Sampling distribution for single-regression Granger causality estimators

Article  (Supplemental Material)

Gutknecht, AJ and Barnett, L (2023) Sampling distribution for single-regression Granger causality estimators. Biometrika.asad009. ISSN 0006-3444

This version is available from Sussex Research Online: http://sro.sussex.ac.uk/id/eprint/110720/

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher’s version. Please see the URL above for details on accessing the published version.

Copyright and reuse:
Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.
Supplementary material for Sampling distribution for single-regression Granger causality estimators

BY A. J. GUTKNECHT
Campus Institute for Dynamics of Biological Networks, Georg-August University, Goettingen, Germany, and MEG Unit, Brain Imaging Center, Goethe University, Frankfurt am Main, Germany
agutkne@uni-goettingen.de

AND L. BARNETT
Sussex Centre for Consciousness Science and Department of Informatics, University of Sussex, Falmer, Brighton, UK
l.c.barnett@sussex.ac.uk

1. THE GENERALIZED $\chi^2$ FAMILY OF DISTRIBUTIONS

Let $Z \sim \mathcal{N}(0, B)$ be a zero-mean $n$-dimensional multivariate-normal random vector with covariance matrix $B$, and $A$ an $n \times n$ symmetric matrix. Then (Jones, 1983) we write $\chi^2(A, B)$ for the distribution of the random quadratic form $Q = Z^T A Z$. If $A = B = I$, then $\chi^2(A, B)$ reduces to the usual $\chi^2(n)$. If $A$ is $m \times m$ and $C$ is $m \times n$, then $\chi^2(A, C B C^T) = \chi^2(C^T A C, B)$.

It is not hard to show (Mohsenipour, 2012) that if $B$ is positive-definite and $A$ symmetric (which will be the case for the generalized $\chi^2$ distributions we encounter), then $\chi^2(A, B) = \chi^2(\Lambda, I)$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1, \ldots, \lambda_n$ the eigenvalues of $BA$, or, equivalently, of $RAR^T$ where $R$ is the right Cholesky factor of $B$ (so that $R^T R = B$). In that case, we have

$$\lambda_1 U_1^2 + \ldots + \lambda_n U_n^2 \sim \chi^2(A, B), \quad U_i \text{ iid } \mathcal{N}(0, 1),$$

so that $\chi^2(A, B)$ is a weighted sum of independent $\chi^2$-distributed variables, and in particular if the $\lambda_i$ are all equal then we have a scaled $\chi^2(n)$ distribution. From (1), moments of a generalized $\chi^2$ variable may be conveniently expressed in terms of the eigenvalues; thus we may calculate that for $Q \sim \chi^2(A, B)$

$$\mathbb{E}[Q] = \mu = \sum_{i=1}^n \lambda_i, \quad \text{var}[Q] = \sigma^2 = 2 \sum_{i=1}^n \lambda_i^2. \quad (2)$$

Empirically, it is found that generalized $\chi^2$ variables (at least for $A$ symmetric and $B$ positive-definite) are very well approximated by $\Gamma$ distributions: specifically, we have $Q \approx \Gamma(\alpha, \beta)$ with shape and scale parameters

$$\alpha = \mu^2 \sigma^{-2}, \quad \beta = \mu^{-1} \sigma^2 \quad (3)$$

respectively.

2. PROOF OF MAIN ARTICLE, PROPOSITION 1

Proof. Let $\theta = [x^T \ y^T]^T$ where $x_i = \theta_i (i = 1, \ldots, d)$ and $y_j = \theta_{d+j} (j = 1, \ldots, s)$. Since by definition $f(0, y) = 0$ for all $y$, we have immediately $\nabla_y f(0, y) = 0$ for all $y$. Treating $y$ as fixed,
we expand \(f(x, y)\) in a Taylor series around \(x = 0\):

\[
f(x, y) = \nabla_x f(0, y)x + \frac{1}{2}x^T\nabla_{xx}^2 f(0, y)x + \frac{1}{2}x^T K(x, y)x,
\]

where for fixed \(y\), \(K(x, y)\) is a \(d \times d\) matrix function of \(x\), and \(\lim_{x \to 0} \|K(x, y)\| = 0\). Now we show that since \(f(x, y)\) is non-negative, we must have \(\nabla_x f(0, y) = 0\) for all \(y\). Suppose, say, \(\nabla_x f(0, y) = -\mathbf{g} < 0\). Setting \(x = \varepsilon > 0\) \([\text{if} \ \nabla_x f(0, y) > 0 \ \text{we take} \ x = -\varepsilon]\) and \(x_2 = \ldots = x_d = 0\), (4) yields

\[
f(x, y) = -\varepsilon \mathbf{g} + \frac{1}{2} \left( \nabla_{x_1}^2 f(0, y) + K_{11}(x, y) \right) \varepsilon^2.
\]

Now since \(\lim_{x \to 0} \|K(x, y)\| = 0\), we can always choose \(\varepsilon\) small enough that \(\frac{1}{2} \left( \nabla_{x_1}^2 f(0, y) + K_{11}(x, y) \right) \varepsilon < \mathbf{g}\), so that \(f(\varepsilon, 0, \ldots, 0, y) < 0\), a contradiction. Thus we have \(\nabla_x f(0, y) = 0\) for all \(y\), proving Proposition 1a.

