Web Appendix for "Adaptive shrinkage in Bayesian vector autoregressive models"

Florian Huber[*1] and Martin Feldkircher[2]

[1]Vienna University of Economics and Business
[2]Oesterreichische Nationalbank (OeNB)

Appendix A   Data overview

Name	Fred mnemonic	Model size $m=3$	$m=7$	$m=21$	Trans
Real Gross Domestic Product	GDPC96	x	x	x	1
Real Personal Consumption Expenditures	PCECC96		x	x	1
Real Gross Private Domestic Investment	GPDIC96		x	x	1
Real private fixed investment: Residential	PRFIx			x	1
Industrial Production Index	INDPRO			x	1
Capacity Utilization: Manufacturing	CUMFNS			x	1
All Employees: Service-Providing Industries	SRVPRD			x	1
Civilian Employment	CE16OV			x	1
Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing	AWHMAN		x	x	2
Personal Consumption Expenditures	PCECTPI			x	1
Gross Domestic Product: Chain-type Price Index	GDPCTPI	x	x	x	1
Gross Private Domestic Investment: Chain-type Price Index	GPDICTPI			x	1
Consumer Price Index for All Urban Consumers: All Items	CPIAUCSL			x	1
Real Average Hourly Earnings of Production and Nonsupervisory Employees: Construction	CES2000000008x		x	x	1
Effective Federal Funds Rate	FEDFUNDS	x	x	x	2
1-Year Treasury Constant Maturity Rate	GS1			x	2
10-Year Treasury Constant Maturity Rate	GS10			x	2
Real M2 Money Stock	M2REALx			x	1
U.S. / U.K. Foreign Exchange Rate	EXUSUKx			x	1
University of Michigan: Consumer Sentiment	UMCSENTx			x	2
S&Ps Common Stock Price Index: Composite	S&P 500			x	1

Notes: Data are obtained from the quarterly variant of the dataset provided by McCracken and Ng (2016). Trans = Transformation codes are as follows: 1 - log difference, 2 - raw data, 3- difference of the log difference

