TI 2026-098/III
Tinbergen Institute Discussion Paper

A Local Nonparametric Test for Granger Causality

Sicco Kooiker [1]

1 Vrije Universiteit Amsterdam

A Local Nonparametric Test for Granger Causality

Sicco Kooiker*

This version: November 26, 2026

Abstract

Granger causality between two time series may only be observable in some periods, and the distribution of economic and financial time series may change over time. The nonparametric Diks–Panchenko (DP) test pools the distributions at all points in time, and measures conditional dependence in the time-averaged distribution. As a result, it can lose power when Granger causality is only present in some periods, and it can reject when the distribution changes over time, although there is no Granger causality at any point in time. We propose a local version of the DP test that weights the observations with exponentially decaying weights around each point in time. We derive its asymptotic normal distribution and derive a rule of thumb for the bandwidth. In simulations, we show that the local DP test has better power when Granger causality is only present in short periods, and that its size is more robust when the distribution changes over time. For S&P 500 returns and volume changes, the local DP test shows that the evidence for Granger causality from returns to volume is present in most of the sample, whereas after volatility filtering the most robust evidence from volume to returns is in the early 1960s.

Keywords: Time-varying Granger causality, Nonparametric Granger causality, Density estimation, Exponential smoothing

JEL Classification: C12, C14, C32, C58, G12

*Kooiker: Vrije Universiteit Amsterdam (email: s.h.kooiker@vu.nl).

1 Introduction

Granger causality (GC), a foundational concept in time series analysis, examines whether one time series provides predictive power for the future values of another time series beyond the information contained in the latter's values, similar to Florens and Mouchart (1982). Initially, this concept was introduced in a two-dimensional parametric setting (Granger, 1969). This method has found significant application in fields such as econometrics, genomics, and neuroscience (Shojaie and Fox, 2022). When combined with advances in (Structural) Vector Autoregressive modeling (Lütkepohl, 2005; Kilian, 2013), GC provides a powerful framework for assessing time-dependent systems. However, two major caveats of these approaches are that they assume the system to be linear and constant over time. In this paper, we address both and develop a local nonparametric test for GC.

Various authors have proposed alternatives to relax the linear restriction in GC testing, among them: GC in the quantile regression domain (Troster, 2018) and GC in parametric nonlinear regression models (Chen et al., 2004; Wismüller et al., 2021). These approaches assume correct model specification. An alternative is nonparametric GC. The negation of GC, known as Granger noncausality (GNC), is a form of conditional independence. One of the earliest tests for nonparametric GNC is proposed by Hiemstra and Jones (1994) (HJ). The HJ test is a modified version of the Baek and Brock (1992) test for conditional independence. Although frequently used, Diks and Panchenko (2006) (DP) point out that the HJ test is biased, as it overrejects the null hypothesis. To address this issue, they introduce an improved test, which we refer to as the DP test, and provide a rule of thumb for selecting the appropriate bandwidth in financial applications.

DP estimates the required probability density functions with the local density estimator (LDE) by Broock et al. (1996), a kernel density estimator (KDE) with a uniform kernel. The DP test pools the distributions at all points in time, and measures conditional dependence in the time-averaged distribution. This has two consequences. When GC is only present in some periods, the causal periods are mixed with the non-causal ones. This weakens the signal and does not reveal when the causality is present. Additionally, when the distributions change over time, potentially in response to an external event, the static DP test can reject although GNC holds at every point in time.

In the parametric econometric literature, time-varying GC is already established (Phillips et al., 2011, 2015a,b; Shi et al., 2020). The studies propose either a rolling or recursive estimation window. Setting the appropriate window size in a rolling window setting is challenging and the rough edges are less suitable for smoothly changing conditions. Alternatively, Harvey

and Oryshchenko (2012) apply the Kalman smoother with exponentially weighted moving average (EWMA) updating equations to KDE, an efficient algorithm that allows for time-varying estimation. While rolling windows require a fixed window size, EWMA allows for smooth transitions between time periods.

We extend LDE estimation using the EWMA scheme to nonparametrically estimate time-varying probability density functions, resulting in a sequence of LDEs over time. Where the DP test takes an average over the resulting dependence measure, we propose to replace this average by EWMA as well. At each time point, we apply an EWMA version of the DP test, generating a sequence of weighted DP statistics. We call the resulting test the local DP test.

We make two contributions. First, we show that, when the distribution changes over time, the static DP test tests GNC in the distribution pooled over the sample, and we characterize when this hypothesis differs from GNC at each point in time. Second, we derive the asymptotic normal distribution of the weighted statistic, following the approach of DP. The Kish effective sample size of the weights takes the role of the sample size and gives the corresponding rule of thumb for the bandwidth.

In a simulation study, we show that, when a common state changes the distributions, the static DP test rejects although GNC holds at every point in time. The local DP test stays close to the nominal level, except near an abrupt break, and locates periods of GC, including short episodes that the static DP test misses. We then revisit the relation between stock returns and trading volume studied by Hiemstra and Jones (1994) and DP, using daily S&P 500 data from 1950 to 2020. The static DP test rejects GNC in both directions. The local DP test shows that the evidence from returns to volume is spread over the whole sample and stays present after volatility filtering, while the evidence from volume to returns is more sensitive to volatility filtering and, after filtering, is concentrated in the early 1960s and the early 2000s.

Our paper also relates to the broader literature on nonparametric tests for GNC. Besides the HJ and DP tests, nonparametric GNC tests are proposed by Bell et al. (1996), Su and White (2008), and Bouezmarni et al. (2012). Like the DP test, these tests measure conditional dependence over the full sample. Machine learning methods for GC detection, such as those in Marinazzo et al. (2008), Vicente et al. (2011), and Tank et al. (2022), are gaining popularity. A problem with these methods is that they do not rely on statistics with a known limit distribution, and therefore do not control the type-I error. GNC is also closely related to transfer entropy (TE), the time series version of conditional mutual information (Wyner, 1978). For Gaussian variables and linear dependence, TE is equivalent to the GC statistic of a bivariate autoregressive model (Barnett et al., 2009), and Diks and Fang (2020) show that an adaptation of the DP

statistic is the first-order Taylor expansion of TE. Finally, the LDE used by DP builds on KDE, which was introduced by Fix and Hodges (1951) and published in its current form by Rosenblatt (1956) and Parzen (1962).

In Section 2, we describe the DP test and show what it tests when the distribution changes over time. In Section 3, we introduce the local DP test. Section 4 derives its asymptotic distribution, the variance estimate and the test, and the bandwidth choice. We present a simulation study to demonstrate the power and size in different settings in Section 5 and apply the test to S&P 500 returns and volume changes in Section 6. Section 7 concludes. The proofs are in Appendix A, the rule of thumb of DP is in Appendix B, additional results for the S&P 500 application are in Appendix C, and the computational details are in Appendix D.

2 Granger causality testing under time variation

2.1 The Diks–Panchenko test

Following the paradigm of GC, we wish to know if Y and X_{-1} are independent given Y_{-1}, where Y_{-1}, X_{-1}, and Y are random variables. In this section, the observations $W_t := (Y_{t-1}, X_{t-1}, Y_t)$, $t = 1, \ldots, n$, are identically but not necessarily independently distributed, as in DP.

The joint density of (Y_{t-1}, X_{t-1}, Y_t) is denoted by f. To allow for concise formulation, the marginal densities of $f(y_{-1}, x_{-1}, y)$ are characterized by their arguments. For example, $f(y_{-1}, x_{-1})$ is the marginal density with respect to Y_{t-1} and X_{t-1}. To test for GC, the hypotheses are

$$\mathcal{H}_0 : \mathbb{P}\left[f(y|Y_{-1}) = f(y|Y_{-1}, X_{-1})\right] = 1 \quad \forall y \in \mathbb{R}, \tag{1}$$

$$\mathcal{H}_1 : \mathbb{P}\left[f(y|Y_{-1}) = f(y|Y_{-1}, X_{-1})\right] < 1 \quad \text{for some } y \in \mathbb{R}. \tag{2}$$

By Bayes' theorem we can rewrite the condition to $\frac{f(y_{-1},y)}{f(y_{-1})} = \frac{f(y_{-1},x_{-1},y)}{f(y_{-1},x_{-1})}$. DP then state this is equivalent to $\frac{f(y_{-1},x_{-1},y)}{f(y_{-1})} = \frac{f(y_{-1},y)}{f(y_{-1})}\frac{f(y_{-1},x_{-1})}{f(y_{-1})}$ for each vector (y_{-1}, x_{-1}, y) in the support of (Y_{-1}, X_{-1}, Y). They use this to construct a class of tests, where they define $\mathcal{H}_0'$ such that:

$$\mathcal{H}_0' : \mathbb{E}\left[\left(\frac{f(Y_{-1}, X_{-1}, Y)}{f(Y_{-1})} - \frac{f(Y_{-1}, Y)}{f(Y_{-1})}\frac{f(Y_{-1}, X_{-1})}{f(Y_{-1})}\right) g(Y_{-1}, X_{-1}, Y)\right] = 0 \tag{3}$$

where $g(Y_{-1}, X_{-1}, Y)$ is a positive weight function. They consider a variety of weight functions. They favor $g(y_{-1}, x_{-1}, y) := f(y_{-1})^2$ over others due to its stability and the U-statistic representation. The U-statistic representation allows for the analytical expression of the asymptotic

distribution. The null hypothesis becomes

$$\begin{aligned} &\mathcal{H}_0'' : q = 0 \\ &\text{where: } q := \mathbb{E}\left[f(Y_{-1}, X_{-1}, Y) f(Y_{-1}) - f(Y_{-1}, Y) f(Y_{-1}, X_{-1})\right]. \end{aligned} \tag{4}$$

The densities have to be estimated. For practical reasons, we follow DP who use LDE. Alternatives are from the class of KDE and can also be implemented. The LDE of the density of a d_Z-variate random vector Z at the observation Z_i is defined by

$$\hat{f}(Z_i; h) := \frac{(2h)^{-d_Z}}{n-1} \sum_{j \neq i} I_{ij}^Z, \tag{5}$$

where $I_{ij}^Z := I(\|Z_i - Z_j\|_\infty \leq h)$ and $\|u\|_\infty := \max_k |u_k|$ denotes the sup norm. For $Z = (Y_{-1}, X_{-1})$ and $Z = (Y_{-1}, Y)$, we write $I_{ij}^{Y_{-1}X_{-1}}$ and $I_{ij}^{Y_{-1}Y}$. The test statistic for testing (4) becomes

$$T_n(h) := \frac{n-1}{n(n-2)} \sum_i \big(\hat{f}(Y_{i-1}, X_{i-1}, Y_i)\hat{f}(Y_{i-1}) - \hat{f}(Y_{i-1}, Y_i)\hat{f}(Y_{i-1}, X_{i-1})\big), \tag{6}$$

where $h := h(n)$. For an appropriate bandwidth sequence, DP show that the statistic is asymptotically normal for independent observations, and argue that this extends to strictly stationary time series under suitable mixing conditions (Denker and Keller, 1983).

