Internet Appendix for Measuring interconnectedness between financial institutions with Bayesian time-varying vector autoregressions

NOT FOR PUBLICATION

Marco Valerio Geraci and **Jean-Yves Gnabo**

Abstract

In this appendix, we describe the priors assumed for the parameters, the sampling algorithm for the posterior distribution, and the approach for computing Bayes factor. We also evaluate the time-varying approach against the classical approach of Granger causality testing in a series of simulation exercises.

This Internet Appendix is structured as follows. In Section A, we describe the prior distributions of the parameters. In Section B, we explain the Markov Chain Monte Carlo (MCMC) sampling algorithm. In Section C, we briefly highlight the estimation of Bayes factor for evaluating the time-varying null hypothesis $\mathcal{H}_{0t}^{ij} : B_t^{(ji)} = 0$. In Section D, we illustrate the results of the simulation exercises assessing the performance of the time-varying parameter approach.

A. Priors

The priors of the initial states of the time-varying parameters, $p(\theta_0)$, $p(\alpha_0)$, $p(\ln h_0)$, $p(\ln q_0)$, are assumed to be normally distributed, independent of each other, independent of $p(\lambda_0 \mid \nu)$ and independent of the hyperparameters, which are the elements of Z_η, Z_ω, S and ν. They are calibrated based on a constant-coefficient VAR(1), with Gaussian errors, estimated with a training sample of 36 monthly observations, over the period 1990-1994.

The prior for the initial states of the time-varying coefficients, $p(\theta_0)$, is,

$$\theta_0 \sim \mathcal{N}(\hat{\theta}_{OLS}, \hat{P}_{OLS}),$$

where $\hat{\theta}_{OLS}$ corresponds to the OLS estimates for the training sample and $\hat{P}_{OLS}$ to four times the covariance matrix $\hat{V}(\hat{\theta}_{OLS})$ of the OLS estimate.

For α_0 and $\ln h_0$ we follow Baumeister and Peersman (2013) and Primiceri (2005). Let $\hat{\Sigma}_{OLS}$ be the estimated covariance matrix of u_t from the time-invariant VAR, and let $C = AD^{1/2}$ be the Choleski factor of $\hat{\Sigma}_{OLS}$, where A is lower diagonal with ones across its diagonal and $D^{1/2}$ is a diagonal matrix. Then, we set

$$\ln h_0 \sim \mathcal{N}(\ln \mu_0, 10 \times I_N),$$

where μ_0 is the vector of diagonal elements of D and I_N is the identity matrix of size N.[1] Although the covariance matrix is chosen arbitrarily, it is set such that the prior is only

[1]For the analysis at the sectorial level, N is equal to the number of sectors, i.e. $N = 4$. For the analysis at the financial institution level, we estimate the connections pairwise using several bivariate TVP-VARs, therefore $N = 2$.

weakly informative.

The prior for the contemporaneous correlations is set to

$$\alpha_0 \sim \mathcal{N}\Big(\tilde{\alpha}_0, \tilde{V}(\tilde{\alpha})\Big),$$

where $\tilde{\alpha}_0$ is the lower diagonal elements of the inverse of A. The covariance matrix, $\tilde{V}(\tilde{\alpha})$, is assumed to be diagonal, and each diagonal element is set to ten times the absolute value of the corresponding element in $\tilde{\alpha}_0$. This is done in order to account for the magnitude of $\tilde{\alpha}_0$, whilst maintaining a very weakly informative prior (Benati and Mumtaz, 2007).

The prior for $\ln q_0$ is set following Baumeister and Benati (2013). Define $\tilde{Q}_0 = \gamma \times \hat{\Sigma}_{OLS}$, with $\gamma = 10^{-4}$, the same value used by Primiceri (2005) in the case of constant Q_t. Then, we set

$$\ln q_0 \sim \mathcal{N}(10^{-2} \times \ln \tilde{q}_0, 10 \times I_{N\cdot(1+N)})$$

where $\tilde{q}_0$ is a vector collecting the elements on the diagonal of $\tilde{Q}_0$.

Regarding λ_0, we follow Jacquier, Polson, and Rossi (2004) and use a conjugate inverse gamma prior,

$$\lambda_0 \mid \nu \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2}\right). \tag{A.1}$$

This means that $\lambda_0 \sim \nu/\chi^2(\nu)$.

For the hyperparameter ν, again we follow Jacquier et al. (2004) and use a discrete uniform prior on $[3, 40]$. A lower bound of 3 assures the existence of a conditional variance.

For computational convenience, we assume the matrices Z_ω and S to be independent. On the other hand, because errors ε_t and η_t are correlated row-by-row, the prior for Ω cannot be independent of Z_η. We follow Jacquier et al. (2004) by considering a reparameterisation of the two matrices.

Consider the submatrix V_i^* obtained by deleting all rows and columns of V except for the i^{th} and the $(i+N)^{\text{th}}$. Then,

$$V_i^* = \begin{bmatrix} 1 & \rho_i\sigma_i \\ \rho_i\sigma_i & \sigma_i^2 \end{bmatrix}.$$

As is done by Jacquier et al. (2004), we transform (ρ_i, σ_i) to (ψ_i, γ_i) as follows:

$$V_i^* = \begin{bmatrix} 1 & \psi_i \\ \psi_i & \gamma_i + \psi_i^2 \end{bmatrix}. \tag{A.2}$$

The transformation is motivated by observing that the volatility innovation η_{it} can be written as

$$\eta_{i\,t} = \ln h_{i\,t} - \ln h_{i\,t-1} = \sigma_i\rho_i\varepsilon_{i\,t} + \sigma_i\sqrt{1-\rho_i^2}\zeta_{i\,t} \qquad \text{with } (\zeta_{i\,t}, \varepsilon_{i\,t}) \sim N(0, I_2).$$

That is, $\psi_i = \sigma_i\rho_i$ can be interpreted as the coefficient in a regression of $\varepsilon_{i\,t}$ on $\eta_{i\,t}$ with error variance $\gamma_i = \sigma_i^2(1-\rho_i)^2$. We use a normal prior for ψ_i and an inverse gamma for γ_i, setting parameters as done by Jacquier et al. (2004). Thus,

$$\psi \mid \gamma_i \sim \mathcal{N}\left(\psi_0\,, \frac{\gamma_i}{p_0}\right),$$

and,

$$\gamma_i \sim \mathcal{IG}(d_0 t_0 = 10^{-4}\,, d_0 = 1),$$

where we set $\psi_0 = 0$ and $p_0 = 2$.

The prior on $(\psi_i\,, \gamma_i)$ induces a prior distribution over $(\rho_i\,, \sigma_i)$. This distribution is diffuse on ρ_i, while ruling out very large correlations. The marginal prior on σ_i is very similar to that used in the basic model with no leverage of Jacquier, Polson, and Rossi (1994).

We use an inverse-Gamma prior for the elements of Z_ω,

$$\sigma_{\omega,i}^2 \sim \mathcal{IG}\left(\frac{10^{-4}}{2}, \frac{10}{2}\right), \qquad \forall i = 1, \ldots, N.$$

The prior has the same mean that was used by Cogley, Primiceri, and Sargent (2010), but it has a smaller variance, analogous to the one used by Baumeister and Benati (2013).