Table A.1: Data overview

Appendix B   Additional empirical results

Fig. B.1: Evolution of ϑ_ψ and ϑ_ϕ over the hold-out period

Posterior mean of ϑ_ψ

Posterior mean of ϑ_ϕ
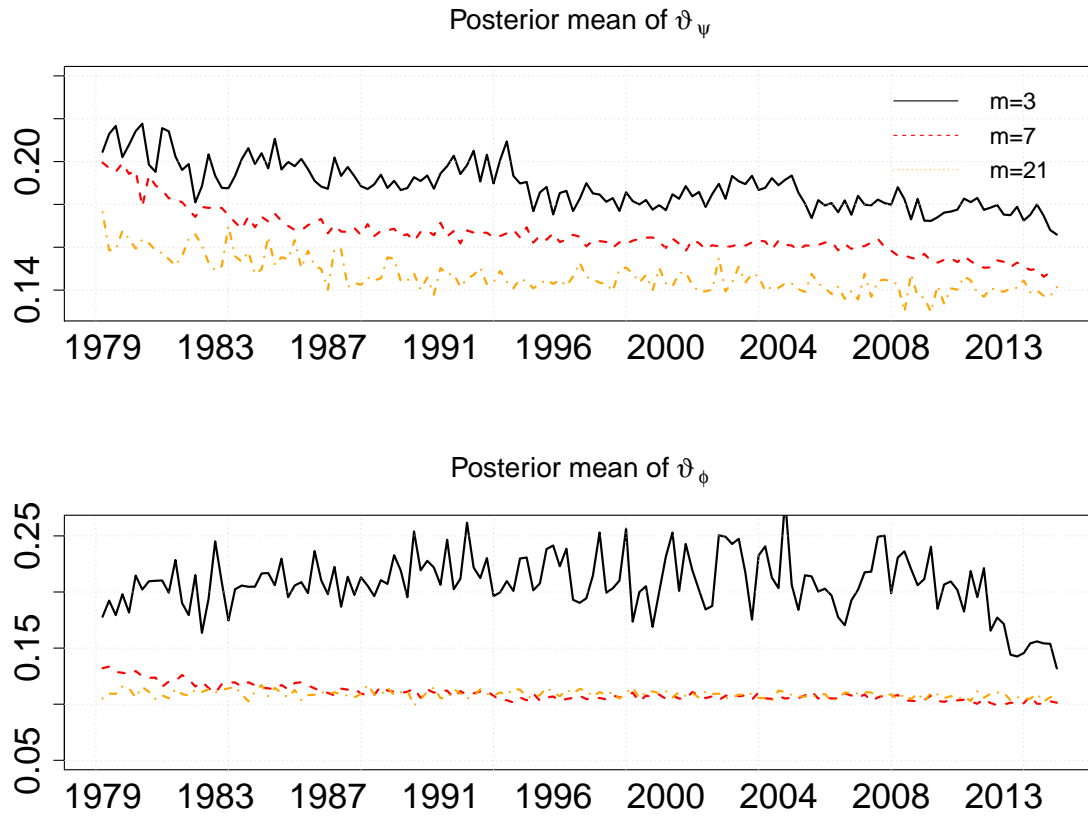
	GDPC96	PCECC96	GPDIC96	PRFIx	INDPRO	CUMFNS	SRVPRD	CE16OV	AWHMAN	PCECTPI	GDPCTPI
NG-global	**529.331**	**546.588**	287.694	**305.904**	438.161	422.881	649.915	579.930	526.676	578.356	653.901
NG-rowwise	527.033	546.230	280.849	303.876	436.556	421.223	660.936	578.439	536.149	574.098	658.506
NG-columnwise	527.828	540.017	288.095	305.161	440.607	**427.505**	662.119	575.942	520.227	**583.744**	656.723
NG-lagwise	527.424	539.890	278.725	305.545	**445.336**	421.452	**665.300**	**588.461**	**544.486**	583.314	**663.689**
SSVS	523.360	506.789	278.556	298.870	424.117	411.446	614.045	545.138	508.915	544.793	619.834
Minnesota	528.809	525.909	**292.413**	297.237	430.114	416.544	615.633	557.709	524.349	557.600	633.379
	GPDICTPI	CPIAUCSL	CES2000000008x	FEDFUNDS	GS1	GS10	M2REALx	EXUSUKx	UMCSENTx	S.P.500	
NG-global	609.524	535.312	561.461	479.871	485.204	484.132	392.861	**238.678**	**186.343**	**200.548**	
NG-rowwise	610.142	522.079	**573.884**	482.062	484.515	485.192	390.432	236.172	178.577	196.890	
NG-columnwise	**622.914**	**541.077**	565.666	477.713	**489.491**	**500.304**	393.440	234.911	188.838	198.265	
NG-lagwise	617.469	525.318	557.263	479.589	488.188	495.938	382.298	233.307	161.971	194.016	
SSVS	575.359	506.216	536.599	471.312	469.139	465.474	359.403	233.647	175.696	191.748	
Minnesota	572.719	518.336	534.961	**487.412**	470.684	467.530	**393.680**	237.635	182.290	199.284	

Notes: Absolute values of cumulative one-step-ahead log predictive scores. The bold figures indicate the best performing model for a given variable. NG-global stands for a vector autoregressive model coupled with the baseline NG prior, NG-rowwise is the VAR model with the rowwise NG specification, columnswise the VAR model with the columnwise NG specification and NG-lagwise denotes the VAR model with the lagwise NG specification. SSVS refers to a VAR coupled with the SSVS prior of George et al. (2008) and Minnesota denotes the hierarchical Minnesota prior VAR.

Table B.1: Out-of-sample performance in terms of the sum of one-step-ahead log predictive scores (LPS): 1979Q3 to 2015Q2.

Appendix C   Additional conditional posterior distributions

The conditional posterior of the autoregressive coefficients $\boldsymbol{\alpha}$ takes a standard form (Kadiyala et al., 1997; Karlsson, 2013), namely a k-dimensional multivariate Gaussian distribution. However, if the number of endogenous variables and the lags thereof becomes large, traditional sampling algorithms based on the full system of equations are prohibitively slow, since the posterior variance-covariance matrix of $\boldsymbol{\alpha}$ has to be inverted within each iteration of the MCMC algorithm. We thus follow Carriero et al. (2015) and exploit the fact that conditional on $\boldsymbol{H}$, the VAR can be written as m unrelated regression models with the errors of the preceding $j-1$ equations being included as additional regressors in the jth equation. This simplifies computation enormously, allowing for large scale models that feature stochastic volatility.