The DP test does not test $\mathcal{H}_0$ itself. $\mathcal{H}_0$ implies $\mathcal{H}_0''$, but not the other way round. Because the term inside the expectation in q can be positive in some regions and negative in others, $q = 0$ can also hold in some cases where the alternative hypothesis is true (Diks and Fang, 2020). The test is therefore not consistent against all alternatives. Diks and Fang (2020) propose an alternative weight function for which the dependence measure is nonnegative and zero only under $\mathcal{H}_0$, but in practice this test is often more conservative than the original DP test. We therefore use the original function $g(y_{-1}, x_{-1}, y)$.

2.2 The static DP test under time variation

When the distribution of W_t changes over time, the static DP test tests GNC in the distribution pooled over the sample, but not GNC at a given point in time. The two hypotheses differ when the date carries information about both X_{-1} and the relation between Y_{-1} and Y.

Let f_t denote the density of W_t. Granger (1969) notes that, for nonstationary series, his definitions can be generalized to causality at a specified time t. The null hypothesis of GNC at

time t is

$$\mathcal{H}_0(t) : \mathbb{P}\left[f_t(y|Y_{-1}) = f_t(y|Y_{-1}, X_{-1})\right] = 1 \quad \forall y \in \mathbb{R}. \tag{7}$$

To see what the static DP test tests, let D be a random date, uniformly distributed on $\{1, \ldots, n\}$ and independent of the observations, and let $(Y_{-1}, X_{-1}, Y) := W_D$. This vector has the pooled density

$$\bar{f} := \frac{1}{n} \sum_{s=1}^{n} f_s. \tag{8}$$

$T_n(h)$ compares every observation with every other observation, regardless of when they occur, and therefore estimates $q(\bar{f})$, where $q(f)$ denotes the measure q in Equation (4) under density f. The static DP test thus tests whether Y and X_{-1} are independent given Y_{-1}, whereas $\mathcal{H}_0(t)$ for every t states that they are independent given (Y_{-1}, D).

GNC at every point in time does not imply GNC under $\bar{f}$, because conditional independence is not preserved under averaging. The hypotheses coincide only under specific conditions.

PROPOSITION 2.1: *Suppose that $\mathcal{H}_0(t)$ holds for every $t = 1, \ldots, n$. Then*

$$\bar{f}(y|y_{-1}, x_{-1}) = \sum_{s=1}^{n} \rho_s(y_{-1}, x_{-1})\, f_s(y|y_{-1}), \qquad \rho_s(y_{-1}, x_{-1}) := \frac{f_s(y_{-1}, x_{-1})}{\sum_{j=1}^{n} f_j(y_{-1}, x_{-1})}, \tag{9}$$

where $\rho_s(y_{-1}, x_{-1}) = \mathbb{P}(D = s | Y_{-1} = y_{-1}, X_{-1} = x_{-1})$. Under $\bar{f}$, Y and X_{-1} are independent given Y_{-1} if (a) $f_s(x_{-1}|y_{-1})$ does not depend on s, or (b) $f_s(y|y_{-1})$ does not depend on s.

Proof. Under $\mathcal{H}_0(s)$, $f_s(y_{-1}, x_{-1}, y) = f_s(y|y_{-1})\, f_s(y_{-1}, x_{-1})$. Averaging over s and dividing by $\bar{f}(y_{-1}, x_{-1})$ gives Equation (9). Similarly, $\bar{f}(y|y_{-1}) = \sum_s \rho_s(y_{-1}) f_s(y|y_{-1})$ with $\rho_s(y_{-1}) := f_s(y_{-1}) / \sum_j f_j(y_{-1})$. Under (a), $\rho_s(y_{-1}, x_{-1}) = \rho_s(y_{-1})$. Under (b), $f_s(y|y_{-1}) = \bar{f}(y|y_{-1})$ for every s and the $\rho_s(y_{-1}, x_{-1})$ sum to one. In both cases, $\bar{f}(y|y_{-1}, x_{-1}) = \bar{f}(y|y_{-1})$. □

Proposition 2.1 tells us that pooled GNC can give different conclusions than GNC at every point in time. Under GNC at every point in time, the static null hypothesis in Equation (1) can only be rejected if X_{-1} carries information about the date and the date carries information about how Y depends on Y_{-1}. One could argue that the date then plays the role of the omitted third series in the example of spurious causality of Granger (1969). For example, if Y becomes more persistent and X becomes more heavy-tailed at the same time, a large $|X_{-1}|$ indicates the later period and helps predict Y given Y_{-1}. As demonstrated in Section 5.2, the static DP test then rejects.

Alternatively, one can approach this from a power perspective. When GC is present only in some periods, the pooled density mixes the causal and the non-causal periods. The dependence in the causal periods is weakened, so that the static DP test can have less power against short episodes of GC than a test at every point in time. Additionally, a rejection of a pooled test does not show when GC is present (Section 5.3). A test for GNC at every point in time can give insight into when GC is active.

The local DP test of Section 3 tests $\mathcal{H}_0(t)$ at each point in time. It gives more weight to observations close to t, and thereby approximately conditions on the date. This has three advantages. First, a change in the distribution that links X and Y through the date affects the local DP test less than the static DP test (Section 5.2). Second, GC that is active only in some periods is not diluted by the other periods. Third, the local DP test shows when GC is active (Section 5.3).

3 The local DP test

In this section, we introduce the local DP test of $\mathcal{H}_0(t)$ (Section 2.2). We use the technique introduced by Harvey and Oryshchenko (2012), who estimate univariate KDE with weighted averages based on exponentially weighted moving averages (EWMA), so that the density estimates adapt to distributional changes. We apply EWMA to the multivariate LDE in the DP statistic. This allows us to test GNC locally, at each point in time.

3.1 The local DP test statistic

We replace the averages in the DP statistic by weighted averages that concentrate on the observations close to time t. Let $\theta_{t,s} \geq 0$, with $\sum_{s=1}^{n} \theta_{t,s} = 1$, be the weight of observation s at time t. The smoothed density estimate at time t is $\hat{f}_{t|n}(z) := (2h)^{-d_Z} \sum_{s=1}^{n} \theta_{t,s} I(\|z - Z_s\|_\infty \leq h)$. For equal weights $\theta_{t,s} = 1/n$, it reduces to the static LDE.

We plug the time-varying density estimates into the DP statistic. At time t, we evaluate the density estimates at observation i with the leave-one-out sum $\hat{f}_{t|n}(Z_i) := (2h)^{-d_Z} \sum_{j \neq i} \theta_{t,j} I_{ij}^Z$, with I_{ij}^Z as in Section 2.1. The DP statistic averages their conditional dependence measure with equal weights over the sample. At time t, we take the expectation in q under the local distribution at that time. We therefore also replace the average over the conditional independence

measure by the weights $\theta_{t,i}$. The time-varying test statistic is

$$T_t(h) := \sum_{i=1}^{n} \theta_{t,i}\big(\hat{f}_{t|n}(Y_{i-1}, X_{i-1}, Y_i)\hat{f}_{t|n}(Y_{i-1}) - \hat{f}_{t|n}(Y_{i-1}, Y_i)\hat{f}_{t|n}(Y_{i-1}, X_{i-1})\big), \qquad t = 1, \ldots, n. \tag{10}$$

For equal weights $\theta_{t,s} = 1/n$, $T_t(h)$ is equivalent to $T_n(h)$: $T_t(h) = \frac{(n-1)(n-2)}{n^2} T_n(h)$ (Lemma A.2).

3.2 Exponential weights and the effective sample size

The EWMA smoother of Harvey and Oryshchenko (2012) estimates the density at time t by a weighted average of the LDE terms $(2h)^{-d_Z} I(\|z - Z_s\|_\infty \leq h)$. In the middle of the sample, the smoother puts weight $\frac{1-\gamma}{1+\gamma}\gamma^{|t-s|}$ on observation s (Whittle, 1984). Near the beginning and end of the sample, these weights do not sum to one. We therefore normalize them at every t,

$$\theta_{t,s} := \frac{\gamma^{|t-s|}}{S_t}, \qquad S_t := \sum_{s=1}^{n} \gamma^{|t-s|} = \frac{1 + \gamma - \gamma^t - \gamma^{n-t+1}}{1 - \gamma}, \qquad s, t = 1, \ldots, n, \tag{11}$$

and $\theta_{t,s} := 1/n$ for $\gamma = 1$. For $\gamma = 1$, all observations receive equal weight and the local DP test reduces to the static DP test. Appendix D points out how the weights can be interpreted as EWMA smoothed density estimates as in Harvey and Oryshchenko (2012).