From (4) we thus have

\[
f(x, y) = \frac{1}{2}x^T \nabla_{xx}^2 f(0, y)x + \frac{1}{2}x^T K(x, y)x.
\]

To see that \(\nabla_{xx}^2 f(0, y)\) must be positive-semidefinite, we assume the contrary. We may then find a unit \(d\)-dimensional vector \(u\) such that \(u^T \nabla_{xx}^2 f(0, y) = -G < 0\). Setting \(x = \varepsilon u\), we may then choose \(\varepsilon\) small enough that \(u^T K(\varepsilon u, y) u = -G < 0\), so that again \(f(x, y)\) is negative and we have a contradiction. Finally, \(\nabla_{xx}^2 f(0, y) = \nabla_{yy}^2 f(0, y) = 0\) for all \(y\) follows directly from (5), and we have established Proposition 1b.

We now prove Proposition 1c using a 2nd-order Delta Method (Lehmann & Romano, 2005). Let \(\theta \in \Theta_0\). Since \(f(\theta)\) and its gradient \(\nabla f(\theta)\) both vanish, the Taylor expansion of \(f(\nabla \theta)\) around \(\theta\) takes the form

\[
f(\nabla \theta) = \frac{1}{2} (\nabla \theta - \theta)^T W(\theta) (\nabla \theta - \theta) + (\nabla \theta - \theta)^T K(\nabla \theta) (\nabla \theta - \theta),
\]

where \(W(\theta) = \nabla^2 f(\theta)\) is the Hessian of \(f\) evaluated at \(\theta\), and \(\lim_{\theta \to \theta_0} \|K(\theta)\| = 0\). By assumption \(N^{\frac{1}{2}} (\nabla \theta - \theta) \xrightarrow{d} Z\) as \(N \to \infty\), where \(Z \sim \mathcal{N}(0, \Omega(\theta))\). Therefore, multiplying both sides of (6) by the sample size \(N\), by the Continuous Mapping Theorem (van der Vaart, 2000), we have

\[
N f(\nabla \theta) \xrightarrow{d} \frac{1}{2} Z^T W(\theta) Z
\]

as \(N \to \infty\), and Proposition 1c follows immediately from (7) and Proposition 1b.
3. Proof of Main Article, Proposition 2

Proof. Following Barnett & Seth (2015), given a VAR($p$) model

$$U_t = \sum_{k=1}^{p} A_k U_{t-k} + \varepsilon_t$$

(9)

with parameters $\theta = (A; \Sigma)$, we create an equivalent innovations-form state-space model (Hannan & Deistler, 2012)

$$Z_{t+1} = \tilde{A} Z_t + K \varepsilon_t,$$

$$U_t = A Z_t + \varepsilon_t,$$

where

$$\tilde{A} = \begin{bmatrix} A_1 & A_2 & \ldots & A_{p-1} & A_p \\ I & 0 & \ldots & 0 & 0 \\ 0 & I & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & I & 0 \end{bmatrix}, \quad A = \begin{bmatrix} A_1 & A_2 & \ldots & A_{p-1} & A_p \end{bmatrix}, \quad K = \begin{bmatrix} I \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$  

$\tilde{A}$ is the $pn \times pn$ state transition matrix (the companion matrix of VAR coefficients) for the VAR($p$) (9), $K$ the $pn \times n$ Kalman gain matrix, and $A$ the $n \times pn$ observation matrix. As before, we use subscript $x$ for the indices $1, \ldots, n_x$, $y$ for the indices $n_x + 1, \ldots, n$, and we use an asterisk to denote “all indices”. The subprocess $X_t$ then satisfies the state-space model

$$Z_{t+1} = \tilde{A} Z_t + K \varepsilon_t,$$

$$X_t = A_{xx} Z_t + \varepsilon_{x,t}.$$  

This state-space model is no longer in innovations form; we can, however (see Barnett & Seth, 2015) derive an innovations-form state-space model for $X_t$ by solving the discrete-time algebraic Riccati equation (DARE)

$$P = \tilde{A} P \tilde{A}^T + \tilde{\Sigma} - (\tilde{A} P A_{xx}^T + \tilde{\Sigma}_{xx})(A_{xx} P A_{xx}^T + \Sigma_{xx})^{-1}(\tilde{A} P A_{xx}^T + \tilde{\Sigma}_{xx})^T$$

(12)

for $P$ (a $pn \times pn$ symmetric matrix), with

$$\tilde{\Sigma} = \begin{bmatrix} \Sigma & 0 & \ldots & 0 \\ 0 & \Sigma & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \Sigma \\ \end{bmatrix}, \quad \tilde{\Sigma}_{xx} = \begin{bmatrix} \Sigma_{xx} \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

which are, respectively, $pn \times pn$ and $pn \times n_x$. We note that under our assumptions, a stabilising solution for (12) exists, and is unique (Solo, 2016). Then

$$Z_{t+1} = \tilde{A} Z_t + K^R \varepsilon_{t}^R,$$

$$X_t = A_{xx} Z_t + \varepsilon_{x,t}^R$$

(13a)

(13b)

is in innovations form, with innovations covariance matrix and Kalman gain matrix

$$\Sigma^R = A_{xx} P A_{xx}^T + \Sigma_{xx},$$

$$K^R = (\tilde{A} P A_{xx}^T + \tilde{\Sigma}_{xx}) \left[ \Sigma^R \right]^{-1}$$

(14a)

(14b)
respectively. The innovations $\varepsilon_t^R$ in (13) are precisely the residuals of the reduced VAR model for $X_t$, and $\Sigma^R = E[\varepsilon_t^R \varepsilon_t^{R \top}]$ implicitly defines $V(\theta)$ as required for the single-regression Granger causality statistic $F_{Y \rightarrow X}(\theta)$ (see Main Article, Section 3 and Section 3.3).