The prior for the different blocks of S are set as follows. In the TVP-VAR including the four sector indices, S is composed of three blocks each assumed to follow an inverted Wishart, with prior degrees of freedom set to the minimum allowed:

$$S_k \sim \mathcal{IW}(\bar{S}_k^{-1}, k+1),$$

where $k = 1, 2, 3$. The scale matrices, $\bar{S}_k$, are diagonal, with diagonal elements set equivalent to 10^{-4} times the absolute value of the relevant diagonal blocks of $\tilde{V}(\tilde{\alpha}_0)$.

In the case of the bivariate TVP-VAR used to estimate pairwise connections between systemically important financial institutions, S has only one block and is a scalar. Then,

$$S \sim \mathcal{IG}(\bar{S}^{-1}, 2),$$

with $\bar{S} = 10^{-3} \times \mid \tilde{\alpha}_0 \mid$.

B. Posterior distribution sampling

We simulate the posterior distribution of the states and the hyperparameters *via* the following MCMC algorithm. In what follows, x^t denotes the entire history of the vector x up to time t–i.e. $x^t \equiv [x_1', x_2', \ldots, x_t']'$– while T is the sample length.

B.1. Drawing the parameter states, $\theta^{\mathbf{T}}$

The conditional distribution of the TVP-VAR parameters, θ^T, can be expressed as:

$$\begin{aligned} & p(\theta^T \mid R^T, \lambda^T, \alpha^T, h^T, q^T, \nu, V) = \\ \text{(A.3)} \qquad & p(\theta_T \mid R^T, \lambda^T, \alpha^T, h^T, q^T, \nu, V) \prod_{t=1}^{T-1} p(\theta_t \mid \theta_{t+1}, R^T, \lambda^T, \alpha^T, h^T, q^T, \nu, V) \end{aligned}$$

Given the prior assumptions above and the state-space model, the conditional densities are normal and can be simulated using the algorithm proposed by Carter and Kohn (1994). We can compute their means and variances through the forward and backward recursions of the Kalman filter and smoother. The last iteration of the filter provides the mean and variance for the first term on the right hand side,

$$p(\theta_T \mid R^T, \lambda^T, \alpha^T, h^T, q^T, \nu, V) = \mathcal{N}(\theta_{T|T}, P_{T|T})$$

A draw from the distribution is used in the backward recursions to simulate the remaining terms in equation (A.3). Conditional on the information in θ_{t+1}, θ_t is conditionally normal with mean and variance given respectively by,

$$\begin{aligned}\theta_{t|t+1} &= \theta_{t|t} + P_{t|t}P_{t+1|t}^{-1}(\theta_{t+1} - \theta_{t|t}),\\ P_{t|t+1} &= P_{t|t} - P_{t|t}P_{t+1|t}^{-1}P_{t|t}\end{aligned}$$

The backward recursions draw sequentially $\theta_{T-1}, \theta_{T-2}, \ldots, \theta_1$ from the conditional distribution,

$$p(\theta_t \mid \theta_{t+1}, R^T, \lambda^T, \alpha^T, h^T, q^T, \nu, V) = \mathcal{N}(\theta_{t|t+1}, P_{t|t+1}),$$

in order to generate a random trajectory, θ^T.

Some studies from the macro literature choose to impose a stability condition so as to exclude explosive paths for B_t in equation (1). This is done by assuming that the probability density of B_t takes a value of zero when the roots of the TVP-VAR polynomial are inside the unit circle. Others, such as Primiceri (2005), do not include this condition, because they assume that the model holds for a finite period of time and not forever. Given that we expect a VAR model on stock returns to have small coefficients (in absolute terms), we follow Primiceri (2005) and do not impose a stability condition.

B.2. Drawing the contemporaneous interactions, $\alpha^{\mathbf{T}}$

Given the data R^T and draws of θ^T, h^T, and λ_t, we can recover $u_t = R_t - X_t'\theta_t$ from Equation (1) and write $A_t u_t/\sqrt{\lambda_t} \equiv A_t\tilde{u}_t = \varepsilon_t^*$, where ε_t^* is the vector of orthogonalised, normally distributed, innovations with known time-varying variance, $H_t = \mathrm{diag}(h_t)$.

From this, a system of unrelated regressions can be estimated to recover A^T according to the following transformed equations:

$$\tilde{u}_{1t} = \varepsilon^*_{1t}$$
$$\tilde{u}_{2t} = -\alpha_{21t}\tilde{u}_{1t} + \varepsilon^*_{2t}$$
$$\vdots$$
$$\tilde{u}_{Nt} = -\alpha_{N\,1,t}\tilde{u}_{1t} - \alpha_{N2,t}\tilde{u}_{2\,t} - \cdots - \alpha_{N\,(N-1),t}\tilde{u}_{N-1\,t} + \varepsilon^*_{N\,t}$$

The coefficients of α_t are drawn using the system above and the Kalman filter and smoother equations explained in previous step.[2]

B.3. Drawing the stochastic volatilities, $\mathbf{h^T}$

Since the stochastic volatilities $h_{i\,t}$ and $h_{j\,s}$ are independent for all $i \neq j$ and t, s, we draw them on a univariate basis for each financial institution $i = 1, \ldots, N$. To do this, we adopt a modified version of the univariate algorithm by Jacquier et al. (1994), developed by Cogley, Morozov, and Sargent (2005), and combine it with elements from Jacquier et al. (2004) in order to account for the leverage effect ρ.

It follows that knowledge of $h_{i\,t-1}$, $h_{i\,t+1}$, ψ_i, γ_i and of the orthogonalised residuals ε^*_{it} (which we recover from R^T, θ^T and α^T) are sufficient statistics for $h_{i\,t}$. It follows that

$$p(h_{it} \mid h_{i\,t-1}, h_{i\,t+1}, \psi_i, \gamma_i, \varepsilon^*_{it}) \propto h_{it}^{-\left(\frac{3}{2} + \frac{\psi_i \varepsilon^*_{it+1}}{\gamma_i \sqrt{h_{i\,t+1}}}\right)} \exp\left[\frac{-\varepsilon^*_{it}}{2h_{it}}\left(1 + \frac{\psi_i^2}{\gamma_i}\right) - \frac{(\ln h_{it} - \mu_{it})^2}{\gamma_i} + \frac{\psi_i \varepsilon^*_{it}\eta_{it}}{\gamma_i \sqrt{h_{it}}}\right],$$

where $\mu_{it} = (\ln h_{i\,t+1} - h_{i\,t-1})/2$.

The non-standard form of the posterior does not allow direct sampling. Instead, we apply the Metropolis accept/reject sampler developed by Cogley et al. (2005) with a log-normal

[2] For the analysis at the sectorial level, N is equal to the number of sectors, i.e. $N = 4$. For the analysis at the financial institution level, we estimate connections pairwise using bivariate TVP-VARs, therefore $N = 2$.

proposal density $\hat{f}$ defined as

$$\hat{f}(h_{i\,t}) \propto h_{i\,t}^{-1} \exp\left[-\frac{(\ln h_{it} - \mu_{i\,t})^2}{2\sigma_i}\right].$$

where σ_i is recovered from drawing ψ_i and γ_i as shown below in Section B.7 of this appendix.

