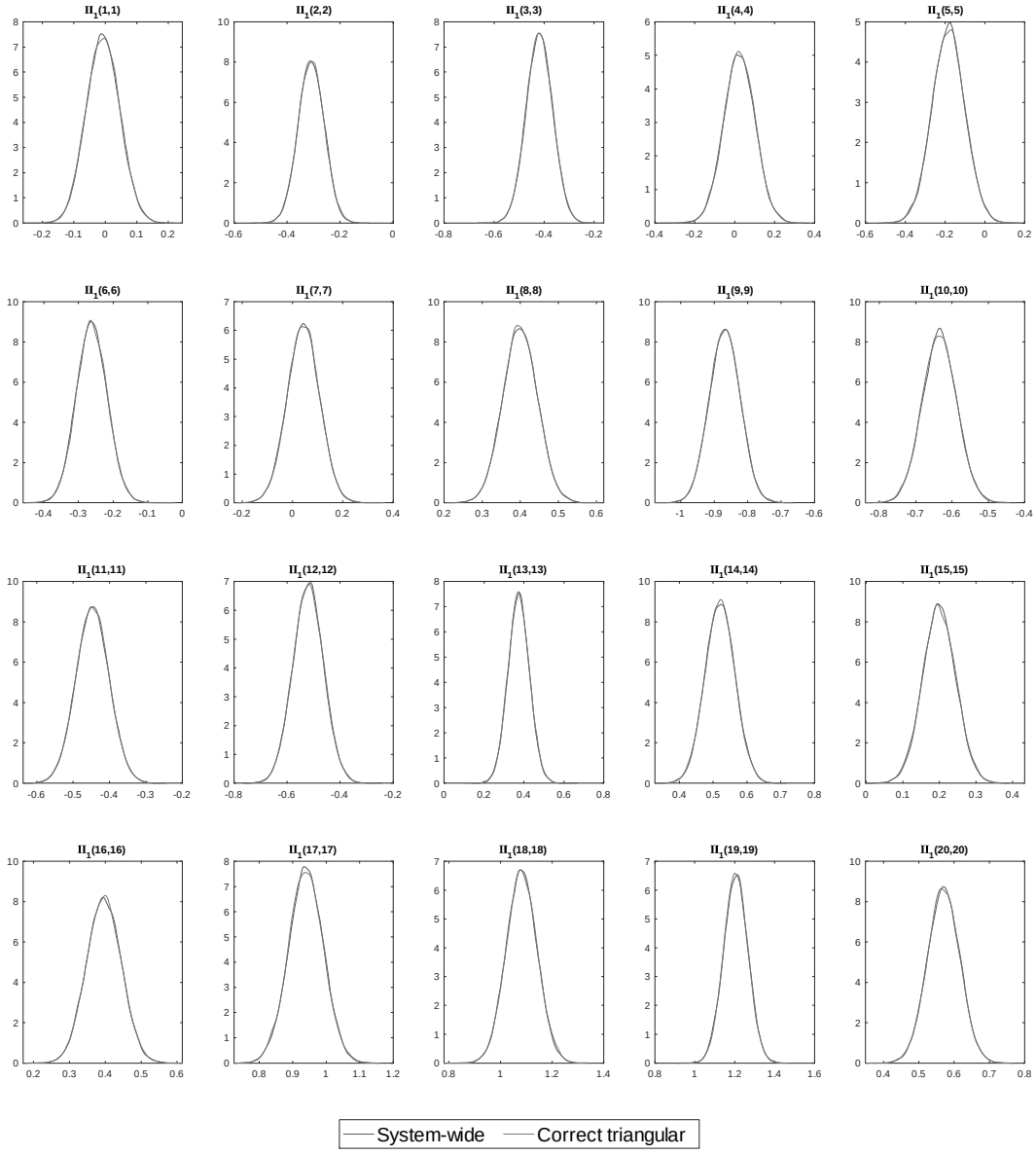


Figure 2: Posterior distribution of the coefficients on the main diagonal of the matrix $\Pi_1$, under the SWA and CTA.