The weights determine how many observations effectively enter the estimates at time t. Since the weights sum to one, the variance of a weighted average of independent variables with a common variance ς^2 is $\varsigma^2 \sum_s \theta_{t,s}^2$. This equals the variance ς^2/m of an unweighted average of $m = (\sum_s \theta_{t,s}^2)^{-1}$ observations. We therefore measure the amount of data at time t by the Kish effective sample size (Kish, 1965),

$$n_K := \Big(\sum_{s=1}^{n} \theta_{t,s}^2\Big)^{-1}, \tag{12}$$

the number of equally weighted observations that gives the same variance as the weighted average. In the middle of the sample, $n_K \approx 4/(1-\gamma)$ (Lemma A.1). At the beginning and end of the sample, the weights are one-sided and n_K is about half as large, so the density estimates there are based on effectively less data. For $\gamma = 1$, $n_K = n$. In this paper, we use a bandwidth that is constant over time. The density estimates within about $n_K/2$ observations of the beginning and end of the sample are therefore less accurate than those in the middle of the sample. Alternatively, one could use a larger bandwidth at the beginning and end of the sample.

The smoothing parameter governs a trade-off in time, similar to the trade-off that the

bandwidth h governs in space. A smaller γ concentrates the weights on observations closer to t, so that the weighted average of the densities f_s stays closer to f_t, and the hypothesis tested stays closer to $\mathcal{H}_0(t)$. It also reduces the effective sample size n_K, which increases the variance.

4 Asymptotic theory and bandwidth choice

This section derives the asymptotic distribution of $T_t(h)$ at a fixed point in time, the corresponding variance estimate and test, and the bandwidth that minimizes its mean squared error. We follow the approach of DP, who derive the asymptotic distribution of $T_n(h)$. As in DP, we derive the result for independent observations and an underlying distribution of the data that does not change over time. Under these assumptions, the derivation of DP carries over with weights, and the sample size n is replaced by the Kish effective sample size n_K. All proofs are in Appendix A.

4.1 Setting and assumptions

We derive the asymptotic distribution of the test statistic at a fixed point $u \in (0, 1)$ in rescaled time, with $t = \lfloor un \rfloor$, as in the literature on locally stationary processes (Dahlhaus, 1997; Vogt, 2012). This keeps the point at which we evaluate the test in the interior of the sample as n grows. Let f_u denote the density of W at time u. For the theorem, we assume that all observations are independent and have this same density (Assumption 4.1). In Remark 4.1, we consider what changes when the observations have time-varying densities.

Let $\mathbb{E}_u$ and $\mathbb{V}_u$ denote the expectation and variance under f_u. With $q(f)$ as in Section 2.2, we write $q_h(f)$ for the expectation of $T_n(h)$ when $W_1, \ldots, W_n$ are independent with density f. The dependence measure at time u is $q(u) := q(f_u)$, and the null hypothesis $\mathcal{H}_0(u)$ of GNC under f_u (Equation (7)) implies $q(u) = 0$. We write $q_h(u) := q_h(f_u)$.

Assumption 4.1 (Independence): *The vectors $W_1, \ldots, W_n$ are independent with density f_u.*

As in DP, we derive the asymptotic distribution for independent observations. This is an approximation, because consecutive vectors W_s and W_{s+1} share the observation Y_s. DP extend their result to dependent data by referring to Denker and Keller (1983), under which the asymptotic variance becomes a long-run variance. Remark 4.2 explains how the result extends to dependent data. In addition, Assumption 4.1 states that the distribution does not change over time. Remark 4.1 discusses the additional bias that arises when the distribution changes over time.

ASSUMPTION 4.2 (Density at time u): *The density f_u and its marginals are bounded and twice continuously differentiable with bounded derivatives.*

Assumption 4.2 contains the conditions of DP at the time point of interest. Under these conditions, DP's Conditions 1–3 hold for the uniform kernel, and $q_h(u) - q(u) = O(h^2)$.

ASSUMPTION 4.3 (The smoothing parameter and the bandwidth):

(i) The time point is $t = \lfloor un \rfloor$ for a fixed $u \in (0, 1)$.

(ii) The smoothing parameter $\gamma = \gamma_n \in (0, 1)$ satisfies $\gamma_n \to 1$ and $n(1 - \gamma_n) \to \infty$.

(iii) The bandwidth is $h = C_h n_K^{-\beta}$ with $C_h > 0$ and $\beta \in (\frac{1}{4}, \frac{1}{3})$.

Assumption 4.3 (i) states that u cannot be equal to 0 or 1. In practice, this means that the beginning and end of the sample are excluded, and that our asymptotic results only apply far enough from the boundaries of the sample.

Assumption 4.3 (ii) is similar to the usual conditions on the bandwidth h, but in time. We let $\gamma_n \to 1$ such that the number of observations in the window grows, $n_K \to \infty$. Since $n(1 - \gamma_n) \to \infty$, the sample size grows faster than the window, such that the window covers a shrinking fraction of the sample, $n_K/n \to 0$.

Assumption 4.3 (iii) is the bandwidth window of DP, with n replaced by n_K. The upper bound $\beta < \frac{1}{3}$ ensures that h does not shrink too fast, so that enough pairs of observations fall within distance h of each other and the noise of the density estimates remains negligible. The lower bound $\beta > \frac{1}{4}$ ensures that h shrinks fast enough to make the bias of the LDE negligible.

4.2 Asymptotic normality

THEOREM 4.1 (Asymptotic normality): *Let Assumptions 4.1–4.3 hold, and let $\sigma^2(u) > 0$. Then*

$$\sqrt{n_K}\,\big(T_t(h) - q(u)\big) \xrightarrow{d} \mathrm{N}\big(0, \sigma^2(u)\big), \tag{13}$$

where $\sigma^2(u)$ is the asymptotic variance in Equation (24) in Appendix A.5. In particular this result gives that under $\mathcal{H}_0(u)$, $\sqrt{n_K}\,T_t(h) \xrightarrow{d} \mathrm{N}\big(0, \sigma^2(u)\big)$.

The proof of Theorem 4.1 is in Appendix A.5. Our weighted theorem is analogous to the unweighted Theorem 1 of DP, with n replaced by n_K. We follow the proof steps of DP. We show that the weighted statistic is a weighted U-statistic and that its variance is that of DP with n replaced by n_K, and we then use Lyapunov's central limit theorem for unequal weights. The rate $\sqrt{n_K} \approx \sqrt{4/(1-\gamma)}$ does not depend on the total sample size n, so the precision of

the test at time t is determined by the number of observations in the window, not by the length of the sample.

REMARK 4.1 (Time-varying distribution): *Assumption 4.1 states that every W_s has density f_u. When the distribution changes over time, so that W_s has density $f_{s/n}$, smoothing over time creates a bias. Let $T_t^*(h)$ be the statistic computed with the same weights $\theta_{t,s}$ from a hypothetical sample $W_1^*, \ldots, W_n^*$ of independent draws from f_u. We define the bias from smoothing over time as*

$$B_t(h) := \mathbb{E}\, T_t(h) - \mathbb{E}\, T_t^*(h), \tag{14}$$

the difference between the expected statistic under the time-varying distribution and under the distribution at time u. For asymptotic normality as in Theorem 4.1 the bias has to vanish faster than the standard deviation, $\sqrt{n_K}\, B_t(h) \to 0$.

Vogt (2012) derives a similar bias term for the Nadaraya–Watson estimator of a regression function that changes smoothly over time. Vogt smooths in time and in space with a symmetric kernel and one bandwidth h for both directions, and shows that the resulting bias is of order h^2 (Vogt, 2012, Theorem 4.2). Our weights are also symmetric around t, and n_K/n plays the role of the bandwidth in time. If a similar result applies, $B_t(h)$ is of order $(n_K/n)^2$, and it is negligible if $\sqrt{n_K}\,(n_K/n)^2 \to 0$, that is, $n_K = o(n^{4/5})$. This means that n_K should not grow too fast, or equivalently, that $n^{4/5}(1-\gamma_n) \to \infty$. The proof of this is left for further work.

REMARK 4.2 (Dependent data): *According to Denker and Keller (1983), for weakly dependent data the static statistic $T_n(h)$ is still asymptotically normal, provided that the covariance among the $r_0(W_i)$ is taken into account in the asymptotic variance (DP, Appendix A.3). Their result is derived without time weighting and is therefore not immediately applicable to $T_t(h)$. For now, we assume that, under the conditions of Denker and Keller (1983, Theorem 1(c)), Theorem 4.1 still holds with $\sigma^2(u)$ replaced by the long-run variance, which we estimate in Equation* (15).

4.3 Variance estimation and the test

To use Theorem 4.1 for testing, we need an estimate of the variance $\sigma^2(u)/n_K$. As in DP, we use a HAC estimator (Newey and West, 1987), but with the weights $\theta_{t,i}$,

$$\hat{\sigma}_t^2 := 9 \sum_{\ell=0}^{L-1} \omega_\ell \sum_{i=\ell+1}^{n} \theta_{t,i}\theta_{t,i-\ell}\big(\hat{r}_{0,t}(W_i) - T_t(h)\big)\big(\hat{r}_{0,t}(W_{i-\ell}) - T_t(h)\big), \tag{15}$$

with Bartlett weights $\omega_0 = 1$ and $\omega_\ell = 2(1 - \ell/L)$ for $\ell \geq 1$, and $L = \lfloor n_K^{1/4} \rfloor$. For dependent data, the higher lags estimate the autocovariances of $r_{0,u}(W_s)$ (Remark 4.2). The number of observations that enters these estimates is of order n_K, not n. This is why the truncation lag L grows with n_K rather than with n. As in Equation (24), the factor 9 arises because the U-statistic is of order three. For $\gamma = 1$, $\hat{\sigma}_t^2$ reduces to the HAC variance of DP.