B.4. Drawing the stochastic volatilities of the states, $\mathbf{q^T}$

Given a draw of θ^T, we can recover the vector of innovation of the states $v_t = \theta_t - \theta_{t-1}$. Given that its variance, $V(\omega_t) = Q_t$, is diagonal, we draw the diagonal elements q_t one by one as done for h_t above. Again we apply the univariate algorithm of Jacquier et al. (1994) with the log-normal proposal density of Cogley et al. (2005).

B.5. Drawing the latent variable, $\lambda_\mathbf{t}$

Notice that, conditional on λ_t, A_t and H_t, the errors $u_t = (R_t - X_t'\theta_t)$ are normal with variance-covariance matrix $\Sigma_t \equiv A_t^{-1} H_t (A_t^{-1})'$.

Then, notice that $p(\lambda^T \mid u^T, \Sigma^T, \nu) = \prod_{t=1}^T p(\lambda_t \mid u_t, \Sigma_t, \nu)$. It follows that

$$p(\lambda_t \mid u_t, \Sigma_t, \nu) \propto p(u_t \mid \lambda_t, \Sigma_t, \nu) p(\lambda_t \mid \nu).$$

Using our conjugate prior in equation (A.1), we have that

$$p(\lambda_t \mid u_t, \Sigma_t, \nu) \sim \mathcal{IG}\left(\frac{u_t'\,\Sigma_t^{-1}\,u_t + \nu}{2}, \frac{\nu + N}{2}\right).$$

B.6. Drawing the hyperparameter, ν

Given A_t, H_t, and ν, the errors u_t are distributed as a multivariate $t(\nu)$ distribution with scale matrix Σ_t. Then, ν is discrete with probability mass proportional to the product of t distribution ordinates:

$$p(\nu \mid \Sigma_t, u_t) \propto p(\nu) p(u_t \mid \Sigma_t, \nu) = p(\nu) \prod_{t=1}^{T} \frac{\Gamma\big[(\nu + N)/2\big]}{\Gamma(\nu/2)\,\nu^{N/2}\,\pi^{N/2}} \left[1 + \frac{1}{\nu}\,u_t'\,\Sigma_t^{-1}\,u_t\right]^{-(\nu+N)/2}.$$

B.7. Drawing the hyperparameters, $\mathbf{\Omega}$ and $\mathbf{Z}_\eta$

Since ρ_i and ρ_j are independent for all $i \neq j$, we proceed by drawing each leverage effect one at a time, following Jacquier et al. (2004).

Given draws of θ^T, α^T, h^T and λ_t we can recover the orthonormal vector of innovations, $\varepsilon_t = A_t H_t^{-1/2} u_t / \lambda_t$. Moreover, given h^T we can recover the vector of innovations from the stochastic volatility equations, $\eta_t = \ln h_t - \ln h_{t-1}$.

Then, let $w_{i\,t} = [\varepsilon_{i\,t}, \eta_{i\,t}]'$ be the vector of innovations and $W_i = \sum_t w_{i\,t} w_{i\,t}'$. Given the re-parametrization of (ρ_i, σ_i) to (ψ_i, γ_i), the conditional posteriors follow by conjugacy of the priors:

$$
\begin{aligned}
p(\psi_i \mid \gamma_i, h_i^T, \varepsilon_i^T) &\sim \mathcal{N}\big(\tilde{\psi}_i\,, \gamma_i/(W_i^{(1\,1)} + p_0)\big) \\
p(\gamma_i \mid h_i^T, \varepsilon_i^T) &\sim \mathcal{IG}\big(v_0 t_0^2 + W_i^{(22).1}\,, v_0 + T - 1\big)
\end{aligned}
$$

with $\tilde{\psi} = \big(W_i^{(1\,2)} + p_0 \psi_0\big)/\big(W_i^{(1\,1)} + p_0\big)$ where $W_i^{(k\,l)}$ denotes the (k,l) element of W_i and $W_i^{(22).1} = W_i^{(2\,2)} - \big(W_i^{(1\,2)}\big)^2/W_i^{(1\,1)}$.

A draw of (ψ_i, γ_i) yields a draw of (ρ_i, σ_i) by $\sigma_i^2 = \psi_i^2 + \gamma_i$ and $\rho_i = \psi_i/\sigma_i$.

B.8. Drawing the hyperparameters, $\mathbf{Z}_\omega$ and $\mathbf{S}$

Given draws of q^T and α^T, we can observe the innovations $\omega_t = \ln q_t - \ln q_{t-1}$ and $\tau_t = \alpha_t - \alpha_{t-1}$. Following conjugacy of the priors, we can draw the elements of Z_ω and the elements of the blocks of S from their respective conditional posterior distributions.

For the empirical investigation at the sector level, we perform 30,000 iterations of the Gibbs sampler and discard the first 10,000 draws. We then keep only the 5th of every draw in order to mitigate autocorrelation among draws. The remaining sequence of 4,000 draws forms a sample of the joint posterior distribution. We use this to estimate the parameters and compute Bayes factor.

Similarly, for the empirical investigation at the financial institution level, for each bivariate TVP-VAR estimated, we perform 6,000 iterations of the Gibbs sampler and discard the first 1,000 draws. Again, we keep only the 5th of every draw. We used an analogous number of draws for the simulation study in Appendix D.

C. Bayesian inference

Denote by Ψ^T all parameters except the states θ^T, i.e., α^T, H^T, Q^T and given hyperparameters governing the priors. Then under the assumption that

$$p(\Psi^T \mid B_t^{(j\,i)} = 0) = p_0(\Psi^T), \tag{A.4}$$

Bayes factor will be given by the SDDR,

$$K_t^{(i\,j)} = \frac{p(B_t^{(j\,i)} = 0 \mid R^T)}{p(B_t^{(j\,i)} = 0)}. \tag{A.5}$$

The assumption given by equation (A.4) requires the prior for Ψ^T in the restricted model, $p_0(\Psi^T)$, to be the same as the prior in the unrestricted model evaluated at the point where the restriction holds, $p(\Psi^T \mid B_t^{(j\,i)} = 0)$. This is amply satisfied if the same prior is used in the restricted and unrestricted model for the parameters that are common to both models, as we do here.

As explained by Koop, Leon-Gonzalez, and Strachan (2010), an estimate of the numerator in equation (A.5) can be calculated using the simulations from the conditional posterior $p(\theta^T \mid R^T, \Psi^T)$. Given our conjugate Normal conditional prior for θ_t, the conditional posterior distribution $p(B_t^{(j\,i)} = 0 \mid R^T, \Psi^T)$ is Normal. By simulating from $p(B_t^{(j\,i)} = 0 \mid R^T, \Psi^T)$ using a Gibbs sampler and averaging across draws, we obtain an estimate of the posterior probability that the null hypothesis holds, $\hat{p}(B_t^{(j\,i)} = 0 \mid R^T)$. Similarly, the denominator can be simulated by using a sequential sampler on the conditional priors $p(B_t^{(j\,i)} = 0 \mid \Psi^T)$, and calculating the average across all draws, $\hat{p}(B_t^{(j\,i)} = 0)$.

