Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Robust Bootstrap Test for Bidirectional Granger Causality via Unexplained Mutual Information

zouhaier dhifaoui

zouhaierdhifaoui@ymail.com

Universite de Sousse Faculte de Medecine de Sousse

Research Article

Keywords: Granger causality, mutual information, likelihood-ratio test, vector autoregression, bootstrap, robust statistics

Posted Date: September 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11060939/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

A Robust Bootstrap Test for Bidirectional Granger Causality via Unexplained Mutual Information

Zouhaier Dhifaoui[1]

[1]University of Sousse, Rue khalifa karoui, Sahloul 4, Tunisia.

Contributing authors: zouhaierdhifaoui@ymail.com;

Abstract

Detecting bidirectional Granger causality in a Gaussian vector autoregression (VAR) via likelihood-ratio or log-determinant statistics faces three gaps: no valid finite-sample test exists for the natural information-theoretic version of this statistic, joint restriction tests cannot distinguish one-directional from genuinely bidirectional causality, and the underlying covariance estimator is highly sensitive to outliers. We close this gap by defining the *unexplained mutual information* (UMI) between a VAR's residuals and observed variables, a log-determinant ratio that coincides with Geweke's (1982) linear-feedback measure and, once scaled, with the classical Gaussian LR statistic. We then build a moving-block bootstrap test imposing the null by construction, an intersection-union procedure isolating genuine bidirectional causality, and a robust variant based on Huber M-estimation withstanding outlier contamination. Monte Carlo experiments confirm correct size and power gains under contamination; an application to S&P 500 returns and VIX illustrates the method.

Keywords: Granger causality, mutual information, likelihood-ratio test, vector autoregression, bootstrap, robust statistics

1 Introduction

Since Granger (1969) turned Wiener's (1956) idea that one series is causal for another whenever its past improves the latter's prediction into an operational econometric procedure, Granger causality has become a standard tool for studying predictive relationships in multivariate time series (Lütkepohl 2007; Pearl 2009). In the Gaussian vector autoregressive (VAR) setting, the question is almost always answered with a classical F- or likelihood-ratio (LR) statistic comparing a restricted specification, in

which one series does not enter the equation for the other, against an unrestricted one that allows both directions of feedback (Sims 1972; Toda and Phillips 1993). Geweke's (1982) log-determinant decomposition of linear dependence offers an equivalent, information-theoretic reading of the same comparison, and it is this reading that we build on here.

Linear Granger-causality testing is not confined to establishing predictability in its own right. It can also serve as a specification diagnostic: Francis et al. (2010), for instance, use the linear causality remaining in VAR residuals to motivate, and later to validate, a nonlinear BEKK-GARCH specification. We keep to the linear Gaussian case throughout, but the connection to such diagnostic uses is one reason a correctly sized and genuinely bidirectional test is worth having in this simpler setting first.

Our starting point is the *unexplained mutual information* (UMI) between the residuals of a VAR model and the observed series, a log-determinant ratio of residual covariance matrices with a natural information-theoretic reading. This quantity turns out to coincide exactly with Geweke's measure of linear feedback and, once scaled by the sample size, with the classical Gaussian likelihood-ratio statistic for the corresponding restriction, so that its behaviour under the null follows a familiar χ^2 distribution rather than requiring new asymptotic theory. Turning this identity into a usable test still takes some care, for three reasons: the sample ratio is bounded below by one whether or not the null holds, so sampling uncertainty has to be handled through resampling rather than a fixed threshold; establishing bidirectional causality is logically stronger than rejecting a single joint restriction, since the latter can be driven by either direction alone; and the Gaussian covariance estimator on which the whole construction rests is well known to be sensitive to a handful of extreme observations, which time series in areas such as finance rarely lack. We address these three points in turn – with a null-imposing bootstrap, an intersection-union procedure, and a robust variant of the test – and evaluate the resulting procedures across VAR designs of increasing order.

The rest of the letter proceeds as follows. Section 2 defines UMI and establishes its identity with Geweke's measure and the Gaussian LR statistic. Section 3 details the null-imposing bootstrap test and the intersection-union procedure for bidirectional causality, together with the practical steps needed to apply them. Section 4 introduces a robust variant of the test for contaminated data. Section 5 reports Monte Carlo evidence on size and power, across lag orders and under outlier contamination. Section 6 applies the test to the joint dynamics of equity returns and market-implied volatility. Section 7 concludes. Proofs are gathered in the Appendix.

2 Unexplained mutual information in Gaussian VAR models

Assumption 1 The bivariate series $\{(Z_{1,t}, Z_{2,t})\}_{t=1}^{T}$ is second-order stationary and jointly Gaussian.

Assumption 2 $\{(Z_{1,t}, Z_{2,t})\}_{t=1}^{T}$ admits a correctly specified bivariate VAR representation of order $p \in \mathbb{N}^{+}$.