We may in fact confirm that

$$\Sigma^R = V(\theta) = A_{xy} \Pi A_{xy}^\top + \Sigma_{xx},$$

where the $p_y \times p_y$ symmetric matrix $\Pi$ is the unique stabilising solution of the reduced DARE

$$\Pi = \check{A}_{yy} \Pi A_{yy}^\top + \check{\Sigma}_{yy} - (\check{A}_{yy} \Pi A_{xy}^\top + \check{\Sigma}_{yx}) (A_{xy} \Pi A_{xy}^\top + \Sigma_{xx})^{-1} (\check{A}_{yy} \Pi A_{xy}^\top + \check{\Sigma}_{yx})^\top,$$

with

$$\check{A}_{yy} = \begin{bmatrix} A_{1,yy} & A_{2,yy} & \cdots & A_{p-1,yy} & A_{p,yy} \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}, \quad A_{xy} = \begin{bmatrix} A_{1,xy} & A_{2,xy} & \cdots & A_{p-1,xy} & A_{p,xy} \end{bmatrix},$$

which are, respectively, $p_y \times p_y$ and $n_x \times p_y$, and

$$\check{\Sigma}_{yy} = \begin{bmatrix} \Sigma_{yy} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \check{\Sigma}_{yx} = \begin{bmatrix} \Sigma_{yx} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

respectively, $p_y \times p_y$ and $p_y \times n_x$. To see this, we may verify by substitution that if

$$\Pi = \begin{bmatrix} \Pi_{11} & \cdots & \Pi_{1p} \\ \vdots & \ddots & \vdots \\ \Pi_{p1} & \cdots & \Pi_{pp} \end{bmatrix}$$

solves the reduced-dimension DARE (16), where the $\Pi_{kl}$ are $n_y \times n_y$, then

$$P = \begin{bmatrix} 0_{n_y \times n_x} & 0_{n_y \times n_y} & \cdots & 0_{n_y \times n_y} \\ 0_{n_y \times n_x} & \Pi_{11} & \cdots & 0_{n_y \times n_y} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{n_y \times n_x} & 0_{n_y \times n_y} & \cdots & \Pi_{pp} \end{bmatrix}$$

solves the original DARE (12), and that $\Sigma^R$ is given by (15). We may also confirm that the Kalman gain matrix (14b) for the reduced DARE is

$$R^R = \begin{bmatrix} I_{n_y \times n_x} \\ L^R_1 \\ 0_{n_y \times n_x} \\ L^R_2 \vdots \\ 0_{n_y \times n_x} \\ L^R_p \end{bmatrix},$$
where

\[ L^R = (\tilde{A}_{yy} \Pi A_{xy} + \tilde{\Sigma}_{yx}) \left[ \Sigma^R \right]^{-1} = \begin{bmatrix} L^R_1 \\ L^R_2 \\ \vdots \\ L^R_p \end{bmatrix} \]

(the \( L^R_k \) are \( n_y \times n_x \)) is the Kalman gain matrix associated with the DARE (16).

\[ \square \]

4. Proof of Main Article, Theorem 1

**Proof.** To calculate \( W_0(\theta) \) we require derivatives up to 2nd order of \( V(\theta) \) (15) with respect to the null-hypothesis parameters (that is, with respect to \( A_{k,ij} \) for \( i \in x, j \in y \)), evaluated for \( \theta \in \Theta_0 \). From (15) we may calculate:

\[ \frac{\partial V_{ii'}}{\partial A_{k,uv}} = \delta_{ui} \left[ \Pi A_{xy} \right]_{k,vi'} + \delta_{ui'} \left[ A_{xy} \Pi \right]_{ki,iv} + \left[ A_{xy} \frac{\partial \Pi}{\partial A_{k,uv}} A_{xy}^T \right]_{i'i'}, \tag{17} \]

where indices \( i, i', u, u' \in x \), indices \( j, j', v, v' \in y \) and \( k, k' = 1, \ldots, p \). Since \( A_{xy} \) vanishes under the null hypothesis, we have

\[ \left. \frac{\partial V_{ii'}}{\partial A_{k,uv}} \right|_{\theta \in \Theta_0} = 0, \tag{18} \]

and from (17) we find

\[ \left. \frac{\partial^2 V_{ii'}}{\partial A_{k,uv} \partial A_{k',u'v'}} \right|_{\theta \in \Theta_0} = \left[ \delta_{ui} \delta_{u'i'} + \delta_{ui'} \delta_{u'i} \right] \Pi_{kk',vv'}. \] \tag{19}

We see then that \( \Pi \) is required only on the null space \( A_{xy} = 0 \), in which case the DARE (16) becomes a discrete Lyapunov (DLYAP) equation for \( \Pi_0 = \Pi_{\theta \in \Theta_0} : \)

\[ \Pi_0 = \tilde{A}_{yy} \Pi_0 \tilde{A}_{yy}^T + \tilde{\Sigma}_{yx} \left[ \Sigma_{xx} \right]^{-1} \tilde{\Sigma}_{yx}^T. \]

We may now calculate the required Hessian. For null parameters \( \theta_\alpha, \theta_\beta \), from the definition \( F_{Y \to X}(\theta) = \log |V(\theta)| - \log |\Sigma_{xx}| \) and using (18) we may calculate

\[ \left. \frac{\partial^2 F_{Y \to X}}{\partial \theta_\alpha \partial \theta_\beta} \right|_{\theta \in \Theta_0} = \text{tr} \left[ \left[ \Sigma_{xx} \right]^{-1} \left. \frac{\partial^2 V}{\partial \theta_\alpha \partial \theta_\beta} \right|_{\theta \in \Theta_0} \right]. \tag{20} \]