The estimator depends on $\hat{r}_{0,t}(W_i)$, an estimate of $r_{0,u}$ for observation i at time t. The function $r_{0,u}$ measures the contribution of a single observation to the statistic. As in DP, we estimate it by the average of the kernel $\kappa(W_i, W_j, W_k)$ over j and k with observation i fixed, but now weighted by $\theta_{t,j}\theta_{t,k}$. This gives

$$\begin{aligned}\hat{r}_{0,t}(W_i) := \tfrac{1}{3}\Big[& \hat{f}_{t|n}(W_i)\hat{f}_{t|n}(Y_{i-1}) - \hat{f}_{t|n}(Y_{i-1}, X_{i-1})\hat{f}_{t|n}(Y_{i-1}, Y_i) \\ & + \sum_{j \neq i} \theta_{t,j} \Big((2h)^{-1} \hat{f}_{t|n}(W_j) I_{ij}^{Y_{-1}} + (2h)^{-3} I_{ij}^{W} \hat{f}_{t|n}(Y_{j-1}) \\ & \qquad - (2h)^{-2} \hat{f}_{t|n}(Y_{j-1}, X_{j-1}) I_{ij}^{Y_{-1}Y} - (2h)^{-2} I_{ij}^{Y_{-1}X_{-1}} \hat{f}_{t|n}(Y_{j-1}, Y_j) \Big) \Big].\end{aligned} \tag{16}$$

It follows that $\sum_i \theta_{t,i} \hat{r}_{0,t}(W_i) \approx T_t(h)$.

The standardized statistic is $\tau_t := T_t(h)/\hat{\sigma}_t$. For independent observations, the variance is $\sigma^2(u)/n_K = 9\mathbb{V}_u(r_{0,u}(W)) \sum_s \theta_{t,s}^2$ and similarly $\hat{\sigma}_t^2$ in Equation (15) without any lags is $9\sum_i \theta_{t,i}^2 (\hat{r}_{0,t}(W_i) - T_t(h))^2$, its weighted sample estimate. Following Newey and West (1987), $n_K \hat{\sigma}_t^2$ is an asymptotically consistent estimate of $\sigma^2(u)$ such that τ_t is asymptotically standard normal under $\mathcal{H}_0(u)$. The test is one-sided and rejects GNC at time t for large values of τ_t, with p-value $1 - \Phi(\tau_t)$.

4.4 Bandwidth choice

The appropriate bandwidth is crucial for density estimation. The bandwidth balances the variance and bias in the estimation. We seek the bandwidth that minimizes the mean squared error of the $T_t(h)$ sequence. DP state that the optimal bandwidth for the test statistic is $h_n^* := C^* n^{-2/7}$. They derive an analytical expression for C^*. However, this is intractable, as it requires knowledge of the underlying distribution of the data. Making parametric assumptions allows for approximating the optimal value C^*. DP approximate C^* for an ARCH process with ARCH coefficient δ. They assume that $Y_{-1} \sim \mathcal{N}(0, 1)$ and that, conditional on Y_{-1}, the variables X_{-1} and Y are independent and $\mathcal{N}(0, 1 + \delta Y_{-1}^2)$ distributed, which is reasonable for small δ. For $\delta = 0.4$, they find $C^* \approx 8$. In their simulations, they use $C = 8.62$, which they find empirically to give fast convergence of the size to the nominal level (Appendix B). We do not address the choice of C^* here. Instead, we ask how the smoothing parameter influences the

optimal bandwidth. In Corollary A.1 in Appendix A.6, we show that the mean squared error of $T_t(h)$ is that of DP with n replaced by n_K, and with the same constant C^*. The γ-specific rule of thumb therefore becomes

$$h^*_{n_K} := C^* n_K^{-2/7}, \tag{17}$$

where $n_K \approx 4/(1-\gamma)$ is the Kish effective sample size in the middle of the sample. For large enough samples this is independent of the total number of observations, due to the exponentially decreasing nature of the weights. The relevant sample size is n_K, because the variance of a weighted average depends on the sum of the squared weights. According to this rule-of-thumb, halving the effective sample size means a 21.9% increase of the optimal bandwidth parameter.

5 Simulation study

We study the finite-sample behavior of the local DP test in three experiments. The goal of the first experiment is to verify the size and power when the distribution does not change over time. The second experiment shows the advantage of the local DP test when the distribution changes over time and the static DP test rejects although GNC holds at every point in time. The third experiment examines what happens to the test when the Granger causal relationship between two variables changes.

The three experiments share a common design. Each simulated sample has $n = 5000$ observations, and, as in DP, each series is standardized to unit variance before testing. We consider $\gamma \in \{0.992, 0.996, 0.998, 0.999\}$, with Kish effective sample sizes $n_K \approx 4/(1-\gamma)$ of 500, 1000, 2000 and 4000 in the middle of the sample,[1] and the static DP test with $\gamma = 1$. The bandwidth follows the rule of thumb of Section 4.4. We use the constant $C = 8.62$ and the cap of DP, so that

$$h = \min\big(8.62\, n_K^{-2/7},\, 1.5\big), \tag{18}$$

with n in place of n_K for the static DP test. DP choose $C = 8.62$ because, in their simulations of the ARCH design below with $a = 0.4$, it gives fast convergence of the size of the static DP test to the nominal level; it is close to the approximate optimal value $C^* \approx 8$ for $a = 0.4$. This gives $h = 1.46$, 1.20, 0.98 and 0.81 for the local DP tests and $h = 0.76$ for the static DP test. All tests are one-sided at the nominal level $\alpha = 0.05$. Theorem 4.1 applies to time points in

[1]The approximation $n_K \approx 4/(1-\gamma)$ requires $n \gg n_K$. For $n = 5000$, the exact values at $t = n/2$ are about 498, 998, 1971 and 3392. We use the approximate values for the bandwidth.

the interior of the sample. In the tables that show rejection rates averaged over time, we only include time points that lie more than $n_K/2$ observations from both ends of the sample.

5.1 Size and power in a stationary environment

The first experiment uses the bivariate ARCH process of DP,

$$X_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}\big(0,\, 1 + aY_{t-1}^2\big), \qquad Y_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}\big(0,\, 1 + aY_{t-1}^2\big), \qquad t = 1, \ldots, n, \tag{19}$$

where X_t and Y_t are independent given the past $\mathcal{F}_{t-1}$. In this process, Y_{t-1} Granger causes X_t through its variance, while X_{t-1} does not Granger cause Y_t. We estimate the size by testing GNC from X_{t-1} to Y_t for $a = 0.4$, the value of DP, for $a = 0.3$ and for $a = 0$, in which case the observations are i.i.d. We estimate the power by testing GNC from Y_{t-1} to X_t for $a = 0.1$, as in DP. Since the distribution does not change over time, the static DP test uses all observations and should perform best in this design.

Table 1 reports the rejection rates. For $a = 0.4$, all tests have a size close to the nominal level, between 4.6% and 5.2%. For $a = 0.3$ and for i.i.d. data, all tests underreject, generally more so for smaller γ. DP chose the constant $C = 8.62$ empirically at $a = 0.4$. For $a = 0.3$, the same constant makes the test conservative.

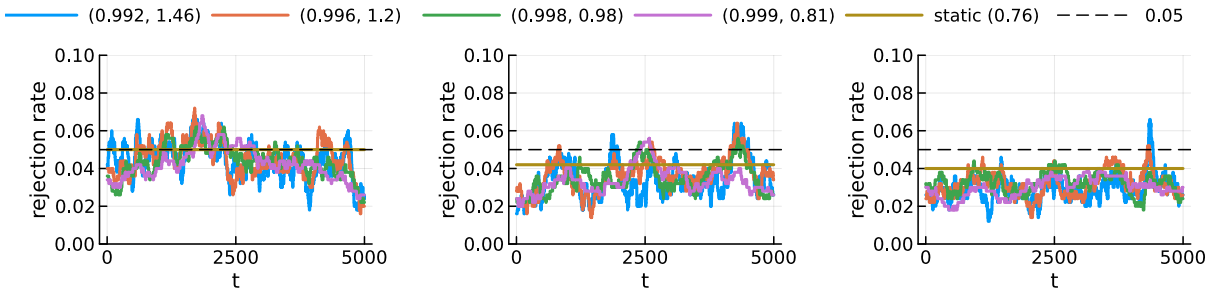
The power increases with γ, from 0.41 to 0.98, and is highest for the static DP test. At each point in time, the local DP test uses effectively n_K observations instead of n. When the distribution does not change over time, this loss of power is the cost of smoothing over time.

TABLE 1: Rejection rates of the local DP test in the ARCH design of DP.

	$\gamma = 0.992$ $n_K = 500$ $h = 1.46$	$\gamma = 0.996$ $n_K = 1000$ $h = 1.2$	$\gamma = 0.998$ $n_K = 2000$ $h = 0.98$	$\gamma = 0.999$ $n_K = 4000$ $h = 0.81$	static n $h = 0.76$
Size, $a = 0.4$	0.046 (0.002)	0.047 (0.003)	0.048 (0.005)	0.052 (0.009)	0.050 (0.010)
Size, $a = 0.3$	0.035 (0.002)	0.038 (0.003)	0.036 (0.004)	0.043 (0.008)	0.042 (0.009)
Size, i.i.d.	0.029 (0.002)	0.032 (0.003)	0.034 (0.004)	0.033 (0.007)	0.040 (0.009)
Power, $a = 0.1$	0.406 (0.006)	0.647 (0.009)	0.874 (0.008)	0.976 (0.006)	0.998 (0.002)

Notes: Interior rejection rates at $\alpha = 0.05$ over $M = 500$ repetitions, with Monte Carlo standard errors in parentheses.

Figure 1 shows the rejection rates at each time point. Near the beginning and end of the sample, where the weights are one-sided, the tests reject slightly less often, but the differences are small.



(a) $a = 0.4$ (b) $a = 0.3$ (c) $a = 0$ (i.i.d.)

FIGURE 1: Size of the local DP test over time in the ARCH design of DP.