Let $\mathbf{Y} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_T)^\top$ collect the observations, $\mathbf{X}$ the matrix of lagged regressors including both series (the *unrestricted* specification), and $\widetilde{\mathbf{X}}$ the matrix containing only own lags (the *restricted* specification imposing no cross-series feedback). The two competing models are

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}^{\text{var}} + \mathbf{U}^{\text{var}}, \qquad \mathbf{Y} = \widetilde{\mathbf{X}}\boldsymbol{\theta} + \mathbf{U}^{\text{r-var}}, \tag{1}$$

estimated by OLS, equivalently Gaussian maximum likelihood.

Proposition 1 (Unexplained mutual information) *Under Assumptions 1–2, for* $\widehat{\mathbf{U}} \in \{\widehat{\mathbf{U}}^{\text{var}}, \widehat{\mathbf{U}}^{\text{r-var}}\}$ *the mutual information between the fitted residuals* $\widehat{\mathbf{U}}$ *and* $\mathbf{Y}$ *is*

$$\mathcal{M}_{\widehat{\mathbf{U}} \leftrightarrow \mathbf{Y}} = \frac{1}{2} \log \left(\frac{|\Sigma(\mathbf{Y})|}{|\Sigma(\mathbf{Y}) - \Sigma(\widehat{\mathbf{U}})|} \right). \tag{2}$$

The proof (Appendix A) follows from the Gaussian entropy decomposition $\mathcal{M}_{\widehat{\mathbf{U}} \leftrightarrow \mathbf{Y}} = \mathcal{H}(\mathbf{Y}) + \mathcal{H}(\widehat{\mathbf{U}}) - \mathcal{H}(\mathbf{Y} \oplus \widehat{\mathbf{U}})$ together with the OLS identity $\Sigma(\mathbf{Y}, \widehat{\mathbf{U}}) = \Sigma(\widehat{\mathbf{U}})$.

2.1 Relation to Geweke's measure and to the likelihood-ratio statistic

Define the ratio and its logarithm,

$$\widehat{R} = \frac{|\widehat{\Sigma}(\widehat{\mathbf{U}}^{\text{r-var}})|}{|\widehat{\Sigma}(\widehat{\mathbf{U}}^{\text{var}})|}, \qquad C = \log(\widehat{R}), \tag{3}$$

where $\widehat{\Sigma}(\cdot)$ denotes the Gaussian maximum-likelihood (i.e. N^{-1}-scaled) residual covariance estimator. By Proposition 1, $C = 2\big(\mathcal{M}_{\widehat{\mathbf{U}}^{\text{r-var}} \leftrightarrow \mathbf{Y}} - \mathcal{M}_{\widehat{\mathbf{U}}^{\text{var}} \leftrightarrow \mathbf{Y}}\big)$ is (twice) the loss in unexplained mutual information incurred by imposing the restriction.

Proposition 2 (Identity with Geweke's feedback measure and the LR statistic) *The quantity* C *in* (3) *coincides exactly with Geweke's (1982) total measure of linear feedback between the two series,* $C = F_{Z_1 \leftrightarrow Z_2}$. *Moreover, when the two models in* (1) *are estimated over the same observations and evaluated at their Gaussian maximum-likelihood estimates, the classical likelihood-ratio statistic for the joint hypothesis that all* r *cross-lag coefficients are zero satisfies*

$$\text{LR} = N\,C = N \log(\widehat{R}) \xrightarrow{d} \chi^2_r \quad \textit{under } H_0, \tag{4}$$

where N *is the effective sample size and* r *the number of cross-lag restrictions* ($r = p_1 + p_2$ *when the restricted model uses own-lag orders* p_1, p_2).

The proof (Appendix A) is the standard Gaussian LR derivation applied to (1); the identity with Geweke's measure follows directly from his log-determinant definition of linear feedback. Proposition 2 shows that C does not carry an independent notion of "causality strength" beyond the classical one: it is a rescaling of the same log-likelihood

comparison that underlies the Gaussian VAR LR test, and its exact finite-sample and asymptotic behaviour is governed by (4). This is a useful state of affairs rather than a limitation: it shows precisely why $C > 0$ signals a better-fitting unrestricted model, and it supplies, at no extra cost, a χ^2 benchmark against which any inferential procedure for C can be judged.