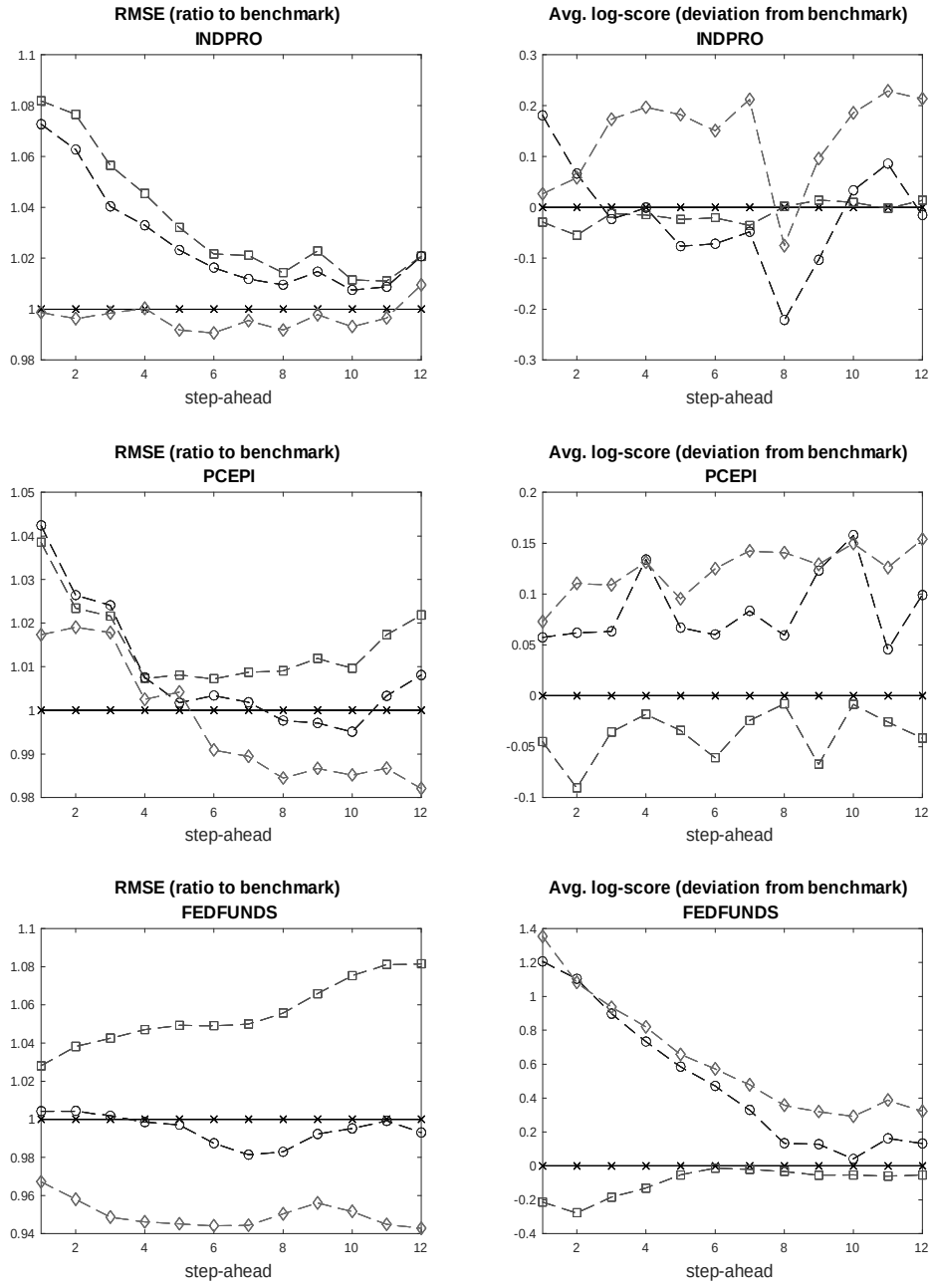


Figure 3: Update of Figure 3 of the paper. Panels on the left-hand side contain results for the point forecasts (relative RMSE of different models vs benchmark). The panels on the right-hand side contain results for the density forecasts (log-score gains of different models vs benchmark). In all the panels, crosses represent the homoskedastic VAR with 20 variables (the benchmark model), the squares represent the homoskedastic VAR with 3 variables, the circles represent the heteroskedastic VAR with 3 variables, and the diamonds represent the heteroskedastic VAR with 20 variables.