Eq. (19) then yields

\[ \left. W_0(\theta) \right|_{[k,uv],[k',u'v']} = 2 \left[ \left[ \Sigma_{xx} \right]^{-1} \right]_{uu'} \left[ \Pi_0 \right]_{kk',vv'}, \]

or

\[ W_0(\theta) = 2 \left[ \Sigma_{xx} \right]^{-1} \otimes \Pi_0, \]

as required. \( \square \)

5. Proof of Main Article, Proposition 4

**Proof.** Dropping the “\( \omega \)” and “\( \theta \)” arguments for compactness where convenient, on the null space \( \Theta_0 \) we have \( \Phi_{xy} = 0 \), and since then \( \Phi \) is lower block-triangular, we have also \( \Psi_{xx} = \)

\[ \square \]
\( [\Phi_{xx}]^{-1}, \Psi_{yy} = [\Phi_{yy}]^{-1}, \) and \( \Psi_{xy} = 0. \) The CPSD for the process \( X_t \) is given by

\[
S_{xx} = [\Psi^* \Psi']_{xx} = \Psi_{xx} \Sigma_{xx} \Psi_{xx} + \Psi_{xy} \Sigma_{yx} \Psi_{xx}^* + \Psi_{xx} \Sigma_{yx} \Psi_{xx} + \Psi_{xy} \Sigma_{yx} \Psi_{xx}. 
\]

On the null space \( S_{xx} = \Psi_{xx} \Sigma_{xx} \Psi_{xx}^* \) so that \( \Phi_{xx} \Sigma_{xx} \Phi_{xx}^* \). We define \( T(\omega) \) to be the \( n_x \times n_x \) (Hermitian) matrix \( T(\omega) = \Psi_{xy}(\omega) \Sigma_{yx}(\omega)^*, \) so that from Main Article, eq. 12, \( f_{Y \to X}(\chi) = \log |S_{xx}(\omega)| - \log |S_{xx}(\omega) - T(\omega)|. \) \( T(\omega) \) vanishes on the null space. We may check that

\[
\frac{\partial \Psi_{pq}}{\partial \theta_{rs}} = \Psi_{pr} \Psi_{sq} e^{-i\omega k} \quad \text{(p, q, r, s = 1, \ldots, n; k = 1, \ldots, p)}, \quad (21)
\]

from which we may calculate (with \( i, i', u, u' \in x, j, j', v, v' \in y)\)

\[
\frac{\partial^2 f_{Y \to X}}{\partial A_{k,rs} \partial A_{j,uv}} = \Psi_{iu} \Psi_{i'u'} [S_{xy|x}]_{uv} e^{-i\omega(k-k')} + \Psi_{iu} \Psi_{i'u'} [S_{yy|x}]_{v'v} e^{i\omega(k-k')}, \quad (22)
\]

so that in particular \( \frac{\partial T}{\partial \theta_0} \bigg|_{\theta_0} = 0 \) for a null parameter \( \theta_0, \) and we find [cf. (20)]

\[
\frac{\partial^2 f_{Y \to X}}{\partial \theta_0 \partial \theta_0} \bigg|_{\theta_0} = \text{tr} \left[ [S_{xx}]^{-1} \frac{\partial^2 T}{\partial \theta_0 \partial \theta_0} \bigg|_{\theta_0} \right] \quad (23)
\]

for null parameters \( \theta_0, \theta_0. \) From (22) and using (21), we may calculate

\[
\frac{\partial^2 T_{i,i'}}{\partial A_{k,uv} \partial A_{j,uv}} \bigg|_{\theta_0} = \Psi_{iu} \Psi_{i'u'} [S_{xy|x}]_{uv} e^{-i\omega(k-k')} + \Psi_{iu} \Psi_{i'u'} [S_{yy|x}]_{v'v} e^{i\omega(k-k')}, \quad (24)
\]

where

\[
S_{yy|x} = \Psi_{yy} - \Sigma_{yy|x} \Psi_{yy} \quad (25)
\]

is the CPSD for a VAR(p) model with parameters \( (A_{yy}, \Sigma_{yy|x}). \) From (23) and (24) we find

\[
\frac{\partial^2 f_{Y \to X}}{\partial A_{k,uv} \partial A_{k',u'v'}} \bigg|_{\theta_0} = \left[ [\Sigma_{xx}]^{-1} \right]_{u'u'} \left\{ \left[S_{yy|x}\right]_{v'v} e^{-i\omega(k-k')} + \left[S_{yy|x}\right]_{v'v} e^{i\omega(k-k')} \right\}
\]

\[
= \left[ [\Sigma_{xx}]^{-1} \right]_{u'u'} \left\{ \left[S_{yy|x} e^{-i\omega(k-k')} + \Sigma_{yy|x} e^{i\omega(k-k')} \right]_{v'v'} \right\}
\]

\[
= 2 \left[ [\Sigma_{xx}]^{-1} \right]_{u'u'} \Re \left\{ \tilde{S}_{yy|x}(kk',uu') \right\},
\]

where

\[
\tilde{S}_{yy|x}(\omega) = Z(\omega) \otimes S_{yy|x}(\omega), \quad Z_{kk'}(\omega) = e^{-i\omega(k-k')}.
\]