Corollary 3 *If $\{Z_{1,t}\}$ and $\{Z_{2,t}\}$ are related by Granger causality in at least one direction, then $\widehat{R} > 1$ in the population limit, i.e. $|\Sigma(\mathbf{U}^{\text{var}})| < |\Sigma(\mathbf{U}^{\text{r-var}})|$ (Appendix A).*

Remark 4 (Why $\widehat{R} > 1$ cannot, by itself, define a test) *Because $\widetilde{\mathbf{X}}$ is nested in $\mathbf{X}$, OLS residual sums of squares are monotone in the regressor set, so $\widehat{R} \geq 1$ holds* in every finite sample*, whether or not H_0 is true. Comparing the sample statistic $\widehat{R}$ to its population boundary of* 1*, and rejecting whenever $\widehat{R} > 1$, therefore has an uncontrolled (and generally very high) rejection frequency even under the null; it is not a valid significance test. A test requires comparing $\widehat{R}$, or a monotone transform of it, to its* sampling *distribution under H_0, which is what* (4) *and the bootstrap procedure of Section 3 provide.*

Remark 5 (Joint restriction versus bidirectional causality) *Rejecting the joint hypothesis that all cross-lag coefficients vanish establishes that Granger causality runs in* at least one *direction; it does not establish that it runs in* both*. A test of bidirectional causality requires the intersection-union combination of the two one-directional hypotheses, $H_0^{(1\to2)}$ and $H_0^{(2\to1)}$, each tested separately, with H_0 (no bidirectional causality) rejected only if both directional nulls are rejected. We implement this alongside the joint test in Sections 3–5.*

3 The UMI causality test: definition and implementation

This section gives a complete, self-contained description of the test, in a form that a practitioner can implement directly: the hypotheses being tested, the null-imposing bootstrap algorithm that calibrates the joint statistic, and the intersection-union procedure that turns the joint test into a genuine test of bidirectional causality.

3.1 Hypotheses

Two distinct hypotheses are of interest, and it matters which one a given procedure actually tests (Remark 5). The *joint* hypothesis, H_0^{joint}, states that all cross-lag coefficients in (1) are zero, i.e. that neither series Granger-causes the other; rejecting it establishes causality in at least one direction. The two *directional* hypotheses are

$$\begin{aligned} H_0^{(2\to1)} &: \text{ the lagged-}Z_2 \text{ coefficients} \\ &\quad\ \text{vanish in the equation for } Z_1, \\ H_0^{(1\to2)} &: \text{ the lagged-}Z_1 \text{ coefficients} \\ &\quad\ \text{vanish in the equation for } Z_2, \end{aligned}$$

and bidirectional causality is the assertion that both $H_0^{(2\to1)}$ and $H_0^{(1\to2)}$ are false simultaneously – an intersection-union alternative, tested by rejecting each directional null on its own merits.

3.2 Null-imposing moving-block bootstrap

The finite-sample distribution of $C = \log \widehat{R}$ under dependence is analytically intractable outside the Gaussian asymptotic regime of (4), so we calibrate the joint test by a residual-based moving-block bootstrap that imposes H_0^{joint} by construction, rather than resampling the raw data. Given the observed series $\mathbf{Y}$ of length N (after allowing for p lags) and a chosen block length ℓ and number of bootstrap replications B:

1. **Fit the restricted model.** Estimate $\widetilde{\mathbf{X}}\boldsymbol{\theta}$ in (1) by OLS, obtaining own-lag coefficient estimates $\widehat{\theta}_1, \widehat{\theta}_2$ (one set per equation) and restricted residuals $\widehat{\mathbf{U}}^{\text{r-var}} = (\widehat{u}_1, \ldots, \widehat{u}_N)$.
2. **Compute the observed statistic.** Fit the unrestricted model as well, and set $C_{\text{obs}} = \log \widehat{R}$ from (3).
3. **Resample the restricted residuals.** For $b = 1, \ldots, B$, draw overlapping blocks of length ℓ from $\{\widehat{u}_t\}_{t=1}^N$ with replacement (moving-block bootstrap) and concatenate them into a resampled sequence $\{\widehat{u}_t^{*(b)}\}_{t=1}^N$ of the same length.
4. **Regenerate synthetic data under H_0^{joint}.** Starting from the first p observed values, generate $\mathbf{Y}^{*(b)}$ recursively from the *restricted* recursion

$$Y_{j,t}^{*(b)} = \widehat{\theta}_j(L)\, Y_{j,t-1}^{*(b)} + \widehat{u}_{j,t}^{*(b)}, \qquad j = 1, 2,$$

   which contains no cross-series lag by construction, so H_0^{joint} holds exactly in every replicate regardless of what holds in the original data.
5. **Refit and recompute.** Fit both the unrestricted and restricted models to $\mathbf{Y}^{*(b)}$ and compute $C^{*(b)} = \log \widehat{R}^{*(b)}$.
6. **Bootstrap p-value.** Set $p = (1 + \#\{b : C^{*(b)} \geq C_{\text{obs}}\})/(B+1)$ and reject H_0^{joint} at level α when $p < \alpha$.

Because H_0^{joint} is imposed by construction in step 4, this procedure targets the correct null distribution of C, unlike a rule based on the sign of $\widehat{R} - 1$ (Remark 4). Two implementation choices matter in practice. The block length ℓ should be at least as large as the estimated own-lag order p, so that the resampling scheme can reproduce the short-range dependence the fitted model is meant to capture; a block length shorter than p tends to understate the true variability of C under H_0. The number of replications B trades off Monte Carlo error in the p-value against computation: $B = 199$ is typically enough for decisions at $\alpha = 0.05$, while $B = 999$ or more is preferable when the p-value is close to α, as in the application of Section 6. Computationally, each bootstrap replicate requires one OLS fit of the restricted and one of the unrestricted model, so the cost of the whole procedure is $O(B)$ least-squares fits of size $O(N)$, which is fast even for B in the thousands.