D. Simulation study

In a series of simulation exercises, we assessed the ability of our time-varying framework to infer the small causal network given in Figure A1. A similar exercise was conducted by Seth (2010) using the same network. We chose this particular network because it is sparse and sparsity is an observed attribute of financial networks.[3]

The network's underlying system is given by

$$
\begin{aligned}
x_{1t} &= \alpha_{1t} + \phi_{1t}\, x_{1\,t-1} + \epsilon_{1t} \\
x_{2t} &= \alpha_{2t} + \phi_{2t}\, x_{2\,t-1} + \beta_{21\,t}\, x_{1\,t-1} + \epsilon_{2t} \\
x_{3t} &= \alpha_{3t} + \phi_{3t}\, x_{3\,t-1} + \beta_{31\,t}\, x_{1\,t-1} + \epsilon_{3t} \\
x_{4t} &= \alpha_{4t} + \phi_{4t}\, x_{4\,t-1} + \beta_{41\,t}\, x_{1\,t-1} + \beta_{45\,t}\, x_{5\,t-1} + \epsilon_{4t} \\
x_{5t} &= \alpha_{5t} + \phi_{5t}\, x_{5\,t-1} + \beta_{54\,t}\, x_{4\,t-1} + \epsilon_{5t}
\end{aligned}
$$

where, $[\epsilon_{1t} \dots \epsilon_{5t}]' = \epsilon_t \sim \mathcal{N}(\mathbf{0}, \Omega)$ and $\Omega = \tau\, I_5$ where τ was set to 0.01. We chose to limit the autoregressive component of the process to one lag, as is done by Barigozzi and Brownlees (2017), so as to keep the simulation exercises computationally manageable.

We performed three different experiments in which the model VAR parameters were allowed to vary according to the following processes:

1. Deterministic fixed constants drawn, at the beginning of each simulation, from a standard uniform distribution.

2. Markov switching between 0 and a random constant drawn, at the beginning of each simulation, from a standard uniform distribution.

3. Smoothly time-varying, according to a unit root process.

For each experiment, we ran 100 simulations each of which involved $T = 300$ time periods.

In this straightforward exercise, we applied our framework without leverage effects ($\rho_i = 0, \forall i$), without heavy tailed errors in the TVP-VAR ($\lambda_t = 1, \forall t$), and with constant time-varying variance-covariance matrix ($\Sigma_t = \Sigma, \forall t$). Moreover, we assume constant variance for the VAR parameters ($Q_t = Q, \forall t$).

We used the framework to infer all possible connections between variables. Paralleling pairwise and conditional Granger causality, this was done in two alternative ways: 1) by

[3] The interested reader may refer to Barigozzi and Brownlees (2017) for an in-depth discussion on sparse networks in finance.

recursively using a bivariate TVP-VAR between every pair of variables, and 2) by running the TVP-VAR on all five variables and inferring connections conditional on the system.

For means of comparison, we also carried out the same simulation exercises, using the classical approach of Granger causality testing (by pairwise and conditional VARs) over rolling windows. For this, we set the level of significance of the tests to 5%.

We assessed the performance of our framework with respect to three standard measures: the mean-squared error (MSE) of the VAR parameter estimates, the receiver-operator characteristic (ROC) curve and the precision-recall (PR) curve. We outline how each measure is computed below.

We compute the MSE of the VAR parameter estimates by taking the sum, across all time periods, of the squared difference between the estimated VAR parameters and the true VAR parameters. This sum is then averaged across all simulations. The formula for the MSE of the cross-parameters $\beta_{i,j,t}$ is given by

$$MSE_C^{TVP} = \sum_{ij \in C} \sum_{t=1}^{T} (\hat{\beta}_{ij\,t}^{TVP} - \beta_{ij\,t})^2 / T,$$

where $C = \{(2,1),(3,4),(3,5),(4,1),(4,5),(5,4)\}$.

For the classical Granger causality approach, parameters are estimated by ordinary least squares (OLS) over rolling windows of size $w = [20, 30, \ldots, 200]'$. Then the MSE is calculated as

$$MSE_{C,w(s)}^{RW} = \sum_{ij \in C} \sum_{t=w(s)+1}^{T} (\hat{\beta}_{ij\,t}^{RW} - \beta_{ij\,t})^2 / \big(T - w(s)\big).$$

The step size for the rolling window calculation is set to 1.

To allow a fairer comparison between MSE^{RW} and the MSE_C^{TVP} across the same time periods, we look at

$$MSE_{C,w(s)}^{TVP} = \sum_{ij \in C} \sum_{t=w(s)+1}^{T} (\hat{\beta}_{ij\,t}^{TVP} - \beta_{ij\,t})^2 / \big(T - w(s)\big),$$

We also compared the performance of our time-varying parameter framework with that of the classical Granger causality approach, by means of the ROC and PR curves.

The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR). In our case, a positive refers to the existence of a connection between the two nodes in question. Then the TPR is the ratio of the number of correctly inferred connections to the number of existing connections. On the other hand, the FPR is the ratio of incorrectly inferred connections to the number of non-existing connections. A high performing test would combine low FPR with high TPR and therefore have a ROC curve in the upper-left corner of the chart.

For time-varying parameter estimation, the ROC curve was calculated using the implied probability from the estimated Bayes factor ($\hat{K}_t^{(ij)}/(1 + \hat{K}_t^{(ij)})$), whereas for the classical Granger causality approach, the p-value was used. All possible pairs, ij, $i \neq j$, were tested and results were aggregated over all time periods and across all simulations.

The PR curve plots the precision, also known as the positive predictive value, against the recall, i.e. the TPR. The precision is the fraction of correctly classified positives, i.e. the ratio of connections correctly inferred to the total number of connections inferred. There exists a one-to-one relationship between the ROC and precision-recall curve. If for a given experiment a curve dominates in ROC space, then it will also dominate in precision-recall space (Davis and Goadrich (2006)). However, looking at the PR curve can provide additional insight in situations like ours, where the number of negatives exceeds by far the number of positives. A high performing test would combine high precision with high TPR and therefore have a PR curve in the upper-right corner of the chart.

Experiment 1: time-invariant connections

For the first experiment, we fix all VAR parameters to constants drawn at the beginning of each simulation.

$$\alpha_{i,t} = a_i, \qquad \phi_{i,t} = f_i, \qquad \beta_{i,j,t} = b_{ij}, \qquad \forall t \in [0, T],$$

where we draw parameters from a standard uniform distribution at the beginning of each simulation, $a_i, f_i, b_{ij} \sim \mathcal{U}(0,1)^3$ for $i = 1, \ldots, 5$ and $(i,j) \in C$.