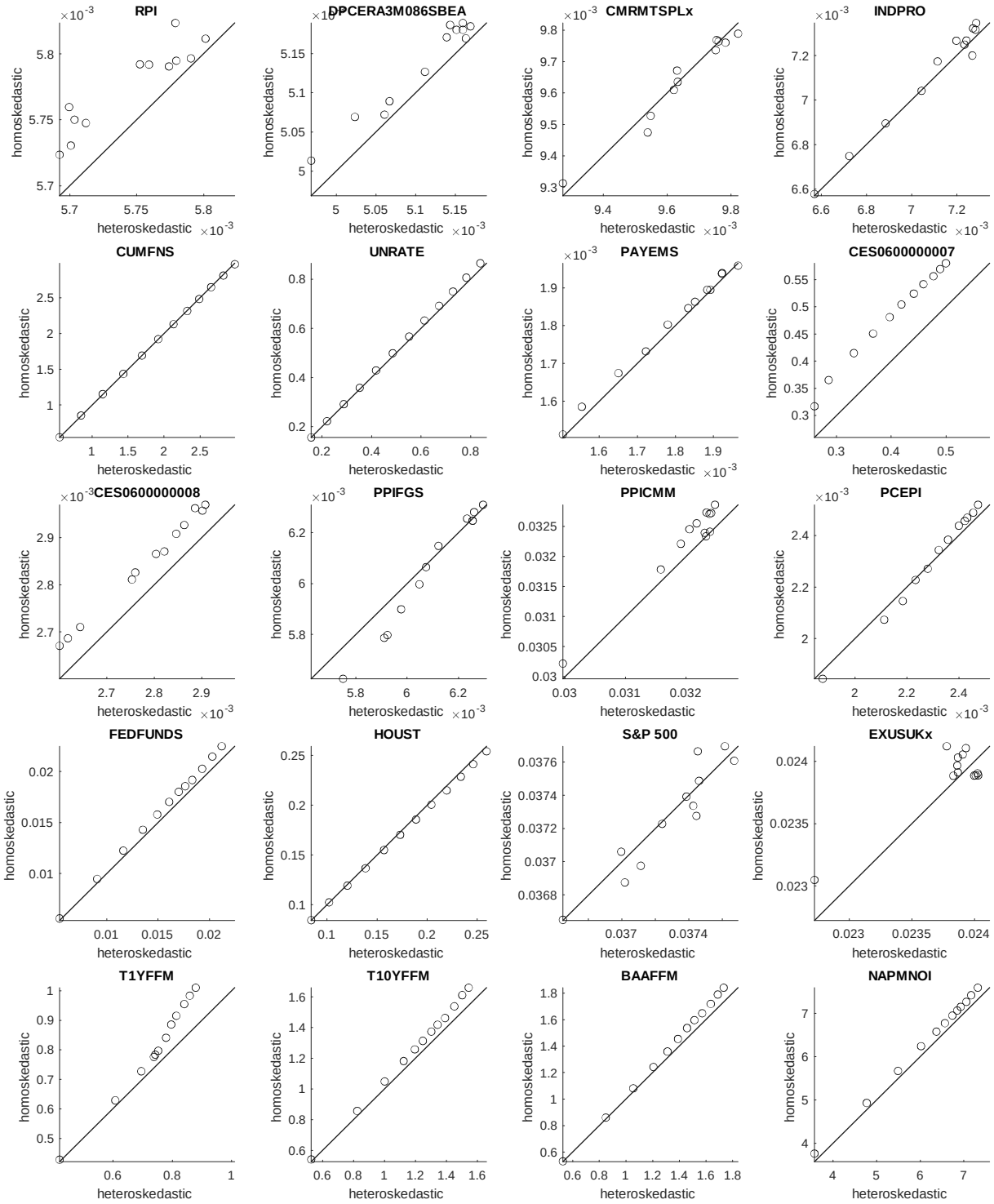


Figure 4: Update of Figure 4 of the paper. Comparison of point forecast accuracy. Each panel describes a different variable. The x axis reports the RMSFE obtained using the BVAR with stochastic volatility (heteroskedastic), the y axis reports the RMSFE obtained using the homoskedastic BVAR. Each point corresponds to a different forecast horizon from 1 to 12 step-ahead (in most cases, a higher RMSFE corresponds to a longer forecast horizon).