3.3 Intersection-union test of bidirectional causality

Rejecting H_0^{joint} does not by itself establish bidirectional causality (Remark 5); it establishes causality in at least one direction. To test bidirectional causality directly, apply the following intersection-union procedure.

1. Test $H_0^{(2\to1)}$ using the single-equation restriction that the lagged-Z_2 coefficients vanish in the equation for Z_1: this can be done with the corresponding t- or F-statistic under Gaussian errors, or, if the same resampling-based calibration is preferred throughout, with the bootstrap of Section 3.2 applied to this single equation. Obtain a p-value $p^{(2\to1)}$.
2. Test $H_0^{(1\to2)}$ analogously, obtaining $p^{(1\to2)}$.
3. Reject the null of no bidirectional causality at level α only if both $p^{(2\to1)} < \alpha$ and $p^{(1\to2)} < \alpha$.

This is a standard intersection-union test: its size is bounded above by α without any correction for multiplicity, because rejection requires both directional nulls to be individually rejected.

4 A robust variant for contaminated data

The construction in Sections 2–3 rests on the Gaussian sample covariance of the VAR residuals, both in the OLS fits and in $\widehat{\Sigma}(\cdot)$. This is exactly what delivers the identity with Geweke's measure and the LR statistic (Proposition 2), but it is also what makes the test sensitive to a small fraction of extreme observations: a single additive outlier can dominate a sample covariance matrix, inflating or deflating $\widehat{R}$ in either direction. Time series in domains such as finance, where the present application lies, routinely contain such observations, so we develop a robust variant of the same test.

The variant keeps the log-determinant statistic and the null-imposing bootstrap of Section 3 unchanged in structure, and replaces only the two estimation steps that are not robust to outliers:

- each equation of the unrestricted and restricted models in (1) is fitted by Huber's (1981) M-estimator (iteratively reweighted least squares with the Huber ψ-function) instead of OLS, which bounds the influence of any single observation on the estimated coefficients;
- the residual covariance matrix $\widehat{\Sigma}(\cdot)$ is replaced by the robust location-and-covariance estimator of Lange et al. (1989), which fits a multivariate t-distribution to the residual pairs by an EM-type algorithm and downweights residual vectors far from the bulk of the data.

Write $\widehat{R}_{\text{rob}}$ and $C_{\text{rob}} = \log \widehat{R}_{\text{rob}}$ for the resulting ratio and its logarithm. The bootstrap algorithm of Section 3.2 applies verbatim to C_{rob}, with the robust fits used in place of OLS at every step, including inside the bootstrap loop; only the estimation, not the resampling logic, changes.

Two points are worth making explicit. First, robustifying the estimator breaks the exact algebraic identity of Proposition 2: C_{rob} is no longer, in general, N^{-1} times a χ^2-distributed likelihood-ratio statistic, so the robust test relies on the bootstrap for

calibration rather than on (4). Second, robustness is not free: bounding the influence of extreme observations also reduces the efficient use of the data when no such observations are present, so the robust test is expected to have somewhat lower power than the Gaussian version on clean data, in line with the well-documented efficiency cost of M-estimation (Maronna et al. 2006).

5 Monte Carlo evidence

We evaluate size and power at nominal level $\alpha = 0.05$ in three sets of experiments: (i) the main comparison of the bootstrap test (UMI-boot), the asymptotic χ^2 likelihood-ratio test (LR), and the intersection-union directional test (IU) in a VAR(1) design; (ii) the same joint tests in VAR(2) and VAR(3) designs, to check that the procedure is not an artefact of the first-order case; and (iii) a comparison of the Gaussian and robust bootstrap tests under outlier contamination.

5.1 Size and power in a VAR(1) design

Data are generated from a bivariate Gaussian VAR(1),

$$\mathbf{Z}_t = \phi_1 \mathbf{Z}_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N\left(\mathbf{0}, \begin{pmatrix} 0.8 & 0.1 \\ 0.1 & 0.8 \end{pmatrix}\right),$$

under four designs for ϕ_1: (i) H_0, diagonal only (diag $= 0.9, 0.7$, off-diagonal $= 0$); (ii) a one-directional alternative ($Z_2 \to Z_1$ only, off-diagonal $0.3/0$); (iii) a weak bidirectional alternative (off-diagonal $0.15/-0.10$); (iv) a strong bidirectional alternative (off-diagonal $0.30/-0.20$). For each of $N \in \{50, 100, 200\}$ we use 1000 Monte Carlo replications and $B = 199$ bootstrap replications with block length $\ell = 5$. Results are given in Table 1