The left panel of Figure A2 shows $MSE_{w(s)}^{RW}$ (light dashed) and $MSE_{w(s)}^{TVP}$ (bold solid).

Notice that $MSE^{RW}_{w(s)}$ is downward sloping in window size. This is because larger windows lead to more precise estimates at the expense of less variability. Since the underlying parameters are constant, $MSE^{RW}_{w(s)}$ decreases quickly with the window size. $MSE^{TVP}_{w(s)}$ is not downward sloping because, unlike the rolling window approach, the time-varying parameter framework uses the whole length of the sample for estimation.

Results show that the time-varying framework performs better than the classical rolling window approach, whether estimation is pairwise (top-left chart) or conditional on the other variables of the system (bottom-left chart). The time-varying parameter framework does well because the Kalman filter and smoother, used for the sampling algorithms, find the best fit with the minimum predictive variance. Even when large rolling windows are used (above 100 observations) the time-varying parameter framework performs comparably well to the classical approach.

We report the performance of our time-varying parameter framework in terms of ROC and PR curves, respectively given in the middle and right panels of Figure A2 (bold solid). We also show the ROC and PR curves associated with the classical Granger causality approach (light dashed) estimated by rolling windows of size 200.[4] This corresponds to two-thirds of the observations in each simulation. It was also one of the best performing window sizes across all three experiments.

The ROC curve for pairwise estimation (top-middle chart) shows that time-varying parameter inference performs comparably well compared to the classical approach with rolling windows. In particular, it does slightly better than the classical approach at low combinations of FPR and TPR. We obtain similar results when testing conditional relationships (bottom-middle chart).

In terms of the PR curve, pairwise time-varying inference does well at combinations with high precision and low recall (upper-right chart). Here the curve associated with time-varying parameters (bold solid) is above that associated with the classical rolling windows approach (light dashed). Similar results were found for conditional testing (lower-right chart). However, at higher combinations of precision and recall, the two approaches perform similarly,

[4] ROC and PR curves calculated at other window sizes have been omitted for space concerns but are available from the authors upon request.

with the PR curve for the classical approach slightly above the time-varying counterpart.

Experiment 2: discretely time-varying connections

For the second experiment, the cross coefficients, $\beta_{ij\,t}$ with subscripts $ij \in C$, of the system were assumed to follow a switching process defined as

$$\beta_{ij\,t} = \begin{cases} 0 & \text{if } s_t^{ij} = 0 \\ b_{ij} & \text{if } s_t^{ij} = 1 \end{cases}$$

where b_{ij} is drawn at the start of the simulation from a standard uniform distribution.

As in the first experiment, the intercept terms α_{it} and autoregressive coefficients ϕ_{it} were drawn from a standard uniform distribution at the beginning of each simulation and were assumed to be constant through time.

Let s_t^{ij} follow a first order Markov chain with the following transition matrix:

$$\mathbf{P} = \begin{bmatrix} \mathbb{P}(s_t^{ij} = 0 \mid s_{t-1}^{ij} = 0) & \mathbb{P}(s_t^{ij} = 1 \mid s_{t-1}^{ij} = 0) \\ \mathbb{P}(s_t^{ij} = 0 \mid s_{t-1}^{ij} = 1) & \mathbb{P}(s_t^{ij} = 1 \mid s_{t-1}^{ij} = 1) \end{bmatrix} = \begin{bmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{bmatrix}$$

where we set $p_{00} = 0.95$ and $p_{11} = 0.90$.

Effectively, the transition matrix holds the probabilities of a link appearing and disappearing between any two nodes $ij \in C$. In the matrix, p_{00} is the probability of no link occurring between two nodes at time t, given that the two nodes were disconnected at time $t-1$. Similarly, p_{11} represents the probability of there being a link between two nodes at t, given that these two nodes were already connected at $t-1$.

The top-left and bottom-left charts of Figure A3 show the MSE for estimates found by, respectively, pairwise estimation and full conditional estimation. When using pairwise testing (top-left chart), results show that estimation using the time-varying parameter framework is more precise than the classical approach only for small window sizes. When the window size increases, the precision of the rolling window estimates increases. On the other hand, with conditional testing (bottom-left chart), the time-varying parameter framework is more precise than the classical approach for all window sizes. This confirms the Monte Carlo simulation results of Baumeister and Peersman (2013).

The ROC curves in the top-middle and bottom-middle charts of Figure A3 highlight the gain obtained by using the time-varying parameter framework for inferring connections. For pairwise inference, the ROC curve lies mostly above the corresponding curve for classical Granger causality testing with rolling windows of size 200. On the other hand, when inferring connections conditionally, classical Granger causality testing performs superiorly to the TVP-VAR.

Similarly, for pairwise inference, the PR curves (top-right and bottom-right charts of Figure A3) show substantial improvements using the time-varying parameter framework. In the case of conditional testing, the PR curve associated with the time-varying parameter framework appears above the corresponding curve for classical testing in areas of the chart with low recall, while it lies slightly below for areas with higher recall. This indicates that, in this case, our framework performs better when the network is sparse.

Experiment 3: smoothly time-varying connections

For the third experiment, the parameters of the system were allowed to evolve according to the following random walk processes,

$$\begin{aligned} \alpha_{i\,t+1} &= \alpha_{it} + v^{\alpha}_{i\,t+1} \\ \phi_{i\,t+1} &= \phi_{it} + v^{\phi}_{i\,t+1} \\ \beta_{ij\,t+1} &= \beta_{ij\,t} + v^{\beta}_{i\,t+1}, \end{aligned}$$

where, $v_{it} = [v^{\alpha}_{it}\,,\, v^{\phi}_{it}\,,\, v^{\beta}_{it}]'$ and $v_{it} \sim \mathcal{N}(\mathbf{0}, \Gamma)$. In turn, the variance of the parameters was set to

$$\Gamma = q^2 \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

where, $q^2 = 0.0002$.

The variance of the parameters was set such that cross coefficients would be more variable than the autoregressive parameters and, in turn, the autoregressive parameters would be more variable than the intercept terms.

As can be noticed from the top-left and bottom-left charts in Figure A4, the time-varying parameters approach results in more precise parameter estimates, in terms of MSE, than the classical approach using rolling windows. Notice that, quite contrary to what was found in Experiment 1, the precision of the classical approach does not increase with window size. Rather, gains from using larger windows reverse for windows of more than 60 observations, showing that precision actually worsens when windows are too large. This highlights the trade-off between higher confidence but less flexibility given by larger windows.

In terms of ROC curves, shown in the top-middle and bottom-middle charts, the curve referring to the time-varying parameters approach lies completely above the one referring to classical Granger causality testing by rolling windows of size 200.[5]

The same result was obtained for the PR curves shown in the top-right and bottom-right charts. Here we notice that there is not much gain from using pairwise inference with recursive bivariate TVP-VARs rather than conditional inference with the full TVP-VAR. On the other hand, with the classical approach we see that using conditional testing leads to a higher PR curve. However, this curve continues to remain below the curve associated with the time-varying parameter framework meaning that our framework is better at inferring connections across all combinations of precision and recall.