Note: Rejection rates at each time point under the null hypothesis, averaged over $M = 500$ repetitions.

5.2 Size under a time-varying null hypothesis

The second experiment illustrates that the static DP test can reject although $\mathcal{H}_0(t)$ holds for every t. This supports the discussion in Section 2.2. We let the null hypothesis of GNC hold at every point in time, but we let the distribution of the data change. The data-generating process is

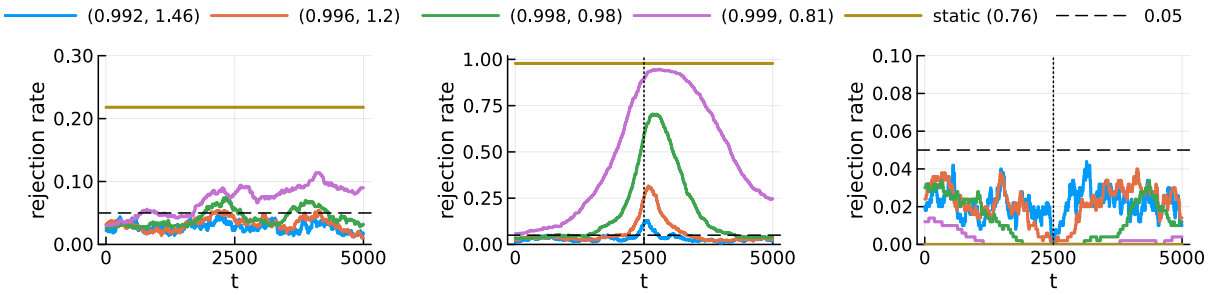
$$\begin{aligned} Y_t &= \phi_t \, Y_{t-1} + \sqrt{1 - \phi_t^2} \, e_t, \\ X_t &= v_t \, \eta_t, \qquad v_t = \begin{cases} 1 & \text{with probability } 1 - \pi_t, \\ 0.5 & \text{with probability } 0.8 \, \pi_t, \\ 2 & \text{with probability } 0.2 \, \pi_t, \end{cases} \end{aligned} \tag{20}$$

where e_t and η_t are independent standard normal and independent of v_t. Y follows an AR(1) process with time-varying persistence and heteroskedasticity. X has heteroskedasticity of a different form, and for larger π_t, X becomes more heavy-tailed. Both series have unit variance at every point in time, since $\mathbb{E}\, v_t^2 = 1$. X is independent of Y, so X_{t-1} does not Granger cause Y_t at any point in time.

We let ϕ_t and π_t depend on a common state $s_t \in [0, 1]$. One can think of s_t as the state of the economy, for example a period of turmoil after a common shock. In that period, Y becomes more persistent and X becomes more heavy-tailed, but X does not cause Y. We consider three designs, each with $n = 5000$ observations. In the first two designs, $\phi_t = 0.8 \, s_t$ and $\pi_t = s_t$, so that persistence and heavy-tailedness increase together. In the first design, *co-moving smooth changes*, the state changes smoothly, with $s_t = t/n$. In the second design, *co-moving break*, the state changes at a break, with $s_t = \mathbf{1}\{t > n/2\}$. In the third design, *opposite break*, we reverse

the second design and let persistence and heavy-tailedness move in opposite directions at the break, with $\phi_t = 0.8\, s_t$, $\pi_t = 1 - s_t$ and $s_t = \mathbf{1}\{t > n/2\}$. As a control, $\phi_t = \pi_t = 0$ gives i.i.d. observations.

These designs show how a regime change that affects both variables at the same time can lead the static DP test to reject. The static DP test pools all observations and ignores s_t. In the pooled distribution, a large or small $|X_{t-1}|$ identifies the change of regime, and the regime determines how strongly Y_t depends on Y_{t-1}. The static DP test therefore detects a relation between X_{t-1} and Y_t that runs through the common state s_t, not from X to Y. This is the familiar problem of an omitted common driver in GC testing. Here s_t is a deterministic function of time, so the local DP test, which compares observations close in time, has the property that it effectively conditions on the state. When persistence and heavy-tailedness change together, the static DP test therefore rejects GNC in the pooled distribution, although GNC holds at every point in time.



(a) Co-moving smooth change (b) Co-moving break (c) Opposite break

FIGURE 2: Rejection rates over time when GNC holds at every time point.

Note: Rejection rates at each time point when X is independent of Y, averaged over $M = 500$ repetitions.

Figure 2 shows the rejection rates over time, and Table 2 reports them for the quarters just before and just after the break. The static DP test rejects GNC in 22% of the samples under the smooth change of the first design and in 98% of the samples under the co-moving break of the second design. When the co-moving break is turned around, the static DP test never rejects.

In the smooth-change simulation experiment, the advantage of the local DP test is clear. The change affects the size of the local DP test less than it affects the size of the static DP test. The local DP test with the smallest smoothing parameter is the least affected by the change. When there is a break, the local DP test also has size problems near the location of the break, but the overrejection and underrejection are smaller for smaller smoothing parameters. The

overrejection and underrejection occur in a band of about n_K observations around the break (Figure 2(b) and (c)). Table 2 shows the average rejection rates in the second quarter, between the 1250th and 2500th observations, and in the third quarter, between the 2500th and 3750th observations. It shows that the test with the smallest smoothing parameter is the least affected by these forms of time variation.

TABLE 2: Size under a time-varying null hypothesis before and after the break.

		$\gamma = 0.992$ $h = 1.46$	$\gamma = 0.996$ $h = 1.2$	$\gamma = 0.998$ $h = 0.98$	$\gamma = 0.999$ $h = 0.81$	static $h = 0.76$
i.i.d.	Q2	0.027 (0.003)	0.030 (0.004)	0.030 (0.005)	0.030 (0.006)	0.034 (0.008)
	Q3	0.030 (0.003)	0.032 (0.004)	0.031 (0.005)	0.027 (0.006)	0.034 (0.008)
Smooth change	Q2	0.031 (0.003)	0.038 (0.005)	0.048 (0.007)	0.065 (0.009)	0.218 (0.018)
	Q3	0.026 (0.003)	0.033 (0.005)	0.042 (0.007)	0.082 (0.010)	0.218 (0.018)
Co-moving break	Q2	0.036 (0.004)	0.061 (0.006)	0.178 (0.011)	0.493 (0.016)	0.978 (0.007)
	Q3	0.045 (0.004)	0.122 (0.008)	0.440 (0.015)	0.891 (0.011)	0.978 (0.007)
Opposite break	Q2	0.020 (0.002)	0.016 (0.003)	0.004 (0.002)	0.000 (0.000)	0.000 (0.000)
	Q3	0.024 (0.003)	0.017 (0.003)	0.002 (0.001)	0.000 (0.000)	0.000 (0.000)

Notes: Rejection rates just before (Q2) and just after (Q3) the break over $M = 500$ repetitions, with Monte Carlo standard errors in parentheses.

5.3 Time-varying Granger causality

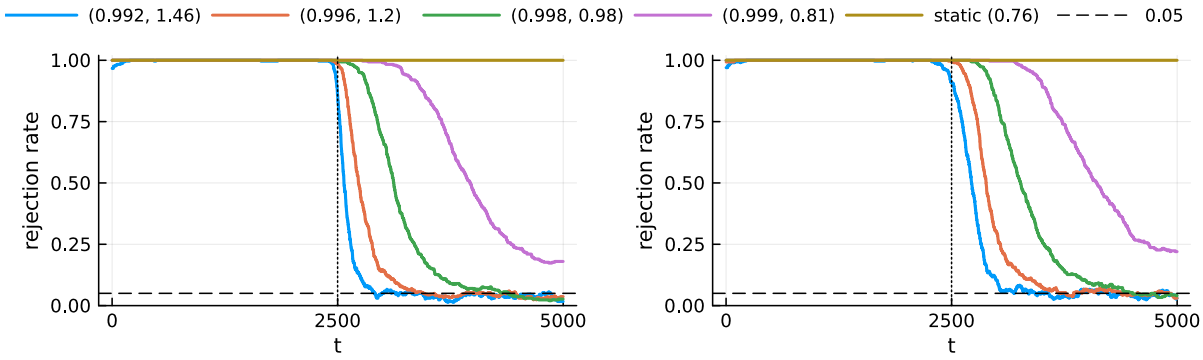
The third experiment examines how the local and static DP tests respond to time-varying Granger causality. We use the ARCH process of Section 5.1, but let the variable that drives the volatility switch from Y to X,

$$X_t,\, Y_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}\big(0,\, \sigma_t^2\big), \qquad \sigma_t^2 = 1 + a\big\{(1 - w_t)\, Y_{t-1}^2 + w_t\, X_{t-1}^2\big\}, \tag{21}$$

with $a = 0.4$. We consider an abrupt switch, $w_t = \mathbf{1}\{t > n/2\}$, and a smooth switch, $w_t = \{1 + \exp(-(t - n/2)/150)\}^{-1}$; the transition takes approximately 900 observations around $t = n/2$. We test GNC from Y_{t-1} to X_t. The alternative hypothesis holds before the switch and the null hypothesis holds after it.

Figure 3 shows the rejection rates over time. Before both switches, all tests reject in almost every sample. After both switches, the static DP test continues to reject at every time point, since it measures dependence in the distribution pooled over both halves of the sample. The local DP tests return to the nominal level. For $\gamma = 0.999$, the smoothing parameter is too large relative to the second half of the sample, and the local DP test still rejects at the end of the sample.