Three findings stand out. First, the asymptotic LR test is moderately oversized at small N (11.0% and 8.5% at $N = 50, 100$ against a nominal 5%), a well-known finite-sample distortion of χ^2 approximations, whereas the bootstrap test controls size closely (at or below nominal, if conservatively so). Second, power of both joint tests (UMI-boot, LR) rises smoothly with N and with the strength of the cross-lag coefficients – exactly the pattern a consistent test should display, and one that a rule based only on the sign of $\widehat{R} - 1$ could never produce, since that ratio exceeds one regardless of sample size (Remark 4). Third, and central to Remark 5, the IU procedure keeps its rejection rate close to its nominal size under the purely one-directional design (11.5%, 7.5%, 5.5% for $N = 50, 100, 200$) even though the joint tests reject that same design very frequently (78–100%): the joint restriction test cannot distinguish one-directional from bidirectional causality, while the intersection-union test can.

5.2 Higher-order VAR designs

The theory of Section 2 and the bootstrap of Section 3 are stated for a general lag order p, but it is worth checking directly that the joint test continues to perform as expected once $p > 1$, since a practitioner rarely knows in advance that a first-order VAR is the

Table 1 Empirical rejection frequencies (%) at nominal level 5%. Size is reported for H_0; power for the three alternatives. UMI-boot: null-imposing bootstrap test (Sec. 3); LR: asymptotic χ^2_2 Gaussian likelihood-ratio test; IU: intersection-union directional test.

N	Design	UMI-boot	LR (χ^2_2)	IU
50	H_0	0.5	11.0	1.0
	One-directional	6.0	78.0	11.5
	Bidirectional, weak	5.0	56.0	17.5
	Bidirectional, strong	47.5	98.5	87.5
100	H_0	0.0	8.5	0.5
	One-directional	21.5	96.0	7.5
	Bidirectional, weak	20.5	83.0	41.5
	Bidirectional, strong	92.5	100.0	98.5
200	H_0	0.5	6.5	0.0
	One-directional	57.0	100.0	5.5
	Bidirectional, weak	70.5	98.0	84.0
	Bidirectional, strong	100.0	100.0	100.0

right specification. We simulate VAR(p) designs for $p = 1, 2, 3$ at $N = 100$, with own-lag persistence spread across lags and a cross-series effect entering only at the first lag (off-diagonal $0.30/-0.20$, zero at further lags), so that the restricted model in each case still imposes $2p$ zero restrictions even though the true cross-effect is short-lived. We use 1000 Monte Carlo replications and $B = 199$ bootstrap replications, a reduced design adequate to check the qualitative pattern. Results are shown in Table 2.

Table 2 Empirical rejection frequencies (%) of the joint UMI-bootstrap test at nominal level 5%, $N = 100$, for VAR designs of order $p = 1, 2, 3$. The restricted model imposes $2p$ zero restrictions in every case.

Design	$p = 1$	$p = 2$	$p = 3$
H_0 (size)	0.0	0.0	0.0
Bidirectional, strong (power)	88.3	98.3	100.0

Size remains at zero rejections in all three designs, if anything conservative, and power stays high and does not deteriorate as p grows, even though the number of nuisance restrictions being tested jointly with the true one doubles and triples. This is consistent with the restricted model correctly imposing zero coefficients at every lag beyond the first, where the true generating process also has none, so the additional restrictions cost little. We take this as evidence that the test is not tailored to the first-order case used in Section 5.1, though we would expect power to fall, as usual, if the true cross effect became weak or diffuse over many lags relative to N.

5.3 Robustness to outlier contamination

We compare the Gaussian bootstrap test (UMI-boot) of Section 3 with its robust counterpart (Robust-UMI-boot) of Section 4 under the H_0 and strong-bidirectional VAR(1) designs of Section 5.1, with and without outlier contamination: 20% of the observations in each series are independently perturbed by additive noise at SNR $= -5$dB. We use ($N = 50$: 500 Monte Carlo, $B = 199$; $N = 100$: 500 Monte Carlo, $B = 199$). Results are given in Table 3.

Table 3 Empirical rejection frequencies (%) at nominal level 5% under the strong-bidirectional VAR(1) design, with and without 20% outlier contamination (SNR $= -5$dB). UMI-boot uses OLS and the Gaussian sample covariance; Robust-UMI-boot uses Huber M-estimation and the multivariate-t covariance of Lange et al. (1989).

N	Contamination	UMI-boot	Robust-UMI-boot
50	None (clean)	34.0	16.0
50	20% outliers	10.0	4.0
100	None (clean)	83.3	46.7
100	20% outliers	10.0	16.7

On clean data the robust test is, as expected, less powerful than the Gaussian one at both sample sizes: bounding the influence of individual observations costs efficiency when there is nothing to be robust to. Under contamination, the picture changes. At $N = 100$, power of the standard test collapses from 83.3% to 10.0%, while the robust test falls much less steeply, from 46.7% to 16.7%, and ends up *more* powerful than the standard test once outliers are present. At $N = 50$ the same qualitative pattern of a smaller relative decline appears for the robust test (from 16.0% to 4.0%, versus 34.0% to 10.0% for the standard test). Overall, the robust variant is the one we would recommend whenever contamination is a realistic concern and N is not very small, while the Gaussian test remains preferable on data believed to be clean, where it is both more powerful and backed by the exact χ^2 benchmark of Proposition 2.