More specifically, note that conditional on $\boldsymbol{H}^{-1}$ and the remaining parameters, the first equation of the VAR can be written as (Carriero et al., 2015)

$$y_{1t} = \boldsymbol{A}_{1\bullet}\boldsymbol{X}_t + s_{1t}^{1/2}\eta_{1t}, \tag{C.1}$$

where the notation $\boldsymbol{A}_{1\bullet}$ indicates that the first row of $\boldsymbol{A}$ is selected and η_{1t} is a standard normally distributed innovation.[1] A generic equation $i = 2, \ldots, m$ is given by

$$y_{it} = \boldsymbol{A}_{i\bullet}\boldsymbol{X}_t + \sum_{j=1}^{i-1} \tilde{h}_{ij}s_{it}^{1/2}\eta_{jt} + s_{it}^{1/2}\eta_{it}, \tag{C.2}$$

with $\tilde{h}_{ij}$ denoting the (negative) free off-diagonal elements of $\boldsymbol{H}^{-1}$. This implies that the full conditional posterior of the ith row of $\boldsymbol{A}$, denoted by $\boldsymbol{A}_{i\bullet}$ takes a particularly simple form, namely that of a standard regression model with heteroscedastic errors.

More precisely, the conditional posterior of $\boldsymbol{A}_{i\bullet}$ is (Carriero et al., 2015)

$$\boldsymbol{A}_{i\bullet}'|\boldsymbol{A}_{1:i-1\bullet}, \boldsymbol{H}, \boldsymbol{\psi}, \lambda_\psi^2, \boldsymbol{S}_{1:T}, \boldsymbol{Y} \sim \mathcal{N}(\overline{\boldsymbol{A}}_{i\bullet}', \overline{\boldsymbol{V}}_{A_i}). \tag{C.3}$$

Hereby, we let $\boldsymbol{A}_{1:i-1\bullet} = (\boldsymbol{A}_{1\bullet}, \ldots, \boldsymbol{A}_{i-1\bullet})$ and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_k)'$. Note that conditional on $\boldsymbol{H}$, the corresponding hyperparameters carry no additional information. The posterior mean and variance are, respectively,

$$\overline{\boldsymbol{A}}_{i\bullet}' = \overline{\boldsymbol{V}}_{A_i}^{-1}\left(\sum_{t=1}^T \boldsymbol{X}_t\tilde{y}_{it}s_{it}^{-0.5} + \underline{\boldsymbol{V}}_{\alpha i}^{-1}\underline{\alpha}_i\right), \tag{C.4}$$

$$\overline{\boldsymbol{V}}_{A_i}^{-1} = \left(\sum_{t=1}^T \boldsymbol{X}_t\boldsymbol{X}_t's_{it}^{-0.5} + \underline{\boldsymbol{V}}_{\alpha i}^{-1}\right). \tag{C.5}$$

---

[1]In what follows we moreover assume that each equation features the same set of explanatory variables $\boldsymbol{X}_t$.

We let $\underline{\boldsymbol{V}}_{\alpha i}$ denote the ith block of a a $k \times k$-dimensional diagonal matrix $\underline{\boldsymbol{V}}_\alpha$ with

$$[\underline{\boldsymbol{V}}_\alpha]_{jj} = 2/\lambda_\psi^2 \psi_j, \; j = 1, \ldots, k. \tag{C.6}$$

and $\tilde{y}_{it} = y_{it} - \sum_{j=1}^{i-1} \tilde{h}_{ij} s_{it}^{1/2} \eta_{jt}$ if $i > 1$ and $\tilde{y}_{it} = y_{it}$ if $i = 1$.