These results, together with those from Experiment 1 and 2, show that our proposed framework provides more precise estimates of connections when the true underlying process is either constant or smoothly changing through time. When changes are abrupt and connections are pairwise estimated, our framework still performs well in terms of detecting connections, but delivers less precise connection-strength estimates than the classical approach. However, when connections are estimated conditionally, our framework performs superiorly in terms of precision, and comparably well in terms of detection, even with abrupt changes in the underlying process.

[5] The results continue to hold for other window sizes.

Tables

Table A1: Sample of Financial Institutions.

Table A1 presents the financial institutions used for the empirical investigation.

Banks		Insurers	Real Estate
AHMANSON (H F) & CO	NORTHERN TRUST CORP	ACE LTD	AMERICAN CAPITAL LTD
AMERICAN EXPRESS CO	OLD KENT FINANCIAL CORP	AETNA INC	AMERICAN TOWER CORP
AMSOUTH BANCORPORATION	PEOPLE'S UNITED FINL INC	AFLAC INC	APARTMENT INVST & MGMT CO
ASSOCIATES FIRST CAP -CL A	PNC FINANCIAL SVCS GROUP INC	ALEXANDER & ALEXANDER	AVALONBAY COMMUNITIES INC
BANK OF AMERICA CORP	PROVIDIAN FINANCIAL CORP	ALLSTATE CORP	BOSTON PROPERTIES INC
BANK OF NEW YORK MELLON CORP	REGIONS FINANCIAL CORP	AMERICAN INTERNATIONAL GROUP	CBRE GROUP INC
BANK ONE CORP	SHAWMUT NATIONAL CORP	AON PLC	CROWN CASTLE INTL CORP
BANKBOSTON CORP	SLM CORP	ASSURANT INC	DDR CORP
BARNETT BANKS INC	STATE STREET CORP	CHUBB CORP	EQUITY OFFICE PROPERTIES TR
BB&T CORP	SUMMIT BANCORP	CIGNA CORP	EQUITY RESIDENTIAL
BENEFICIAL CORP	SUNTRUST BANKS INC	CINCINNATI FINANCIAL CORP	GENERAL GROWTH PPTYS INC
BOATMENS BANCSHARES INC	SYNOVUS FINANCIAL CORP	CNA FINANCIAL CORP	HCP INC
CAPITAL ONE FINANCIAL CORP	U S BANCORP	CNO FINANCIAL GROUP INC	HEALTH CARE REIT INC
CHASE MANHATTAN CORP -OLD	VISA INC	CONTINENTAL CORP	HFS INC
CIT GROUP INC	WACHOVIA CORP	COVENTRY HEALTH CARE INC	HOST HOTELS & RESORTS INC
CIT GROUP INC-OLD	WACHOVIA CORP-OLD	EXPRESS SCRIPTS HOLDING CO	KIMCO REALTY CORP
CITICORP	WASHINGTON MUTUAL INC	GENWORTH FINANCIAL INC	MACERICH CO
CITIGROUP INC	WELLS FARGO & CO	HANCOCK JOHN FINL SVCS INC	PROLOGIS INC
COMERICA INC	WELLS FARGO & CO -OLD	HARTFORD FINANCIAL SERVICES	PUBLIC STORAGE
COMMERCE BANCORP INC/NJ	WESTERN UNION CO	HUMANA INC	SIMON PROPERTY GROUP INC
COMPASS BANCSHARES INC	ZIONS BANCORPORATION	JEFFERSON-PILOT CORP	VENTAS INC
CONCORD EFS INC		LINCOLN NATIONAL CORP	VORNADO REALTY TRUST
CORESTATES FINANCIAL CORP		LOEWS CORP	WYNDHAM WORLDWIDE CORP
COUNTRYWIDE FINANCIAL CORP	**Brokers**	MCLENNAN COS	
DISCOVER FINANCIAL SVCS INC	AMERIPRISE FINANCIAL INC	MBIA INC	
FANNIE MAE	BEAR STEARNS COMPANIES INC	METLIFE INC	
FEDERAL HOME LOAN MORTG CORP	BLACKROCK INC	MGIC INVESTMENT CORP/WI	
FIFTH THIRD BANCORP	CME GROUP INC	PROGRESSIVE CORP-OHIO	
FIRST CHICAGO NBD CORP	E TRADE FINANCIAL CORP	PROVIDENT COS INC	
FIRST FIDELITY BANCORP	FEDERATED INVESTORS INC	PRUDENTIAL FINANCIAL INC	
FIRST HORIZON NATIONAL CORP	FRANKLIN RESOURCES INC	SAFECO CORP	
FIRST INTERSTATE BNCP	GOLDMAN SACHS GROUP INC	TORCHMARK CORP	
FLEETBOSTON FINANCIAL CORP	INTERCONTINENTALEXCHANGE GRP	TRAVELERS COS INC	
GOLDEN WEST FINANCIAL CORP	INVESCO LTD	U S HEALTHCARE INC	
GREAT WESTERN FINANCIAL	JANUS CAPITAL GROUP INC	UNITEDHEALTH GROUP INC	
HSBC FINANCE CORP	LEGG MASON INC	UNUM GROUP	
HUDSON CITY BANCORP INC	LEHMAN BROTHERS HOLDINGS INC	USF&G CORP	
HUNTINGTON BANCSHARES	MERRILL LYNCH & CO INC	USLIFE CORP	
JPMORGAN CHASE & CO	MORGAN STANLEY	WELLPOINT HEALTH NETWRKS INC	
KEYCORP	NASDAQ OMX GROUP INC	XL GROUP PLC	
M & T BANK CORP	NYSE EURONEXT		
MARSHALL & ILSLEY CORP	PAINE WEBBER GROUP		
MASTERCARD INC	PRICE (T. ROWE) GROUP		
MBNA CORP	PRINCIPAL FINANCIAL GRP INC		
MELLON FINANCIAL CORP	SCHWAB (CHARLES) CORP		
MERCANTILE BANCORPORATION			
MONEYGRAM INTERNATIONAL INC			
NATIONAL CITY CORP			
NORTH FORK BANCORPORATION			

Figures

Figure A1: The Causal Network

Figure A1 shows the causal network used for the simulation study.