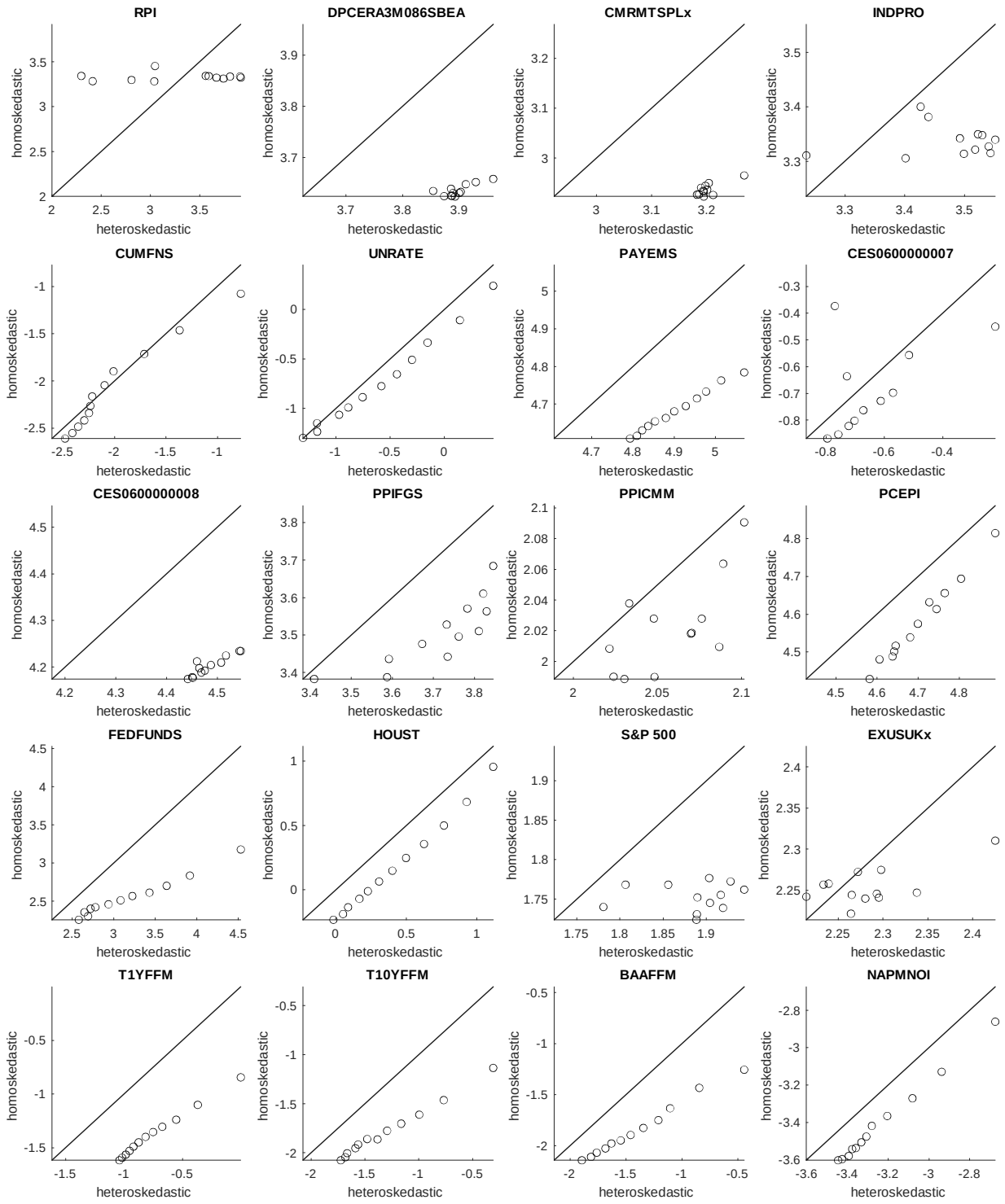


Figure 5: Update of Figure 5 of the paper. Comparison of density forecast accuracy. Each panel describes a different variable. The x axis reports the (log) density score obtained using the BVAR with stochastic volatility (heteroskedastic), the y axis reports the (log) density score obtained using the homoskedastic BVAR. Each point corresponds to a different forecast horizon from 1 to 12 step-ahead.