The same holds true for the covariance parameters in $\boldsymbol{H}$, which are obtained by running $m - 1$ univariate regression models with stochastic volatility (Cogley and Sargent, 2005). To see this, note that conditional on $\boldsymbol{A}$, we can rewrite the model as follows

$$\boldsymbol{H}\boldsymbol{\varepsilon}_t = \boldsymbol{S}_t^{1/2}\boldsymbol{\eta}_t, \tag{C.7}$$

with $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{mt})'$ and $\boldsymbol{S}_t = \boldsymbol{S}_t^{1/2}\boldsymbol{S}_t^{1/2}$. The first equation of Eq. (C.7) is the identity,

$$\varepsilon_{1t} = s_{1t}^{1/2}\eta_{1t}. \tag{C.8}$$

For $i = 2, \ldots, m$, it is easy to show that a typical equation is

$$\varepsilon_{it} = \sum_{j=1}^{i-1} h_{ij}\varepsilon_{jt} + s_{it}^{1/2}\eta_{it}, \tag{C.9}$$

which, again, is a standard regression model with heteroscedastic innovations.

The conditional posterior distribution associated with the covariance parameters of equation i is given by

$$\boldsymbol{h}_i | \boldsymbol{\phi}, \lambda_\phi^2, \boldsymbol{S}_{1:T}, \boldsymbol{A}, \boldsymbol{Y} \sim \mathcal{N}(\overline{\boldsymbol{h}}_i, \overline{\boldsymbol{V}}_{ih}), \tag{C.10}$$

with $\boldsymbol{h}_i = (h_{i1}, \ldots, h_{ii-1})'$ and $\boldsymbol{\phi}_i = (\phi_{i1}, \ldots, \phi_{ii-1})'$. The (inverse) posterior variance equals

$$\overline{\boldsymbol{V}}_{ih}^{-1} = \left(\sum_{t=1}^T \boldsymbol{\xi}_{it}\boldsymbol{\xi}_{it}' s_{it}^{-1/2} + \underline{\boldsymbol{V}}_{ih}^{-1} \right)^{-1}. \tag{C.11}$$

Here, we let $\boldsymbol{\xi}_{it} = (\varepsilon_{1t}, \ldots, \varepsilon_{i-1t})'$ denote the errors of the preceding $i - 1$ equations and $\underline{\boldsymbol{V}}_{ih}$ is the prior variance matrix of the Normal-Gamma prior.

The posterior mean is

$$\overline{\boldsymbol{h}}_i = \overline{\boldsymbol{V}}_{ih}^{-1} \left(\sum_{t=1}^T \boldsymbol{\xi}_{it}\varepsilon_{it} s_{it}^{-1/2} \right). \tag{C.12}$$

Finally, the (conditional) posterior distribution of the full history of log-volatilities has no well known form. We thus opt for the algorithm proposed in Kastner and Frühwirth-Schnatter (2014) to simulate the full history of log-volatilities and the coefficients of the state equation (which are based on simple Gibbs and Metropolis Hastings steps).

Appendix D   Derivations

D.1   Derivations related to the baseline Normal-Gamma prior

To derive Eq. (2.8), note that due to the hierarchical nature of the model, the conditional posterior of ψ_i is independent from the data. Combining the likelihood with the prior yields

$$p(\psi_i|\vartheta_\psi, \lambda_\psi^2, \alpha_i) \propto \psi_\alpha^{-1/2} \exp\left(-\frac{\alpha_i^2}{2\psi_i}\right) \times \psi_i^{(\vartheta_\psi-1)} \exp\left(-\frac{\vartheta_\psi \lambda_\psi^2 \psi_i}{2}\right), \tag{D.1}$$

$$\propto \psi_i^{(\vartheta_\psi-0.5)-1} \exp\left(-(\alpha_i^2/\psi_i + \vartheta_\psi \lambda_\psi^2 \psi_i)/2\right), \tag{D.2}$$

where we exploit the scaling property of the Gamma distribution to rewrite the prior in Eq. (2.4) as

$$\alpha_i|\psi_i \sim \mathcal{N}(0, \psi_i), \ \psi_i \sim \mathcal{G}(\vartheta_\psi, \vartheta_\psi \lambda_\psi^2/2). \tag{D.3}$$

Equation (D.2) is the kernel of the GIG distribution described in Eq. (2.8).

We derive Eq. (2.10) by combining the Gamma likelihood with the prior and simplifying

$$p(\lambda_\psi^2|\boldsymbol{\psi}, \vartheta_\psi) \propto (\lambda_\psi^2)^{(k\vartheta_\psi+c_{\psi0})-1} \times \exp\left(-(c_{\psi1} + \vartheta_\psi/2\sum_{j=1}^{k}\psi_j)\lambda_\psi^2\right), \tag{D.4}$$

which is the kernel of a Gamma density with shape parameter equal to $k\vartheta_\psi + c_{\psi0}$ and rate parameter given by $c_{\psi1} + \vartheta_\psi/2\sum_{j=1}^{k}\psi_j$.