The second design examines GC that holds only during a short episode. We use Process (21)



(a) Abrupt switch

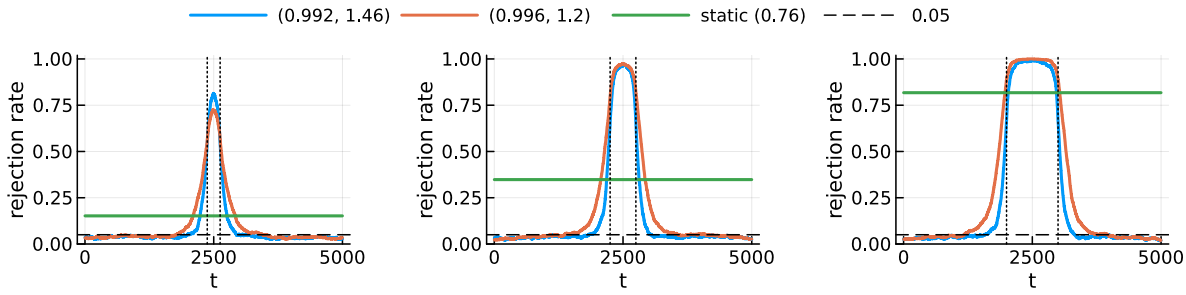
(b) Smooth switch

FIGURE 3: Rejection rates over time when the direction of GC switches.

Note: Rejection rates at each time point of the test for GNC from Y_{t-1} to X_t, which holds only after $t = n/2$, averaged over $M = 300$ repetitions.

with $a = 0.3$, and set $w_t = 0$ for $|t - n/2| < L/2$ and $w_t = 1$ otherwise. So Y Granger causes X only in an episode of L observations around $t = n/2$. Outside the episode, X drives the volatility of both series, so the marginal distributions are the same over the whole sample. We consider episodes of $L \in \{250, 500, 1000\}$ observations, which cover 5%, 10% and 20% of the sample.

Figure 4 shows the rejection rates over time. The static DP test detects the episode reliably only when it covers a large part of the sample, since the observations outside the episode weaken the dependence within it. At the center of the episode, the local DP tests reject in most samples, also for the shortest episode. For $L = 250$, $\gamma = 0.992$ rejects more often than $\gamma = 0.996$, because its effective sample size is closer to the length of the episode. Away from the episode, the local DP tests return to the nominal level. The local DP test thus detects short episodes of GC that the static DP test misses.



(a) $L = 250$ (b) $L = 500$ (c) $L = 1000$

FIGURE 4: Rejection rates over time when GC holds only during an episode.

Note: Rejection rates at each time point of the test for GNC from Y_{t-1} to X_t, which fails only for $|t - n/2| < L/2$ (between the dotted lines), averaged over $M = 1020$ repetitions.

6 Empirical results

Hiemstra and Jones (1994) used linear and nonparametric tests to study the relationship between daily Dow Jones returns and changes in trading volume. Their nonparametric test found strong evidence that past volume changes help predict returns. Diks and Panchenko (2006) revisited this question using S&P 500 data from a comparable period. Both nonparametric tests found strong evidence that returns predict volume changes, while the DP test gave considerably weaker evidence in the reverse direction. Both studies answer the question for the sample as a whole. We ask how the predictive dependence in each direction varies over time, and which periods drive the full-sample evidence.

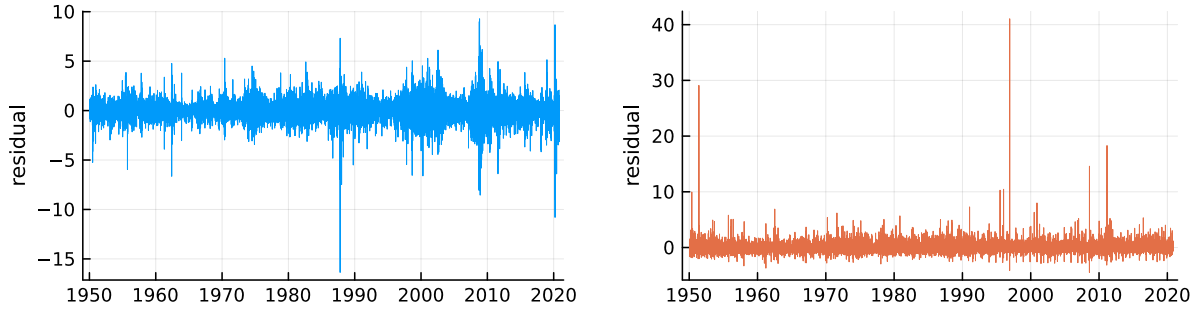
We use daily S&P 500 returns and volume changes from 18 January 1950 to 4 November 2020, with $n = 17{,}816$ test dates. Following DP, we adjust the series for month-of-the-year and day-of-the-week effects and use a first-order vector autoregression (VAR(1)) to prefilter the adjusted series. We prefilter to remove the fast moving signal that is not well captured by the nonparametric estimation. All tests are applied to the VAR residuals and test predictive dependence in the prefiltered series. We denote returns to volume by R→V and volume to returns by V→R. In the main text we use $\gamma = 0.996$ as the reference setting. Appendix C reports the results for $\gamma = 0.998$ and $\gamma = 0.992$, and we discuss them where they change the interpretation.

6.1 Static full-sample results

Table 3 reports the full-sample tests. With bandwidth $h = 0.53$ from Equation (18), the static DP test rejects GNC in both directions at $p < 0.001$. We additionally include a parametric benchmark, the one-lag causality-in-variance test of Cheung and Ng (1996), which tests for

correlation between lagged and current squared residuals after standardizing each series by its own GARCH(1,1) volatility. It does not reject in either direction. The DP rejections together with the weaker evidence from this benchmark leave several explanations open. The DP test may detect distributional features that the benchmark omits, it may have more power against the same alternative, or it may respond to a changing distribution. The volatility clustering in the VAR residuals (Figure 5) makes the last explanation relevant, as pooling distributions that change over time can produce static rejections (Section 2.2).

Following DP, we also divide the return residuals by their EGARCH(1,1) volatility (Nelson, 1991). On the filtered series, the static DP test rejects from returns to volume, but not from volume to returns. The static DP result agrees with the one-lag result of DP.



(a) Returns

(b) Volume changes

FIGURE 5: Prefiltered S&P 500 returns and volume changes.

TABLE 3: Full-sample GC tests for S&P 500 returns and volume changes.

	VAR residuals		EGARCH-filtered	
Test	R → V	V → R	R → V	V → R
DP, $h = 0.53$	< 0.001	< 0.001	< 0.001	0.239
Causality in variance (Cheung–Ng), one lag	0.679	0.230	—	—

Notes: p-values of the static DP test and the one-lag causality-in-variance test of Cheung and Ng (1996) in both directions. EGARCH-filtered: return residuals divided by their EGARCH(1,1) volatility, volume residuals unchanged. The Cheung–Ng test already standardizes each series by its own GARCH(1,1) volatility.

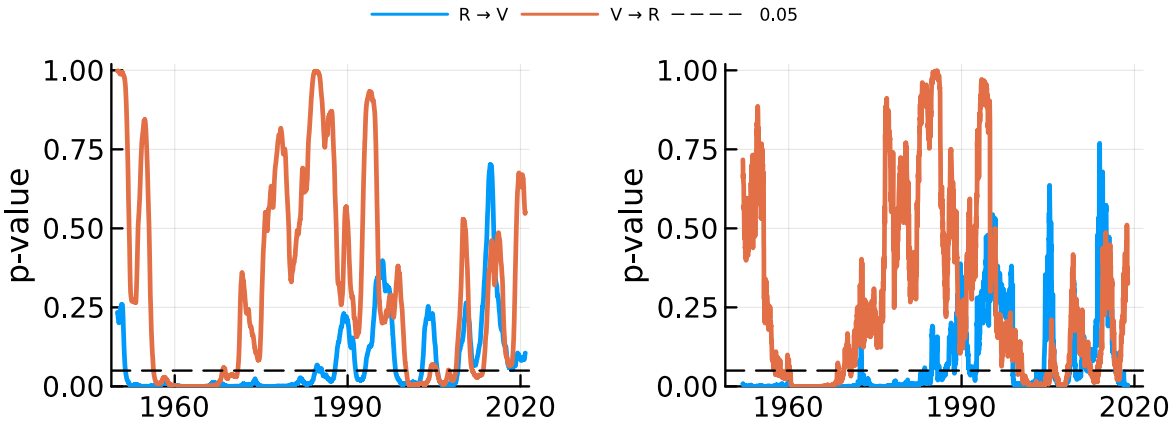
6.2 Local results and moving-window benchmarks

Figure 6(a) plots the local DP p-values at the reference setting $\gamma = 0.996$, with $h = 1.20$ from Equation (18). These p-values are pointwise. We use rejection percentages to describe the time pattern without interpreting them as a joint significance level for entire periods. Table 4 reports the percentages. The local DP test rejects from returns to volume on about two thirds of the dates, spread over the whole sample. It rejects from volume to returns on about a third of

the dates, mainly in 1956–1971 and 2000–2003. In these periods, the test often rejects in both directions at the same date. The larger and smaller smoothing parameters change the rejection percentages, but rejections stay more present in the returns to volume direction than in the volume to returns direction (Appendix C).

To see how much the exponential weights matter, we compare the local DP test with the static DP test applied to centered moving windows. A window of 1,001 observations matches the interior effective sample size for $\gamma = 0.996$, and we use the same bandwidth. The window moves five trading days at a time. For the comparison in Table 4, we evaluate both tests on the same dates: the window centers that are interior dates of the local DP test, which lie more than $n_K/2$ observations from either end of the sample. Figure 6(b) shows that the moving-window test locates similar episodes, but the results are less smooth. This results in more oscillating behavior between rejections and nonrejections. Table 6 in Appendix C gives the percentage of dates on which the local DP test and the moving-window DP test give the same decision. For the two longer matched window lengths, the two DP implementations have more similar rejection rates. For the shortest window of 501 observations, the moving-window test rejects more often from returns to volume than the local DP test with $\gamma = 0.992$.

The Cheung–Ng test on the same 1,001-observation windows rejects on a minority of dates in both directions (Figure 10 in Appendix C). The DP tests therefore find considerably more evidence in both directions than this benchmark.