6 Application: stock returns and volatility dynamics

The robust-to-outliers bidirectional test developed in Section 4 is applied to the joint dynamics of equity returns and market-implied volatility, a relationship for which the finance literature furnishes two distinct, and in principle simultaneously operative, causal mechanisms. The leverage-effect hypothesis of Black (1976) and Christie (1982) posits a return-to-volatility channel: a decline in equity value mechanically raises financial leverage, which in turn elevates the conditional variance of subsequent returns. The volatility-feedback hypothesis of French et al. (1987) and Campbell and Hentschel (1992), subsequently embedded by Bekaert and Wu (2000) within an asymmetric conditional-variance framework, instead posits a volatility-to-return channel:

an upward revision in anticipated volatility must be compensated by a higher equity risk premium, which depresses contemporaneous prices. Because these mechanisms operate in opposite directions, their joint operation constitutes a canonical instance of bidirectional Granger causality, motivating the empirical illustration below.

The analysis uses the first difference of the logarithm of the S&P 500 index and of the CBOE Volatility Index (VIX), Whaley's (2000) "investor fear gauge" and the most widely used empirical proxy for market-implied volatility (Giot 2005). The two series comprise $N = 68$ daily observations spanning 16 January–21 April 2026, obtained from the FRED database. Differencing the logarithm of each index induces stationarity and avoids applying the VAR framework directly to the highly persistent levels of the underlying series. The series are shown in Figure 1, together with the outlier observations detected by the MAD method in each series (one per series, marked with a gold circle). Both transformed series fail to reject normality under the robust-to-outliers Jarque–Bera test (S&P 500: statistic $= 4.451$, $p = 0.108$; VIX: statistic $= 1.809$, $p = 0.404$), and both reject the unit-root null under the outlier-robust KPSS test (Darne 2004) (S&P 500: statistic $= 0.081$, $p = 0.1$; VIX: statistic $= 0.066$, $p = 0.1$), supporting the distributional and stationarity assumptions of the VAR specification.

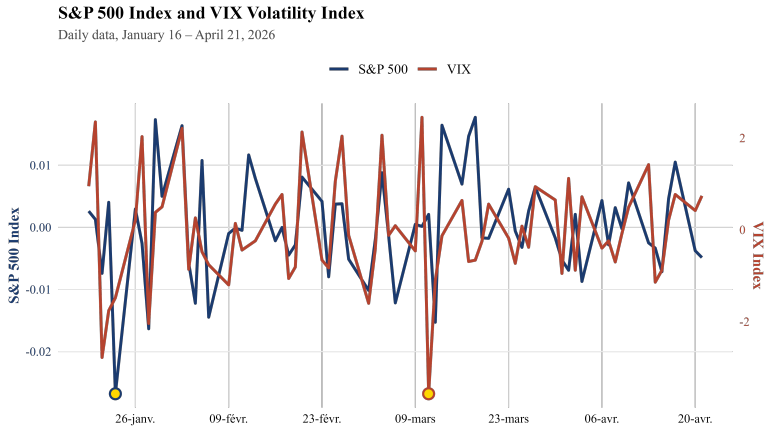


Fig. 1 First-differenced logarithms of the S&P 500 index and the VIX, 16 January–21 April 2026.

Lag-order selection by the Akaike and final prediction error criteria unanimously selects $p = 3$. The bootstrap block length $\ell = 5$ therefore exceeds p, satisfying the requirement $\ell \geq p$ of Section 3.2 for the resampling scheme to preserve the estimated autoregressive dependence.

The robust UMI-boot test ($B = 999$ replications) yields a statistic of 1.914 and a p-value of 0.035, rejecting the null of bidirectional Granger noncausality. This rejection does not by itself imply bidirectional causality, since it only establishes that at least one direction is active. We therefore applied the robust UI test of Section 3.3, based on robust VAR regression and robust covariance estimation as in Section 5.3. The robust UI test ($B = 999$ replications) gives a statistic of 0.916 from the S&P 500 to

the VIX ($p = 0.002$), and -0.156 from the VIX to the S&P 500 ($p = 0.897$), indicating significant causality running from returns to volatility, but not the reverse.

This asymmetric pattern favours the leverage-effect channel of Black (1976) and Christie (1982) over the volatility-feedback channel in this sample, in line with the high-frequency evidence of Bollerslev et al. (2006) that leverage effects dominate on impact while feedback effects operate more slowly, and consistent with Bae et al. (2007) and Hibbert et al. (2008), who attribute the empirical dominance of the return-to-volatility direction to asymmetric investor reactions rather than to leverage alone.