The \( pm_y \times pm_y \) Hermitian matrix \( \tilde{S}_{yy|x}(\omega) \) is the CPSD for the companion VAR(1) (Main Article, eq. 2) of a VAR(p) model with parameters \( (A_{yy}, \Sigma_{yy|x}), \) and as such may be thought of as the spectral counterpart of the autocovariance matrix \( \tilde{\Gamma}_{yy|x} \) of Main Article, Section 4.2. Thus for \( \omega \in [0, 2\pi], \) we have \( W_0(\omega; \theta) = [\Sigma_{xx}]^{-1} \otimes \Re \{ \tilde{S}_{yy|x}(\omega) \}, \) so that

\[
W_0(\mathcal{F}; \theta) = [\Sigma_{xx}]^{-1} \otimes \Re \{ \tilde{S}_{yy|x}(\mathcal{F}) \}
\]
with
\[
\hat{S}_{yy|x}(\mathcal{F}) = \frac{1}{|\mathcal{F}|} \int_{\mathcal{F}} \hat{S}_{yy|x}(\omega) \, d\omega
\]  
(26)

as required. □

6. PROOF OF MAIN ARTICLE, THEOREM 3

**Lemma 1.** Suppose given a sequence of pairs of real-valued random variables \((X_n, Y_n)\) such that \(X_n \xrightarrow{d} X\) and \(Y_n \xrightarrow{n} c\), where \(c\) is a constant. Then
\[
\text{pr}[X_n \leq Y_n] \longrightarrow \text{pr}[X \leq c]
\]  
(27)
as \(n \longrightarrow \infty\).

**Proof.** By Slutsky’s Lemma (Lehmann & Romano, 2005), we have \(X_n - Y_n \xrightarrow{d} X - c\). Thus for any \(\varepsilon > 0\) there exists \(n_0 \in \mathbb{N}\) such that for all \(n \geq n_0\) we have \(|\text{pr}[X_n - Y_n \leq 0] - \text{pr}[X - c \leq 0]| < \varepsilon\), or, equivalently \(|\text{pr}[X_n \leq Y_n] - \text{pr}[X \leq c]| < \varepsilon\), which establishes (27). □

**Proof of Main Article, Theorem 3.** Main Article, eq. 38 is equivalent to \(P_I(\theta; \alpha) = 1 - \text{pr} \left[ NF \leq \Phi^{-1}_0(1 - \alpha) \right] \). Since \(\hat{\theta}\) is a consistent estimator for \(\theta\) and the projection of \(\hat{\theta}\) onto \(\Theta_0\) is continuous—and maps any \(\theta \in \Theta_0\) to itself—by the Continuous Mapping Theorem the projection of \(\hat{\theta}\) onto \(\Theta_0\) converges in probability to \(\theta\). It is not hard to verify that the inverse CDF \(\Phi^{-1}(\cdots)\) evaluated at \(1 - \alpha\) is continuous in \(\theta\) argument, so that again by the Continuous Mapping Theorem we have \(\Phi^{-1}_0(1 - \alpha) \xrightarrow{p} \Phi^{-1}_\theta(1 - \alpha)\). By Main Article, Theorem 1, \(NF \xrightarrow{d} Q_\theta\), where \(Q_\theta\) is a (generalized \(\chi^2\)) random variable with CDF \(\Phi_\theta\). Applying Lemma 1 to the pair-sequence \((NF, \Phi^{-1}_\theta(1 - \alpha))\) we have
\[
P_I(\theta; \alpha) \longrightarrow 1 - \text{pr}[Q_\theta \leq \Phi^{-1}_\theta(1 - \alpha)] = 1 - \Phi_\theta(\Phi^{-1}_\theta(1 - \alpha)) = \alpha
\]as required. □

7. WORKED EXAMPLE: THE GENERAL BIVARIATE VAR(1)

Consider the bivariate VAR(1)
\[
\begin{align*}
X_t &= a_{xx}X_{t-1} + a_{xy}Y_{t-1} + \varepsilon_{xt}, \quad (28a) \\
Y_t &= a_{yx}X_{t-1} + a_{yy}Y_{t-1} + \varepsilon_{yt}, \quad (28b)
\end{align*}
\]
with parameters
\[
A = \begin{bmatrix} a_{xx} & a_{xy} \\ a_{yx} & a_{yy} \end{bmatrix}, \quad \Sigma = \text{E}[\varepsilon_t \varepsilon_t^\top] = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix},
\]
so that \(\theta = (a_{xx}, a_{xy}, a_{yx}, a_{yy}, \sigma_{xx}, \sigma_{xy}, \sigma_{yx}, \sigma_{yy})\), and the null hypothesis \(H_0\) (Main Article, eq. 17) is \(a_{xy} = 0\). The transfer function is then \(\Psi(\omega) = \Phi(\omega)^{-1}\) with \(\Phi(\omega) = I - Az\), and the factorisation \(S(\omega) = \Psi(\omega)\Sigma \Psi(\omega)^\ast\) (Main Article, eq. 8) of the CPSD \(S(\omega)\) holds for \(\omega \in [0, 2\pi]\).

Setting \(\Delta(\omega) = |\Phi(\omega)|\) (determinant), we have
\[
\Psi(\omega) = \begin{bmatrix} 1 - a_{xx}z & -a_{xy}z \\ -a_{yx}z & 1 - a_{yy}z \end{bmatrix}^{-1} = \Delta(\omega)^{-1} \begin{bmatrix} 1 - a_{yy}z & a_{xy}z \\ a_{yx}z & 1 - a_{xx}z \end{bmatrix}.
\]
This leads to
\[ S(\omega) = |\Delta(\omega)|^{-2} \left[ \begin{array}{cc} 1 - a_{yy}z & a_{xy}z \\ a_{yx}z & 1 - a_{xx}z \end{array} \right] \left[ \begin{array}{cc} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{array} \right] \left[ \begin{array}{c} 1 - a_{yy}\bar{z} \\ a_{xy}\bar{z} \end{array} \right] \]
onumber
on \( z = 1 \), where \( z = e^{-i\omega} \) and \( \bar{z} \) is the complex conjugate.