(a) Local DP, $\gamma = 0.996$, $h = 1.20$ (b) Moving-window DP, $m = 1{,}001$, $h = 1.20$

FIGURE 6: Local and moving-window DP tests for S&P 500 returns and volume changes.

Note: Pointwise p-values on the VAR residuals in both directions. Moving-window values are plotted at the center of each window, which moves five trading days at a time.

TABLE 4: Percentage of dates with a rejection of GNC at the reference setting.

	All dates			1956–1971		2000–2003	
Specification	R → V	V → R	Joint	R → V	V → R	R → V	V → R
Panel A: Local DP, $\gamma = 0.996$, $h = 1.20$							
VAR residuals	66.5	34.9	29.8	100.0	88.8	72.1	95.5
EGARCH-filtered	48.4	15.6	12.5	76.7	38.3	72.1	92.5
EWMA-standardized	50.9	15.8	14.7	97.1	43.8	78.6	89.1
Panel B: Moving-window DP, $m = 1{,}001$, $h = 1.20$							
VAR residuals	68.2	30.0	26.0	100.0	67.8	100.0	87.1
EGARCH-filtered	52.1	12.5	10.6	90.8	32.5	71.6	69.2
EWMA-standardized	51.4	13.2	12.7	98.4	36.6	85.1	74.6
Local and moving-window DP both reject	61.1	26.5	22.6	100.0	66.2	72.1	87.1
Panel C: Cheung–Ng test, one lag, $m = 1{,}001$							
VAR residuals	18.9	4.3	2.5	27.6	11.1	0.0	4.0

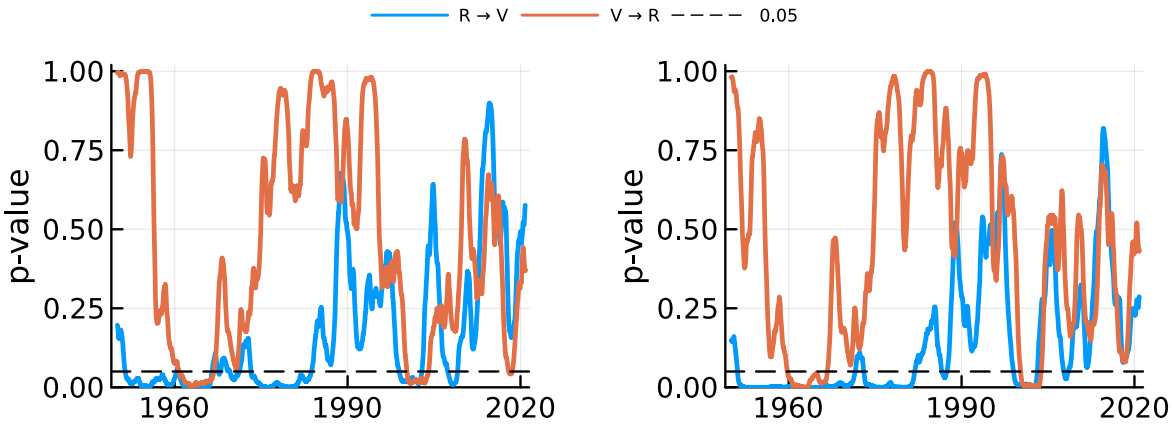
Notes: Percentage of dates with $p < 0.05$, evaluated on a common grid: the centers of the 1,001-observation moving windows, every fifth trading day, that are interior dates of the local DP test with $\gamma = 0.996$. Joint means a rejection in both directions at the same date. EGARCH-filtered: return residuals divided by their EGARCH(1,1) volatility, volume residuals unchanged. EWMA-standardized: both residual series divided by their EWMA volatility.

6.3 Robustness to time-varying volatility

The simulations in Section 5.2 show that when an unobserved state changes the distributions of both series, pooling observations across states can lead to rejections in both directions. A common change in volatility is one way this mechanism could operate in the S&P 500 data. Within a window that spans a rise in volatility, a large lagged observation in either series signals the high-volatility state and can therefore be recognized as GC when we pool over that period. Although the local DP test is more robust against distributional changes, the *co-moving break* design in Section 5.2 shows that changes that are too abrupt can still lead to rejections. As a countermeasure, we remove part of the fast-moving changes with a volatility filter and check whether the local DP test still rejects in each direction afterwards.

Dividing each residual by its estimated volatility puts observations from calm and turbulent periods on a common scale. We use exponentially weighted moving average (EWMA) volatility estimates with decay parameter $\lambda = 0.94$ for both series. As a second filter, we follow DP and divide only the return residuals by their EGARCH(1,1) volatility. EWMA adapts locally, whereas the EGARCH parameters are estimated over the full sample. As DP point out, volatility filtering can change the dependence between the series, and a misspecified volatility model can leave conditional heteroskedasticity in the filtered residuals. Figure 7 shows the local DP test after each filter.

After either filter, the local DP test continues to reject from returns to volume on about



(a) EGARCH-filtered returns

(b) EWMA-standardized residuals

FIGURE 7: Local DP test after volatility filtering.

Note: Pointwise p-values of the local DP test with $\gamma = 0.996$ and $h = 1.20$ in both directions. Panel (a): return residuals divided by their EGARCH(1,1) volatility, volume residuals unchanged. Panel (b): both residual series divided by their EWMA volatility with $\lambda = 0.94$. Figure 6(a) shows the test on the unfiltered VAR residuals.

half of the dates. In 1956–1971, the local DP test rejects on nearly all dates before and after EWMA standardization. The EGARCH filter lowers the rejection rate in this period, but the test still rejects on most dates. In 2000–2003, the local DP test rejects on most dates after both filters. There is strong evidence for Granger causality from returns to volume and the evidence is present throughout a large part of the sample.

In the direction from volume to returns, both filters more than halve the rejection rate over the full sample. After filtering, the local DP test rarely rejects outside 1956–1971 and 2000–2003. In 1956–1971, the rejection rate falls from 88.8% before filtering to 38.3% after EGARCH filtering and 43.8% after EWMA standardization. Still, after either filter, the local DP test rejects on almost every date from about 1960 to 1966. In 2000–2003, the test rejects on 95.5% of dates before filtering and on around 90% of dates after either filter. Interestingly, the local DP test rejects GNC in both directions during this period. This episode coincided with the bursting of the dot-com bubble, the 2001 U.S. recession, and the switch from fractional to decimal pricing on U.S. stock exchanges (Ofek and Richardson, 2003; National Bureau of Economic Research, 2003). One possible explanation is that these events affected both series abruptly and are therefore still recognized as GC in the local test, while in fact GNC could hold at every point in time. In the simulation experiment in Section 5.2, however, a break leads the local DP test to overreject in a band of about n_K observations around the break (Figure 2(b)). The rejections in 1960–1966 last about 1,500 trading days, longer than $n_K \approx 1{,}000$, and are

therefore difficult to explain by a single break. The rejections in 2000–2003 last about n_K observations. For $\gamma = 0.992$, with $n_K \approx 500$, the band around a break is narrower. After either filter, the local DP test then rejects only briefly around 2001, whereas it still rejects on a large part of the dates between 1960 and 1966 (Figure 8(f) and (i) in Appendix C). The rejections from volume to returns in 2000–2003 are therefore also consistent with a common change in the distributions of both series around 2001, whereas the rejections in 1960–1966 are more robust to the choice of γ.

The static DP test rejects GNC in both directions over the full sample. The local DP test shows that the evidence from returns to volume changes is spread across much of the sample and remains after volatility filtering. In the reverse direction, the evidence is concentrated in two episodes. Filtering shortens the episode in 1956–1971 to about 1960–1966, and the local DP test still rejects on most dates in 2000–2003 after either filter. The local DP test thus identifies two episodes of predictability from volume changes to returns that the full-sample DP test misses after EGARCH filtering, although the episode in 2000–2003 may reflect a common change in both series.

7 Conclusion

We present the local DP test, a nonparametric test for GNC at each point in time. The DP test pools the distributions at all points in time, and measures conditional dependence in the time-averaged distribution. As a result, it loses power when GC is only present in some periods, and it can reject when the distribution changes over time, although GNC holds at every point in time. The local DP test replaces the averages in the DP statistic by exponentially weighted averages around each point in time, based on the EWMA density estimates of Harvey and Oryshchenko (2012).

Following the approach of DP, we derive the asymptotic normal distribution of the local statistic at a fixed point in time, in which the Kish effective sample size $n_K \approx 4/(1-\gamma)$ takes the role of the sample size. The same substitution gives the rule of thumb for the bandwidth, $h^*_{n_K} = C^* n_K^{-2/7}$. A smaller γ keeps the pooled distribution closer to the distribution at time t, but it also reduces the effective sample size.

The simulations show both sides of this trade-off. When the distribution does not change over time, the static DP test has the most power. When a common state changes the distributions of both series, the static DP test rejects although GNC holds at every point in time, while the local DP test stays close to the nominal level except near a break. When GC is only present in some periods, the local DP test locates these periods, including short episodes that

the static DP test misses. For prefiltered S&P 500 returns and volume changes, the static DP test rejects GNC in both directions. The local DP test shows that the evidence from returns to volume is spread over the whole sample and survives volatility filtering, while the evidence from volume to returns is more sensitive to volatility and is concentrated in the early 1960s and the early 2000s once volatility is accounted for, although the early 2000s episode may reflect a common change in both series. Parametric methods may often be preferred due to their higher power when correctly specified. If the local DP test identifies GC missed by these methods, it could indicate the presence of a different, potentially time-varying, nonlinear type of GC.

Several questions remain for future research. We derive the asymptotic distribution for independent observations with a distribution that does not change over time, and leave the proofs for time-varying and dependent data for future research. The asymptotic distribution is pointwise in time. To determine whether GC is present during longer time periods, a bootstrap method similar to Phillips et al. (2015a) could be employed to account for the dependence between the test statistics. Finally, the method can easily be extended to higher-order lags, to conditioning on other variables as in Diks and Wolski (2016), and to a kernel density estimator with a Gaussian kernel as suggested by Diks and Fang (2020).