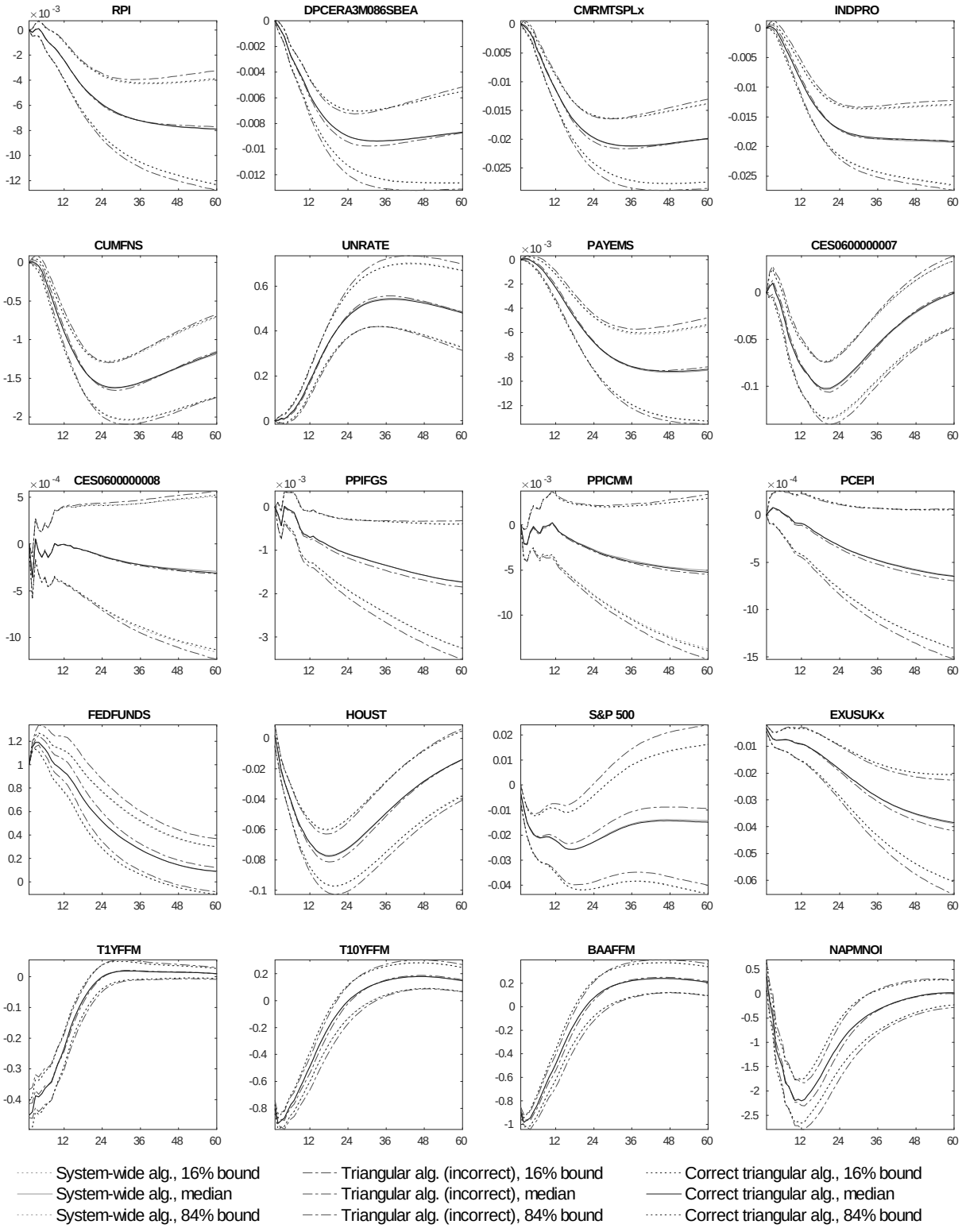


Figure 6: Comparison of impulse responses computed using the alternative algorithms. Model with monthly data, N=20 and p=13.

## 3 The Correct Triangular Algorithm for Homoskedastic VARs

This section outlines the implementation of the correct triangular algorithm to the following homoskedastic VAR:

$$y_t = \Pi' x_t + \epsilon_t,$$

where $\epsilon_t \sim iid\, N(0, \Sigma), t = 1, \ldots, T$. The coefficient priors are Gaussian $\pi^{(j)} \sim N(\underline{\mu}_{\pi^{(j)}}, \underline{\Omega}_{\pi^{(j)}})$, $j = 1, \ldots, N$, and they are independent across equations. We first use an LDL decomposition (sometimes referred to as a square-root-free Cholesky decomposition) to factor the reduced-form error covariance matrix as $\Sigma = A^{-1}\Lambda A^{-1'}$, where $A^{-1}$ is a lower triangular matrix with ones on its main diagonal and $\Lambda$ is a diagonal matrix with generic $j$-th element $\lambda_j$.[1] Then, the full posterior density of $\pi^{(j)}$ has a multivariate normal distribution