We wish to calculate the Granger causality \( F_Y \rightarrow X \). If \( v \) is the residuals variance for the VAR representation of the subprocess \( X_t \), then the Granger causality is just \( F_Y \rightarrow X = \log v - \log \sigma_{xx} \).

To solve for \( v \) we could use the reduced DARE (16), but here we use an explicit spectral factorisation for the CPSD \( S_{xx}(\omega) \) of \( X_t \).

Let \( \psi(\omega) \) be the transfer function of the process \( X_t \), which satisfies the spectral factorisation \( S_{xx}(\omega) = v|\psi(\omega)|^2 \). We may now calculate (we denote terms we don’t need by “…”).

\[
S(\omega) = |\Delta(\omega)|^{-2} \left[ \begin{array}{c} 1 - a_{yy}z \\ a_{yx}z \end{array} \right] \left[ \begin{array}{cc} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{array} \right] \left[ \begin{array}{c} 1 - a_{yy}\bar{z} \\ a_{xy}\bar{z} \end{array} \right] 
\]

We now calculate (note that \( z + \bar{z} = 2 \cos \omega \)):

\[
S_{xx}(\omega) = |\Delta(\omega)|^{-2} \left\{ \left(1 - a_{yy}z\right)\left[\sigma_{xx}(1 - a_{yy}z) + \sigma_{xy}\sigma_{yy}\bar{z}\right] + a_{xy}\sigma_{yx}(1 - a_{yy}z) + \sigma_{yy}\left[\sigma_{xx}(1 - a_{yy}\bar{z}) + \sigma_{yx}\sigma_{yy}\bar{z}\right] \right\} 
\]

\[
= |\Delta(\omega)|^{-2} \left\{ \left(1 - a_{yy}z\right)^2 + \sigma_{xy}\sigma_{yy}[(1 - a_{yy}z)^2 + (1 - a_{yy}\bar{z})\bar{z} + \sigma_{xy}\sigma_{yy}(z + \bar{z}) - 2a_{yy} + \sigma_{yy}\bar{z}] \right\} 
\]

and finally

\[
S_{xx}(\omega) = |\Delta(\omega)|^{-2} \left( P - Q \cos \omega \right) 
\]

where we have set

\[
P = \sigma_{xx}(1 + a_{yy}^2) - 2\sigma_{xy}\sigma_{yy} + \sigma_{yy}a_{xy}^2, \quad Q = 2(\sigma_{xx}\sigma_{yy} - \sigma_{xy}\sigma_{yy}) 
\]

The form of this expression suggests that the transfer function \( \psi(\omega) \) should take the form

\[
\psi(\omega) = \Delta(\omega)^{-1}(1 - bz) 
\]

for some constant \( b \); this implies that \( X_t \) is VARMA(2,1). Then \( |\psi(\omega)|^2 = |\Delta(\omega)|^{-2}(1 + b^2 - 2b \cos \omega) \) and the spectral factorisation \( S_{xx}(\omega) = v|\psi(\omega)|^2 \) now reads

\[
v(1 + b^2 - 2b \cos \omega) = P - Q \cos \omega. 
\]

This must hold for all \( \omega \), so we have

\[
v(1 + b^2) = P, \quad vb = \frac{1}{2}Q. 
\]

We may now solve for \( v \). The second equation gives \( v^2b^2 = \frac{1}{4}Q^2 \), so multiplying the first equation through by \( v \) we obtain the quadratic equation for \( v \)

\[
v^2 - Pv + \frac{1}{4}Q^2 = 0, 
\]

with solutions

\[
v = \frac{1}{2} \left\{ P \pm (P^2 - Q^2)^{\frac{1}{2}} \right\}. 
\]

\(^1\) For a complex variable \( w \), \( |w| \) denotes the norm \( \langle w\bar{w} \rangle^\frac{1}{2} \).
We need to take the “+” solution, as this yields the correct (zero) result for the null case $a_{xy} = 0$, so that

$$F_{Y \rightarrow X} = \log \frac{1}{2} \left\{ P + (P^2 - Q^2)^{\frac{1}{2}} \right\} - \log \sigma_{xx}.$$  

Besides the residuals covariances $\Sigma$, only the $Y \rightarrow X$ (i.e., causal) autoregression coefficient $a_{xy}$ and the $Y$ autoregression coefficient $a_{yx}$ appear in the expression for $F_{Y \rightarrow X}$. We note that, given any $F > 0$ and a set of model parameters excluding $a_{xy}$, there are in general two possible values of $a_{xy}$ which yield $F_{Y \rightarrow X} = F$, except in cases where no $a_{xy}$ exists due to the stability constraint on the spectral radius, which requires that $|a_{xx}a_{yy} - a_{xy}a_{yx}| < 1$.

From Main Article, eq. 12, the spectral Granger causality from $Y \rightarrow X$ is

$$f_{Y \rightarrow X}(\omega) = \log(P - Q \cos \omega) - \log \left( P - Q \cos \omega - a_{xy}^2 \sigma_{yy|x} \right),$$

where $\sigma_{yy|x} = \sigma_{yy} - \sigma_{xy}^2 \sigma_{xx}^{-1} = \sigma_{yy}(1 - \kappa^2)$, with $\kappa = \sigma_{xy}(\sigma_{xx}\sigma_{yy})^{-\frac{1}{2}}$ the residuals correlation.