The derivation of Eq. (2.11) closely resembles the derivation of Eq. (2.8). Finally, the derivation of Eq. (2.13) is analogous to the derivation of Eq. (2.10).

D.2   Derivations related to the three extensions of the baseline Normal-Gamma prior

As noted in Section 2, the relevant conditional posterior distributions can still be used with only minor alterations.

First, due to the presence of different global shrinkage parameters that are either row-, column- or lag-specific we have to modify $\underline{\boldsymbol{V}}_\alpha$ accordingly. For instance, in the case of the rowwise specification, the mp elements relating to the m different equations of $\underline{\boldsymbol{V}}_\alpha$ feature an equation-specific shrinkage parameters $\lambda_{\psi i}^2$ and ϑ_{psii} for $i = 1, \ldots, m$. In the case of the columnwise specification the prior variance matrix has to be modified to take into account mp different shrinkage parameters. Finally, for the lagwise specification we have to adapt $\underline{\boldsymbol{V}}_\alpha$ such that elements associated with $\boldsymbol{A}_j$ $(j = 1, \ldots, p)$ feature lag-specific shrinkage parameters $\lambda_{\psi j}$ and $\vartheta_{\psi j}$. Under these

slight modifications of the prior variance-covariance matrix, the conditional posterior of A remains the same.

For all modifications we let $\mathcal{A}_j^{(n)}$ for specification $n = 1, 2, 3$ denote a generic index set that selects the appropriate elements of α (i.e., all elements of the jth row, jth column or the jth lag). The modified counterpart of Eq. (2.8) is then given by

$$\psi_i | \vartheta_\psi, \lambda_\psi^2, \alpha_i \sim \mathcal{GIG}\left(\vartheta_\psi - \frac{1}{2}, \vartheta_{\psi j}\lambda_{\psi j}^2, \alpha_i^2\right) \quad \text{for} \quad i = 1, \ldots, k. \tag{D.5}$$

Similarly, we adapt Eq. (2.10) as

$$\lambda_{\psi j}^2 | \vartheta_{\psi j}, \psi \sim \mathcal{G}\left(c_{\psi 0} + \vartheta_{\psi j} q_j^{(n)}, c_{\psi 1} + \vartheta_\psi/2 \sum_{i \in \mathcal{A}_j^{(n)}} \psi_i\right). \tag{D.6}$$

Here, we let $q_j^{(n)} = \#(\mathcal{A}_s^{(n)})$ denote the cardinality of $\mathcal{A}_j^{(n)}$.

For the third (lagwise) case the specific structure of $\lambda_{\psi j}^2 = \prod_{i=1}^j \zeta_i$ calls for additional derivations. For the first lag, we combine the Gamma likelihood with the Gamma prior to obtain,

$$p(\zeta_1 | \vartheta_{\psi 1}, \psi) \propto \zeta_1^{(\vartheta_{\psi 1} m^2 + d_1) - 1} \times \exp\left\{-\left(\vartheta_{\psi 1}/2 \sum_{j \in \mathcal{A}_j^{(3)}} \psi_j + l_1\right)\zeta_1\right\}, \tag{D.7}$$

which is the kernel of a Gamma distribution with parameters $(\vartheta_{\psi 1} m^2 + d_1)$ and $\vartheta_{\psi 1}/2 \sum_{j \in \mathcal{A}_j^{(3)}} \psi_j + l_1$.

For higher lag orders $\mathfrak{g} = 2, \ldots, p$ one can show that the conditional posterior is again Gamma distributed,

$$p(\zeta_\mathfrak{g} | \vartheta_{\psi\mathfrak{g}}, \psi, \lambda_{\psi\mathfrak{g}-1}^2) \propto \zeta_\mathfrak{g}^{(\vartheta_{\psi\mathfrak{g}} m^2 + d_\mathfrak{g}) - 1} \times \exp\left\{-\left(\vartheta_{\psi\mathfrak{g}}\lambda_{\psi\mathfrak{g}-1}^2/2 \sum_{j \in \mathcal{A}_\mathfrak{g}^{(3)}} \psi_j + l_\mathfrak{g}\right)\zeta_\mathfrak{g}\right\}, \tag{D.8}$$

with parameters $\vartheta_{\psi\mathfrak{g}} m^2 + d_\mathfrak{g}$ and $(\vartheta_{\psi\mathfrak{g}}\lambda_{\psi\mathfrak{g}-1}^2/2 \sum_{j \in \mathcal{A}_\mathfrak{g}^{(3)}} \psi_j + l_\mathfrak{g})$. For all specifications, the acceptance probability for the MH step needs to be modified to take into account that we sample different $\vartheta_{\psi j}$.