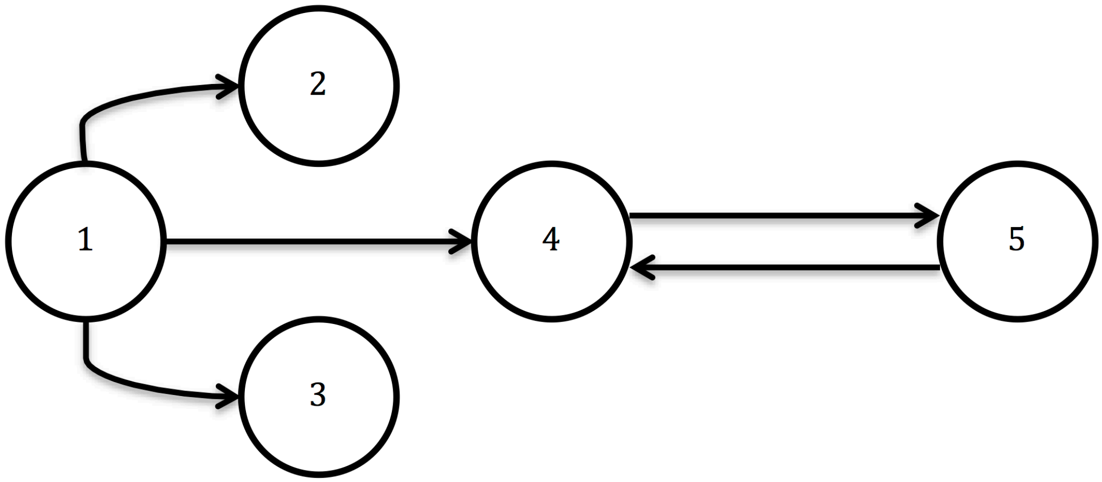


Figure A2: Results for Experiment 1

Figure A2 shows the results of Experiment 1, for which the underlying parameters are constant. The light dashed line relates to results obtained using classical Granger causality testing over rolling windows, while the bold solid line relates to results obtained using the proposed time-varying parameter framework. The upper panel shows results for pairwise estimation and inference. The lower panel shows results using conditional estimation and inference. The left column figures show the mean squared error of cross-parameter estimates. The middle and right column figures show the ROC curves and the PR curves for inference of the underlying network, with rolling window size set at 200 observations for the rolling window approach.

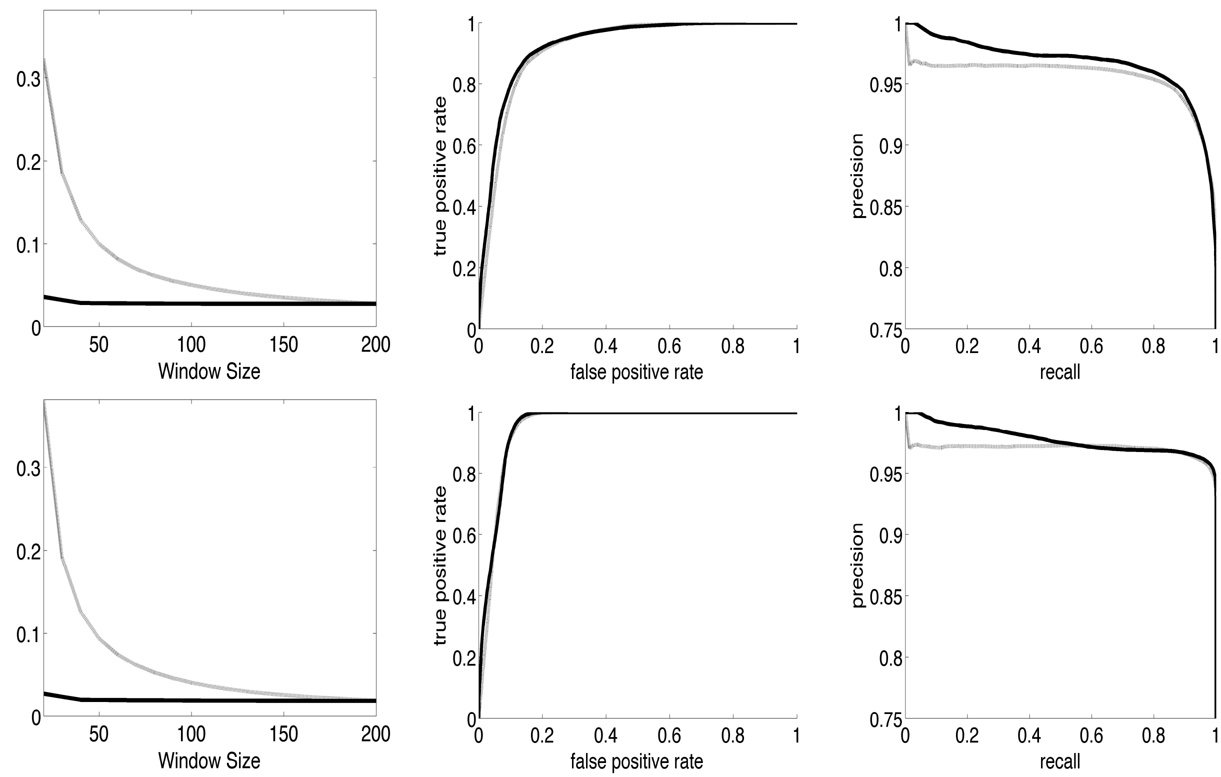


Figure A3: Results for Experiment 2

Figure A3 shows the results of Experiment 2, for which the underlying parameters follow a regime switching process. The light dashed line relates to results obtained using classical Granger causality testing over rolling windows, while the bold solid line relates to results obtained using the proposed time-varying parameter framework. The upper panel shows results for pairwise estimation and inference. The lower panel shows results using conditional estimation and inference. The left column figures show the mean squared error of cross-parameter estimates. The middle and right column figures show the ROC curves and the PR curves for inference of the underlying network, with rolling window size set at 200 observations for the rolling window approach.