For the sampling distributions, we shall also need the (inverse of) the covariance matrix $\Gamma_0$ of the process $[X_t^T \ Y_t^T]^T$ on the null space $a_{xy} = 0$. Solving the DLYAP equation $\Gamma_0 - A\Gamma_0 A^T = \Sigma$ for

$$\Gamma_0 = \begin{bmatrix} p & r \\ r & q \end{bmatrix}$$

yields

$$p = \left(1 - a_{xx}^2\right)^{-1} \sigma_{xx},$$

$$r = \left(1 - a_{xx}a_{yy}\right)^{-1} \sigma_{xy} + a_{xx}a_{yx} \left(1 - a_{xx}^2\right)^{-1} \sigma_{xx},$$

$$q = \left(1 - a_{yy}^2\right)^{-1} \left\{ \sigma_{yy} + 2a_{yy}a_{xy}(1 - a_{xx}a_{yy})^{-1} \sigma_{xy} \\
+ a_{xy}^2(1 + a_{xx}a_{yy})(1 - a_{xx}^2)^{-1} (1 - a_{xx}a_{yy})^{-1} \sigma_{xx} \right\},$$

and we have in particular

$$\omega_{yy} = [\Gamma_0^{-1}]_{yy} = \frac{p}{pq - r^2}. \tag{30}$$

Note also that in the null case $a_{xy} = 0$, the spectral radius is $\rho = \max(|a_{xx}|, |a_{yy}|)$.

We apply Main Article, Theorem 1 to calculate the asymptotic distribution of the single-regression estimator $\hat{\Gamma}_{Y \rightarrow X}^{SR}$ on the null space. Noting that for model order $p = 1$ we have $\hat{\Gamma} = \Gamma_0$, and setting $\Gamma_0^{-1} = [\omega_{ij}]$, we have $[\Gamma^{-1}]_{yy} = [\Gamma_0^{-1}]_{yy} = \omega_{yy}$, where $\omega_{yy}$ is given by (30). Solving the DLYAP equation (Main Article, eq. 32) for $\Gamma_{yy|x}$, we find that $\Gamma_{yy|x} = (1 - \kappa^2)\sigma_{yy}/(1 - a_{yy}^2)$, and the single eigenvalue of $[\hat{\Gamma}^{-1}]_{yy}\hat{\Gamma}_{yy|x}$ is $\lambda = (1 - \kappa^2)\sigma_{yy}\omega_{yy}/(1 - a_{yy}^2)$.

By Main Article, Theorem 1 the asymptotic distribution of the single-regression estimator is thus a scaled $\chi^2(1)$:

$$N\hat{F}_{Y \rightarrow X}^{SR} \xrightarrow{d} \lambda \cdot \chi^2(1) = \Gamma \left( \frac{1}{2}, 2\lambda \right),$$

and the $\Gamma$-approximation in this case is exact.

We also calculate the spectral Granger causality from $Y \rightarrow X$ at $\omega \in [0, 2\pi]$ as

$$f_{Y \rightarrow X}(\omega; \theta) = \log(P - Q \cos \omega) - \log \left\{ P - Q \cos \omega - (1 - \kappa^2)\sigma_{xx}a_{xy}^2 \right\}.$$
We find then that
\[ S_{yy\omega}(\omega) = (1 - \kappa^2)\sigma_{yy} (1 - 2a_{yy} \cos \omega + a^2_{yy})^{-1}. \]  

(31)

In this case, since the model order is \( p = 1 \), the point-frequency null hypothesis (Main Article, eq. 39) coincides with the time-domain null hypothesis Main Article, eq. 17 (i.e., \( a_{xy} = 0 \)), so that from Main Article, Theorem 2 we have
\[ N \hat{f}_Y \rightarrow \chi^2(\omega) \xrightarrow{d} \lambda(\omega) \cdot \chi^2(1), \]
where \( \lambda(\omega) = (1 - \kappa^2)\sigma_{yy} \omega_{yy} (1 - 2a_{yy} \cos \omega + a^2_{yy})^{-1} \), and the asymptotic distribution for the band-limited estimator may then be calculated as per (26) by integrating (31) across the appropriate frequency range.²

8. A RANDOM SAMPLING SCHEME FOR VAR MODEL PARAMETER SPACE

Consider, for a given number of variables \( n \) and model order \( p \), the parameter space \( \Theta = \{(A, \Sigma) : A \text{ is } n \times pn \text{ with } \rho(A) < 1, \Sigma \text{ is } n \times n \text{ positive-definite}\} \) of VAR(\( p \)) models. Firstly, we note that the residuals covariance matrix \( \Sigma \) can be taken to be a correlation matrix; this can always be achieved by a rescaling of variables leaving Granger causalities invariant. Further Granger causality invariances under linear transformation of variables (Barrett et al., 2010) allow further effective dimensional reduction of \( \Theta \); however, even under these general transformations, and under the constraint \( \rho(A) < 1 \), the quotient space of \( \Theta \) has finite Lebesgue measure; thus we cannot generate uniform variates (it is questionable whether this would in any case be appropriate to a given empirical scenario). Here we utilize a practical and flexible scheme for generation of variates on \( \Theta \), parametrized by spectral radius \( \rho \), log-generalized correlation \( \gamma = -\log |\Sigma| + \sum_i \log \Sigma_{ii} \), and population Granger causality \( F = F_Y \rightarrow \chi(\theta) \), all of which have a critical impact on Granger causality sampling distributions.