7 Concluding remarks

The unexplained mutual information statistic for detecting bidirectional Granger causality is, up to a known scaling, identical to Geweke's linear feedback measure and to the classical Gaussian likelihood-ratio statistic, which supplies a χ^2 benchmark for the joint hypothesis. Because the underlying sample ratio is bounded below by one whether or not the null holds, a valid test requires comparing it to its actual sampling distribution rather than to this population boundary; we obtain that distribution through a bootstrap procedure that imposes the null hypothesis explicitly, detailed in Section 3 together with the practical choices (block length, number of replications) it involves. Distinguishing this joint restriction test from a genuine test of bidirectional causality matters in practice: our simulations show the joint test cannot separate one-directional from bidirectional causality, while the intersection-union combination of directional tests can, at a cost in power that a practitioner should weigh explicitly. The same simulations confirm the procedure is not specific to first-order VAR designs, retaining correct size and rising power in VAR(2) and VAR(3) specifications.

Because the Gaussian covariance estimator at the core of the test is sensitive to outliers, we also developed a robust variant based on Huber M-estimation and a multivariate-t covariance estimator. It is less powerful than the Gaussian test on clean data, as expected of any robust procedure, but recovers a clear power advantage once contamination is present and the sample is not too small, making it the more suitable choice whenever outliers are a realistic concern in the data at hand.

Extending the framework to non-Gaussian innovations, to higher-dimensional VAR systems, and to alternative robust covariance estimators (e.g. minimum covariance determinant or S-estimators) are natural next steps, along with a larger-scale simulation study to sharpen the finite-sample comparison between the Gaussian and robust tests reported here.

Use of generative AI tools

The author(s) used a large language model to assist with English-language editing and LaTeX formatting during the preparation of this manuscript; all methodological content, derivations, and numerical results were produced and verified by the author(s).

Appendix A Proofs

Proof of Proposition 1.

For the linear Gaussian model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with OLS residual $\widehat{\boldsymbol{\epsilon}}$, the normal equations give $\Sigma(\mathbf{Y}, \widehat{\boldsymbol{\epsilon}}) = \Sigma(\mathbf{Y}) - \Sigma(\mathbf{Y}, \mathbf{X})\Sigma(\mathbf{X})^{-1}\Sigma(\mathbf{X}, \mathbf{Y}) = \Sigma(\widehat{\boldsymbol{\epsilon}})$. For jointly Gaussian $\mathbf{Y}, \widehat{\boldsymbol{\epsilon}}$, $\mathcal{H}(\mathbf{Y}) = \frac{1}{2}\log|\Sigma(\mathbf{Y})| + \text{const.}$, and using the block-determinant identity for the joint covariance of $(\mathbf{Y}, \widehat{\boldsymbol{\epsilon}})$ together with $\Sigma(\mathbf{Y}, \widehat{\boldsymbol{\epsilon}}) = \Sigma(\widehat{\boldsymbol{\epsilon}})$ gives $|\Sigma(\mathbf{Y} \oplus \widehat{\boldsymbol{\epsilon}})| = |\Sigma(\widehat{\boldsymbol{\epsilon}})|\,|\Sigma(\mathbf{Y}) - \Sigma(\widehat{\boldsymbol{\epsilon}})|$. Substituting into $\mathcal{M}_{\widehat{\boldsymbol{\epsilon}} \leftrightarrow \mathbf{Y}} = \mathcal{H}(\mathbf{Y}) + \mathcal{H}(\widehat{\boldsymbol{\epsilon}}) - \mathcal{H}(\mathbf{Y} \oplus \widehat{\boldsymbol{\epsilon}})$ and cancelling the constants yields (2). The same argument applied separately to $\widehat{\mathbf{U}}^{\text{var}}$ and $\widehat{\mathbf{U}}^{\text{r-var}}$ gives the VAR and restricted-VAR versions used in (3). ■

Proof of Proposition 2.

Geweke's (1982) total measure of linear feedback between Z_1 and Z_2 is defined as the log-ratio of the generalized variance of the residuals from the model excluding cross-series lags to that of the model including them, i.e. $F_{Z_1 \leftrightarrow Z_2} = \log\big(|\Sigma(\mathbf{U}^{\text{r-var}})|/|\Sigma(\mathbf{U}^{\text{var}})|\big)$, which is exactly C in (3). For the LR identity, the Gaussian log-likelihood of model (1), concentrated at the ML (OLS) estimates, is $\ell(\widehat{\Sigma}) = -\frac{N}{2}\log|\widehat{\Sigma}| + \text{const.}$, so $\text{LR} = 2\big(\ell^{\text{var}} - \ell^{\text{r-var}}\big) = N\big(\log|\widehat{\Sigma}(\widehat{\mathbf{U}}^{\text{r-var}})| - \log|\widehat{\Sigma}(\widehat{\mathbf{U}}^{\text{var}})|\big) = N\log\widehat{R} = NC$. Under H_0 and standard regularity conditions for nested Gaussian models, $\text{LR} \xrightarrow{d} \chi^2_r$ with r equal to the number of zero restrictions imposed. ■

Proof of Corollary 3.