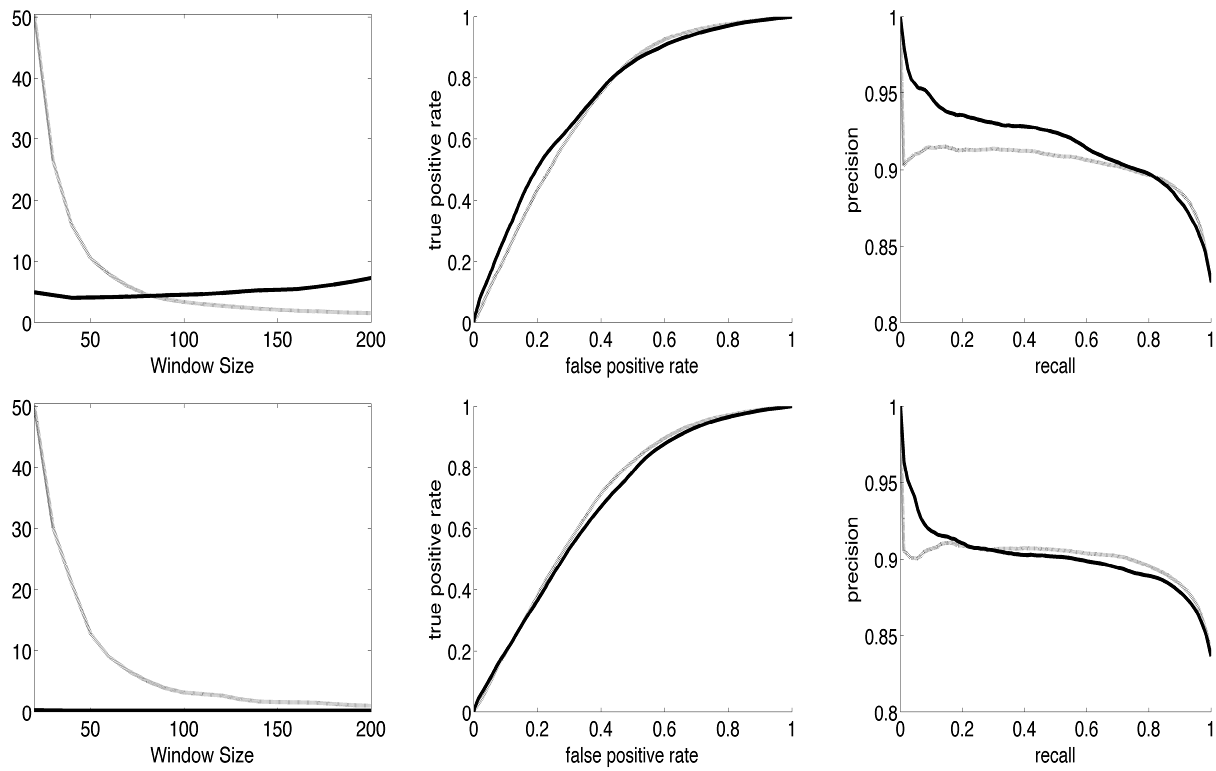
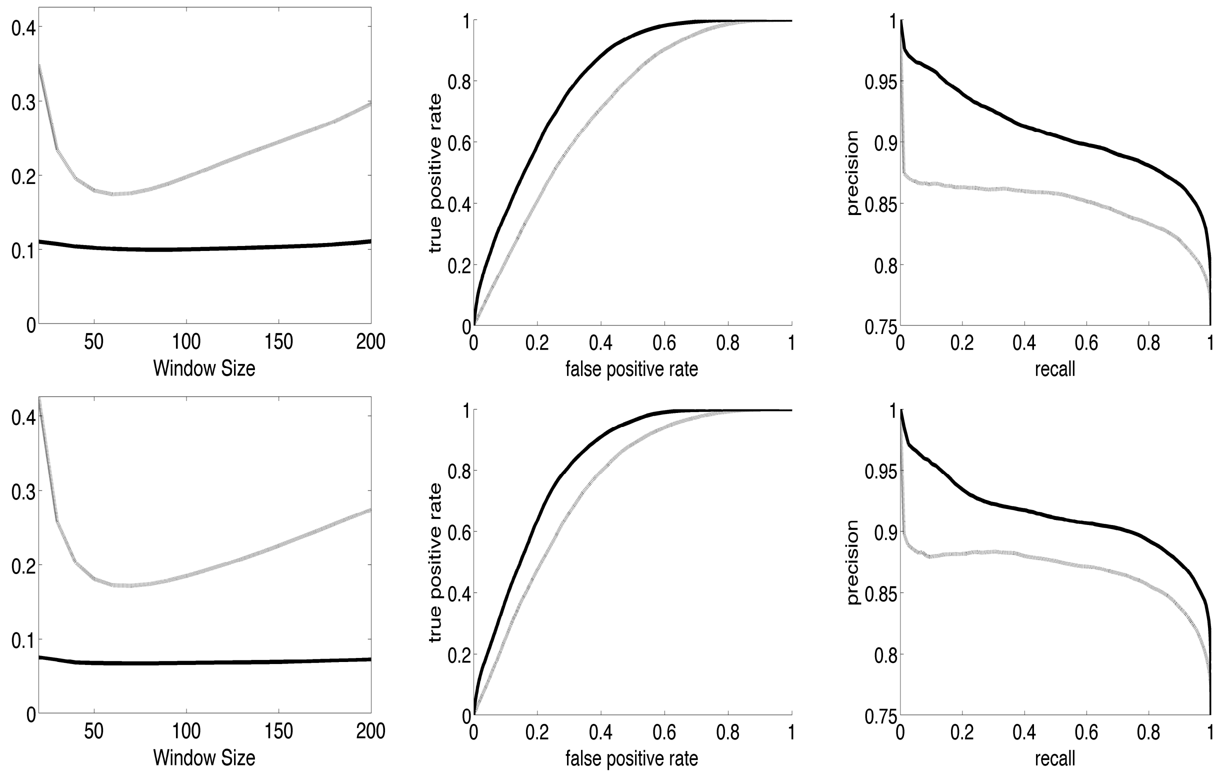


Figure A4: Results for Experiment 3

Figure A4 shows the results of Experiment 3, for which the underlying parameters follow a random walk process. The light dashed line relates to results obtained using classical Granger causality testing over rolling windows, while bold solid line relates to results obtained using the proposed time-varying parameter framework. The upper panel shows results for pairwise estimation and inference. The lower panel shows results using conditional estimation and inference. The left column figures show the mean squared error of cross-parameter estimates. The middle and right column figures show the ROC curves and the PR curves for inference of the underlying network, with rolling window size set at 200 observations for the rolling window approach.



References

Barigozzi, M., and C. Brownlees. "NETS: Network Estimation for Time Series." (2017). ECARES Working Paper, Université libre de Bruxelles.

Baumeister, C., and L. Benati. "Unconventional monetary policy and the Great Recession: Estimating the macroeconomic effects of a spread compression at the zero lower bound." *International Journal of Central Banking*, 9 (2013), 165–212.

Baumeister, C., and G. Peersman. "Time-Varying Effects of Oil Supply Shocks on the U.S. Economy." *American Economic Journal: Macroeconomics*, 5 (2013), 1–29.

Benati, L., and H. Mumtaz. "US evolving macroeconomic dynamics: A structural investigation." (2007). ECB Working Paper No. 746.

Carter, C. K., and R. Kohn. "On Gibbs Sampling for State Space Models." *Biometrika*, 81 (1994), 541–553.

Cogley, T.; S. W. Morozov; and T. J. Sargent. "Bayesian fan charts for UK inflation: Forecasting and sources of uncertainty in an evolving monetary system." *Journal of Economic Dynamics & Control*, 29 (2005), 1893–1925.

Cogley, T.; G. E. Primiceri; and T. J. Sargent. "Inflation-Gap persistance in the US." *American Economic Journal: Macroeconomics*, 2 (2010), 43–69.

Davis, J., and M. Goadrich. "The relationship between precision-recall and roc curves." In *23rd international conference on Machine learning*. 233–240.

Jacquier, E.; N. G. Polson; and P. E. Rossi. "Bayesian analysis of stochastic volatility models." *Journal of Business & Economic Statistics*, 12 (1994), 371–389.

Jacquier, E.; N. G. Polson; and P. E. Rossi. "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors." *Journal of Econometrics*, 122 (2004), 185–212.

Koop, G.; R. Leon-Gonzalez; and R. W. Strachan. "Dynamic Probabilities of Restrictions in State Space Models: An Application to the Phillips Curve." *Journal of Business and Economic Statistics*, 28 (2010), 370–379.

Primiceri, G. "Time Varying Structural Vector Autoregressions and Monetary Policy." *Review of Economic Studies*, 72 (2005), 821–852.

Seth, A. K. "A matlab toolbox for granger causal connectivity analysis." *Journal of Neuroscience Methods*, 186 (2010), 262–273.