To generate a random correlation matrix \( \Sigma \) of dimension \( n \) with given generalized correlation \( \gamma \), we use the following algorithm:

1. Starting with an \( n \times n \) matrix with components iid \( \sim N(0, 1) \), compute its QR-decomposition \([Q, R]\). The matrix \( M_{ij} = Q_{ij} \cdot \text{sign}(R_{jj}) \) is then a random orthogonal matrix.
2. Create a random \( n \)-dimensional variance vector \( v \) with components \( v_i \) iid \( \sim \chi^2(1) \). The matrix \( V = M \cdot \text{diag}(v) \cdot M^T \) is then positive-definite, and for the corresponding correlation matrix \( \Sigma_{ij} = V_{ij} (V_{ii} V_{jj})^{-\frac{1}{2}} \) we have \( \gamma^* = -\sum_i \log v_i + \sum_i \log V_{ii} \).
3. If necessary, repeat steps 1,2 until \( \gamma^* \geq \gamma \) (this may fail if \( \gamma \) is too large).
4. Using a binary chop, find a constant \( c \) such that, iteratively replacing \( v \leftarrow v + c, \gamma^* \) falls within an acceptable tolerance of \( \gamma \) (this generally converges). The correlation matrix \( \Sigma \) is then returned.

For a VAR coefficients matrix sequence \( A = [A_1 \ A_2 \ \ldots \ A_p] \), the spectral radius \( \rho(A) \) is given by Main Article, eq. 3. If \( \lambda \) is a constant, it is easy to show that if \( A' \) is the sequence \( [\lambda A_1 \ \lambda^2 A_2 \ \ldots \ \lambda^p A_p] \), then \( \rho(A') = \lambda \rho(A) \). Thus any VAR coefficients sequence may be exponentially weighted so that its spectral radius takes a given value. Such weighting, however, we may use \( \int (1 - 2a \cos \omega + a^2)^{-1} d\omega = 2(1 - a^2)^{-1} \tan^{-1} \left( \frac{1 + a \tan \omega}{1 - a \tan \omega} \right) \).

² Although the space of \( n \times n \) correlation matrices has finite measure.
³ For Gaussian covariance matrices, log-generalized correlation coincides with multi-information (Studeny & Vejnarová, 1998). If \( R = (\rho_{ij}) \) is a correlation matrix with all \( \rho_{ij} \ll 1 \) for \( i \neq j \), then \( -\log |R| \approx \sum_{i<j} \rho_{ij}^2 \).
has the side-effect of exponential decay of the $A_k$ with lag $k$, which is, anecdotally, unrealistic\(^5\).

We observe empirically that we can compensate for this decay reasonably consistently across number of variables and model orders by scaling all coefficients by $A_k$ by $\exp \left\{ - \left( pw \right)^{\frac{1}{2}} \right\}$ for some constant $w$; here we choose $w = 1$, which generally achieves a more realistic gradual and approximately linear decay. To generate a random VAR model with given generalized correlation $\gamma$ and given spectral radius $\rho$, our procedure is as follows:

1. Generate a random correlation matrix $\Sigma$ with generalized correlation $\gamma$ as described above.
2. Generate $p n \times n$ coefficient matrices $A_k$ with components iid $\sim \mathcal{N}(0, 1)$. The $A_k$ are the weighted uniformly by $\exp \left\{ - \left( pw \right)^{\frac{1}{2}} \right\}$.

To enforce the null condition $A_{k,xy} = 0$,

3. Set all $A_{k,xy}$ components to zero.
4. Scale the $A_k$ coefficients sequence exponentially by an appropriate constant $\lambda$, so as to achieve the given spectral radius $\rho$.

To instead enforce a given (non-null) population Granger causality value $F$,

3. Scale the $A_{k,xy}$ components uniformly by a constant $c$.
4. Scale the $A_k$ coefficients sequence exponentially by an appropriate constant $\lambda$, so as to achieve the given spectral radius $\rho$.

Under steps 3, 4 the population Granger causality depends monotonically on $c$; consequently,

5. Perform a binary chop on $c$, iterating steps 3, 4 until the Granger causality is within an acceptable tolerance of $F$ (this generally converges quickly).

In all simulations except for the bivariate model (Section 7), we used $\gamma = 1$; spectral radii and population Granger causality values are as indicated in the plots. Convergence tolerances were set to $(\text{machine } \varepsilon)^{1/2} \approx 1.5 \times 10^{-8}$ under the IEEE 754-2008 binary64 floating point standard.

**References**

BARNETT, L. & SETH, A. K. (2015). Granger causality for state-space models. *Phys. Rev. E (Rapid Communications)* **91**, 040101(R).

BARRATT, A. B., BARNETT, L. & SETH, A. K. (2010). Multivariate Granger causality and generalized variance. *Phys. Rev. E* **81**, 041907.

HANNAN, E. J. & DEISTLER, M. (2012). *The Statistical Theory of Linear Systems*. Philadelphia, PA, USA: SIAM.

JONES, D. A. (1983). Statistical analysis of empirical models fitted by optimization. *Biometrika* **70**, 67–88.

LEHMANN, E. L. & ROMANO, J. P. (2005). *Testing Statistical Hypotheses*. New York, NY, USA: Springer Science+Business Media, LLC, 3rd ed.

MOHSENIPOUR, A. A. (2012). *On the Distribution of Quadratic Expressions in Various Types of Random Vectors*. Ph.D. thesis, The University of Western Ontario, Electronic Thesis and Dissertation Repository, 955.

SOLO, V. (2016). State-space analysis of Granger-Geweke causality measures with application to fMRI. *Neural Comput.* **28**, 914–949.

STUDENÝ, M. & VĚJNAROVÁ, J. (1998). The multiinformation function as a tool for measuring stochastic dependence. In *Learning in Graphical Models*, M. I. Jordan, ed. Dordrecht: Springer Netherlands, pp. 261–297.

VAN DER VAART, A. W. (2000). *Asymptotic Statistics*, vol. 3. Cambridge University Press.

---

\(^5\) At least, in the authors’ experience, for neural or econometric data.