If Granger causality holds in at least one direction, the unrestricted model attains strictly lower population residual variance in the corresponding equation(s), so $\Sigma(\mathbf{U}^{\text{var}}) \preceq \Sigma(\mathbf{U}^{\text{r-var}})$ with strict inequality in trace. Writing $\Sigma(\mathbf{Y}) = R^\top R$ and $\widetilde{B} = R^{-1}\Sigma(\mathbf{U}^{\text{r-var}})R^{-\top}$, $\widetilde{C} = R^{-1}\Sigma(\mathbf{U}^{\text{var}})R^{-\top}$, both positive semi-definite with eigenvalues in $[0, 1)$, the trace inequality together with $|\Sigma(\mathbf{U})| = |\Sigma(\mathbf{Y})|\,|\widetilde{(\cdot)}|$ gives $|\Sigma(\mathbf{U}^{\text{var}})| < |\Sigma(\mathbf{U}^{\text{r-var}})|$, i.e. $\widehat{R} \to R > 1$. ■

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

Data will be made available on request.

Declaration of interests

The authors have no conflicts of interest to disclose.

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

□ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

References

Aït-Sahalia, Y., Fan, J., Li, Y., 2013. The leverage effect puzzle: disentangling sources of bias at high frequency. Journal of Financial Economics 109, 224–249.

Bae, K.-H., Kim, D., Nelson, C.R., 2007. Why are stock returns and volatility negatively correlated? Journal of Empirical Finance 14, 41–58.

Badshah, I.U., 2013. Quantile regression analysis of the asymmetric return-volatility relation. Journal of Futures Markets 33, 235–265.

Bekaert, G., Wu, G., 2000. Asymmetric volatility and risk in equity markets. Review of Financial Studies 13, 1–42.

Bollerslev, T., Litvinova, J., Tauchen, G., 2006. Leverage and volatility feedback effects in high-frequency data. Journal of Financial Econometrics 4, 353–384.

Black, F., 1976. Studies of stock price volatility changes. Proceedings of the 1976 Meetings of the Business and Economic Statistics Section, American Statistical Association, 177–181.

Campbell, J.Y., Hentschel, L., 1992. No news is good news: An asymmetric model of changing volatility in stock returns. Journal of Financial Economics 31, 281–318.

Christie, A.A., 1982. The stochastic behavior of common stock variances: value, leverage and interest rate effects. Journal of Financial Economics 10, 407–432.

Darné, O., 2004. The effects of additive outliers on stationarity tests: a monte carlo study, Economics Bulletin, AccessEcon, vol. 3(16), 1–8.

Francis, B.B., Mougoué, M., Panchenko, V., 2010. Is there a symmetric nonlinear causal relationship between large and small firms? Journal of Empirical Finance 17, 23–38.

French, K.R., Schwert, G.W., Stambaugh, R.F., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Geweke, J., 1982. Measurement of linear dependence and feedback between multiple time series. Journal of the American Statistical Association 77, 304–313.

Giot, P., 2005. Relationships between implied volatility indexes and stock index returns. Journal of Portfolio Management 31(3), 92–100.

Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438.

Hacker, R.S., Hatemi-J, A., 2006. Tests for causality between integrated variables using asymptotic and bootstrap distributions: theory and application. Applied Economics 38, 1489–1500.

Hibbert, A.M., Daigler, R.T., Dupoyet, B., 2008. A behavioral explanation for the negative asymmetric return–volatility relation. Journal of Banking & Finance 32, 2254–2266.

Huber, P.J., 1981. Robust Statistics. Wiley, New York.

Künsch, H.R., 1989. The jackknife and the bootstrap for general stationary observations. Annals of Statistics 17, 1217–1241.

Lange, K.L., Little, R.J.A., Taylor, J.M.G., 1989. Robust statistical modeling using the t distribution. Journal of the American Statistical Association 84, 881–896.

Lütkepohl, H., 2007. New Introduction to Multiple Time Series Analysis. Springer, Berlin.

Maronna, R.A., Martin, R.D., Yohai, V.J., 2006. Robust Statistics: Theory and Methods. Wiley, Chichester.

Pearl, J., 2009. Causality: Models, Reasoning, and Inference, 2nd ed. Cambridge University Press, Cambridge.

Sims, C.A., 1972. Money, income, and causality. American Economic Review 62, 540–552.

Toda, H.Y., Phillips, P.C.B., 1993. Vector autoregressions and causality. Econometrica 61, 1367–1393.

Whaley, R.E., 2000. The investor fear gauge. Journal of Portfolio Management 26(3), 12–17.

Wiener, N., 1956. The theory of prediction, in: Beckenbach, E.F. (Ed.), Modern Mathematics for the Engineer. McGraw-Hill, New York.