Steps (4) to (6) of the MCMC algorithm presented in Appendix E have to be modified to draw distinct $\lambda_{\psi j}^2$ and $\vartheta_{\psi j}$ for each variant of the prior. These steps are straightforward to implement and do not increase the computational burden considerably.

Appendix E Full conditional MCMC algorithm

A relatively straightforward MCMC scheme can be devised by iteratively drawing from the conditional posterior distributions described in the previous subsections. We present the exact algorithm for the case of a single global shrinkage parameter only. The needed alterations for the other variants of the NG prior are straightforward to implement and only steps (4) to (6) are affected. For the baseline prior setup, the MCMC algorithm cycles through the following steps

Step 0 Initialize all parameters of the model by using the corresponding OLS estimates or by drawing from the prior

Step 1 Draw $\boldsymbol{\alpha}$ using the algorithm put forth in Carriero et al. (2015) on an equation-by-equation basis. The exact conditional posterior distribution for the regression coefficients is given in Eq. (C.3).

Step 2 Draw $\boldsymbol{h}_i = (h_{i1}, \ldots, h_{ii-1})'$ for $i = 2, \ldots, m$ from simple normally distributed posterior distributions given by Eq. (C.10).

Step 3 Update the full history of log-volatilities $\boldsymbol{S}_{1:T} = (\boldsymbol{S}_1, \ldots, \boldsymbol{S}_T)$ and the coefficients of Eq. (2.3) for each equation $j = 1, \ldots, m$ using the algorithm proposed in Kastner and Frühwirth-Schnatter (2014) and implemented in Kastner (2015).

Step 4 Draw ψ_i for $i = 1, \ldots, k$ element-wise from Eq. (2.8) and ϕ_{ij} for $i = 2, \ldots, m;\ j = 1, \ldots, m - 1$ from Eq. (2.11).

Step 5 Draw λ_ψ^2 and λ_ϕ^2 from Eq. (2.10) and Eq. (2.13).

Step 6 Draw ϑ_ψ and ϑ_ϕ with a univariate random walk Metropolis Hastings step. We follow Griffin and Brown (2010) by proposing $\vartheta_j^* = \exp(\kappa_j z_j)\vartheta_j$ for $j \in \{\psi, \phi\}$. κ_j is set such that the acceptance probability lies between 20 and 40 percent and z_j is a standard normally distributed random variable.

After discarding a suitable amount of draws the algorithm is repeated N times. In all simulations and the empirical applications we find that convergence is typically smooth, with convergence diagnostics indicating rapid convergence towards the stationary distribution.

References

Carriero A, Clark TE, Marcellino M et al. (2015) Large vector autoregressions with asymmetric priors and time varying volatilities. *Queen Mary, University of London Working Paper Series, 759*

Cogley T and Sargent TJ (2005) Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic dynamics* 8(2), 262–302

George EI, Sun D and Ni S (2008) Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics* 142(1), 553–580

Griffin JE and Brown PJ (2010) Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis* 5(1), 171–188

Kadiyala KR, Karlsson S et al. (1997) Numerical methods for estimation and inference in Bayesian VAR models. *Journal of Applied Econometrics* 12(2), 99–132

Karlsson S (2013) Forecasting with Bayesian vector autoregressions. *Handbook of Economic Forecasting* 2, 791–897

Kastner G (2015) Dealing with stochastic volatility in time series using the R package stochvol. *Journal of Statistical Software. URL http://cran. r-project. org/web/packages/stochvol/vignettes/article. pdf* (forthcoming)

Kastner G and Frühwirth-Schnatter S (2014) Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis* 76, 408–423

McCracken MW and Ng S (2016) FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics* 34(4), 574–589