$$(\pi^{(j)} \,|\, y, \Sigma, \pi^{(-j)}) \sim \mathcal{N}(\overline{\mu}_{\pi^{(j)}}, \overline{\Omega}_{\pi^{(j)}}),$$

with

$$\overline{\Omega}^{-1}_{\pi^{(j)}} = \underline{\Omega}^{-1}_{\pi^{(j)}} + \textstyle\sum_{i=j}^{N} \frac{a_{i,j}^2}{\lambda_i} \sum_{t=1}^{T} x_t x_t' \tag{1}$$

$$\overline{\mu}_{\pi^{(j)}} = \overline{\Omega}_{\pi^{(j)}} \left( \underline{\Omega}^{-1}_{\pi^{(j)}} \underline{\mu}_{\pi^{(j)}} + \textstyle\sum_{i=j}^{N} \frac{a_{i,j}}{\lambda_i} \sum_{t=1}^{T} x_t z_{i,t} \right), \tag{2}$$

where $z_{j+l,t} = \tilde{y}_{j+l,t} - \sum_{i \neq j, i=1}^{j+l} a_{j+l,i} x_t' \pi^{(i)}$, for $l = 0, \ldots, N-j$, and $a_{i,i} = 1$.

Next, we provide an alternative expression for the moments (1) and (2) which works more efficiently in the commonly-used Matlab software package. Let $L^{0.5} = 1_T \otimes (\lambda_j^{0.5}, \ldots, \lambda_N^{0.5})$, where $1_T$ is a $T \times 1$ column of ones. Moreover, define

$$\begin{aligned} \underset{T(N-j+1)\times 1}{Y^{(j)}} &= vec((y - X\Pi^{[j=0]})A^{(j:N,1:N)'})./vec(L^{0.5}), \\ \underset{T(N-j+1)\times k}{X^{(j)}} &= (A^{(j:N,j)} \otimes X)./vec(L^{0.5}) \end{aligned}$$

where ./ is the Matlab element-by-element division operator. Then, we obtain an equivalent expression for the posterior moments appearing in (1) and (2):

$$\begin{aligned} \overline{\Omega}^{-1}_{\pi^{(j)}} &= (\underline{\Omega}^{-1}_{\pi^{(j)}} + X^{(j)\prime} X^{(j)}), \\ \overline{\mu}_{\pi^{(j)}} &= \overline{\Omega}_{\pi^{(j)}} (\underline{\Omega}^{-1}_{\pi^{(j)}} \underline{\mu}_{\pi^{(j)}} + X^{(j)\prime} Y^{(j)}). \end{aligned}$$

It is also possible to draw from the multivariate normal distribution posterior density of $\pi^{(j)}$ given by $(\pi^{(j)} \,|\, y, \Sigma, \pi^{(-j)}) \sim \mathcal{N}(\overline{\mu}_{\pi^{(j)}}, \overline{\Omega}_{\pi^{(j)}})$ with a Cholesky decomposition of $\Sigma$. In

[1] This decomposition can be computed either with a function such as the LDL function in Matlab or from a normalization of the Cholesky decomposition.

this case, let $A^{\dagger}$ denote the Cholesky decomposition of $\Sigma$, so that $\Sigma = A^{\dagger}A^{\dagger\prime}$, and define $\tilde{A} = A^{\dagger -1}$, with elements $\tilde{a}_{i,j}$ in row $i$, column $j$. In this case, $\tilde{y}_t$ is defined as $\tilde{y}_t = \tilde{A}y_t$, and $z_{j+l,t} = \tilde{y}_{j+l,t} - \sum_{i\neq j, i=1}^{j+l} \tilde{a}_{j+l,i}x_t'\pi^{(i)}$, for $l = 0, \ldots, N-j$. Then the posterior moments of the conditional normal distribution take the form:

$$\overline{\Omega}_{\pi^{(j)}}^{-1} = \underline{\Omega}_{\pi^{(j)}}^{-1} + \sum_{i=j}^{N} \tilde{a}_{i,j}^2 \sum_{t=1}^{T} x_t x_t', \tag{3}$$

$$\overline{\mu}_{\pi^{(j)}} = \overline{\Omega}_{\pi^{(j)}} \left( \underline{\Omega}_{\pi^{(j)}}^{-1} \underline{\mu}_{\pi^{(j)}} + \sum_{i=j}^{N} \tilde{a}_{i,j} \sum_{t=1}^{T} x_t z_{i,t} \right). \tag{4}$$