References

Baek, E. G. and Brock, W. A. (1992). A nonparametric test for independence of a multivariate time series. *Statistica Sinica*, 2(1):137–156.

Barnett, L., Barrett, A. B., and Seth, A. K. (2009). Granger causality and transfer entropy are equivalent for Gaussian variables. *Physical Review Letters*, 103(23):238701.

Bell, D., Kay, J., and Malley, J. (1996). A non-parametric approach to non-linear causality testing. *Economics Letters*, 51(1):7–18.

Billingsley, P. (1995). *Probability and Measure*. Wiley, New York, 3rd edition.

Bouezmarni, T., Rombouts, J. V. K., and Taamouti, A. (2012). Nonparametric copula-based test for conditional independence with applications to Granger causality. *Journal of Business & Economic Statistics*, 30(2):275–287.

Broock, W. A., Scheinkman, J. A., Dechert, W. D., and LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, 15(3):197–235.

Chen, Y., Rangarajan, G., Feng, J., and Ding, M. (2004). Analyzing multiple nonlinear time series with extended Granger causality. *Physics Letters A*, 324(1):26–35.

Cheung, Y.-W. and Ng, L. K. (1996). A causality-in-variance test and its application to financial market prices. *Journal of Econometrics*, 72(1–2):33–48.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, 25(1):1–37.

Denker, M. and Keller, G. (1983). On U-statistics and v. Mises' statistics for weakly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 64(4):505–522.

Diks, C. and Fang, H. (2020). A consistent nonparametric test for Granger non-causality based on the transfer entropy. *Entropy*, 22(10).

Diks, C. and Panchenko, V. (2006). A new statistic and practical guidelines for nonparametric Granger causality testing. *Journal of Economic Dynamics and Control*, 30(9-10):1647–1669.

Diks, C. and Wolski, M. (2016). Nonlinear Granger causality: Guidelines for multivariate analysis. *Journal of Applied Econometrics*, 31(7):1333–1351.

Fix, E. and Hodges, J. L. (1951). Discriminatory analysis. Nonparametric discrimination: Consistency properties. Project 21-49-004, Report 4, USAF School of Aviation Medicine, Randolph Field, Texas.

Florens, J.-P. and Mouchart, M. (1982). A note on noncausality. *Econometrica*, 50(3):583–591.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438.

Harvey, A. and Oryshchenko, V. (2012). Kernel density estimation for time series data. *International Journal of Forecasting*, 28(1):3–14.

Hiemstra, C. and Jones, J. D. (1994). Testing for linear and nonlinear Granger causality in the stock price-volume relation. *The Journal of Finance*, 49(5):1639–1664.

Kilian, L. (2013). Structural vector autoregressions. In Hashimzade, N. and Thornton, M. A., editors, *Handbook of Research Methods and Applications in Empirical Macroeconomics*, Chapters, chapter 22, pages 515–554. Edward Elgar Publishing.

Kish, L. (1965). *Survey Sampling*. Wiley, New York.

Koopman, S. J. and Harvey, A. C. (2003). Computing observation weights for signal extraction and filtering. *Journal of Economic Dynamics and Control*, 27(7):1317–1333.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Marinazzo, D., Pellicoro, M., and Stramaglia, S. (2008). Kernel-Granger causality and the analysis of dynamical networks. *Physical Review E*, 77(5):056215.

National Bureau of Economic Research (2003). Business cycle dating committee announcement, July 17, 2003. National Bureau of Economic Research, Cambridge, MA.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Ofek, E. and Richardson, M. (2003). DotCom mania: The rise and fall of Internet stock prices. *The Journal of Finance*, 58(3):1113–1137.

Parzen, E. (1962). On the estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):1065–1076.

Phillips, P. C. B., Shi, S., and Yu, J. (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review*, 56(4):1043–1077.

Phillips, P. C. B., Shi, S., and Yu, J. (2015b). Testing for multiple bubbles: Limit theory of real time detectors. *International Economic Review*, 56(4):1079–1134.

Phillips, P. C. B., Wu, Y., and Yu, J. (2011). Explosive behavior in the 1990s NASDAQ: When did exuberance escalate asset values? *International Economic Review*, 52(1):201–226.

Rosenblatt, M. (1956). On some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27(3):832–837.

Shi, S., Hurn, S., and Phillips, P. C. B. (2020). Causal change detection in possibly integrated systems: Revisiting the money–income relationship. *Journal of Financial Econometrics*, 18(1):158–180.

Shojaie, A. and Fox, E. B. (2022). Granger causality: A review and recent advances. *Annual Review of Statistics and Its Application*, 9(1):289–319.

Su, L. and White, H. (2008). A nonparametric Hellinger metric test for conditional independence. *Econometric Theory*, 24(4):829–864.

Tank, A., Covert, I., Foti, N., Shojaie, A., and Fox, E. B. (2022). Neural Granger causality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4267–4279.

Troster, V. (2018). Testing for Granger-causality in quantiles. *Econometric Reviews*, 37(8):850–866.

Vicente, R., Wibral, M., Lindner, M., and Pipa, G. (2011). Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of Computational Neuroscience*, 30(1):45–67.

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *The Annals of Statistics*, 40(5):2601–2633.

Whittle, P. (1984). *Prediction and regulation by linear least-square methods*. Blackwell, Oxford, 2 edition.

Wismüller, A., Dsouza, A. M., Vosoughi, M. A., and Abidin, A. (2021). Large-scale nonlinear Granger causality for inferring directed dependence from short multivariate time-series data. *Scientific Reports*, 11(1):7817.

Wyner, A. (1978). A definition of conditional mutual information for arbitrary ensembles. *Information and Control*, 38(1):51–59.

A Proofs

This appendix proves the results of Section 4. The proofs follow DP's argument for independent and identically distributed observations in their Appendix A.1, with weights. Appendix A.1 collects properties of the weights. Appendix A.2 writes the statistic as a weighted U-statistic. Appendices A.3 and A.4 give the Hoeffding decomposition and the mean and variance of the weighted statistic. Appendix A.5 proves Theorem 4.1, and Appendix A.6 states and proves Corollary A.1 on the optimal bandwidth.

For readability, we fix t and write $\theta_s := \theta_{t,s}$, and $p_m := \sum_s \theta_s^m$ for the power sums of the weights, so that $n_K = 1/p_2$. All sums run over $\{1, \dots, n\}$, and a sum over i, j, k all different runs over all ordered (i, j, k) in which no two indices are equal. C denotes a constant that does not depend on n, h or the indices, and may change from line to line.

A.1 The weights

LEMMA A.1 (Weights): *Let Assumption 4.3 (i)–(ii) hold. Then, for n large enough:*

(a) $\sum_s \theta_s = 1$ *and* $n_K = \dfrac{(1+\gamma)^3}{(1-\gamma)(1+\gamma^2)}\big(1+o(1)\big) = \dfrac{4}{1-\gamma}\big(1+o(1)\big)$.

(b) The largest weight $\theta_{\max} := \max_s \theta_s$ *satisfies* $\theta_{\max} \le 3/n_K$, *and* $p_m \le (3/n_K)^{m-2}\, n_K^{-1}$ *for* $m \ge 2$.

(c) The weights summed over all i, j, k that are all different are

$$c_t := \sum_{i,j,k \ \text{all different}} \theta_i \theta_j \theta_k = 1 - 3p_2 + 2p_3 = 1 + O(n_K^{-1}).$$

(d) The squared weights satisfy

$$\sum_{i \neq j} \theta_i^2 \theta_j^2 = n_K^{-2}\big(1 + O(n_K^{-1})\big), \qquad \sum_{i,j,k \ \text{all different}} \theta_i^2 \theta_j^2 \theta_k^2 = n_K^{-3}\big(1 + O(n_K^{-1})\big).$$

(e) For each s, let $a_s := \sum_{j,k \neq s,\ j \neq k} \theta_j \theta_k$ *be the sum of $\theta_j \theta_k$ over all j and k that are different from each other and from s. Then* $a_s = 1 - 2\theta_s + 2\theta_s^2 - p_2$, *and* $|a_s - 1| \le C/n_K$ *for all s.*

Proof. Part (a). The weights sum to one by Equation (11). For the second argument, we state the effective sample size in terms of γ. By definition,

$$n_K = \frac{1}{\sum_s \theta_s^2} = \frac{S_t^2}{\sum_s \gamma^{2|t-s|}},$$

because $\theta_s = \gamma^{|t-s|}/S_t$. Both sums are geometric series. Summing them on both sides of t gives

$$S_t = \sum_s \gamma^{|t-s|} = \frac{1+\gamma-\gamma^t-\gamma^{n-t+1}}{1-\gamma}, \qquad \sum_s \gamma^{2|t-s|} = \frac{1+\gamma^2-\gamma^{2t}-\gamma^{2(n-t+1)}}{1-\gamma^2}.$$

The terms γ^t and γ^{n-t+1} come from the two ends of the sample. We show that they go to zero. By Assumption 4.3 (i), t lies in the interior of the sample. The distances from t to the start and to the end of the sample, $m = t$ and $m = n-t+1$, are therefore both at least cn for some $c > 0$. Since $\log x \le x - 1$ for $x > 0$, we have

$$\gamma^m = \exp(m \log \gamma) \le \exp\big(-m(1-\gamma)\big) \le \exp\big(-cn(1-\gamma)\big).$$

By Assumption 4.3 (ii), $n(1-\gamma) \to \infty$, so the exponent $-cn(1-\gamma)$ goes to $-\infty$ and $\gamma^m \to 0$. Hence γ^t and γ^{n-t+1} go to zero, and so do their squares γ^{2t} and $\gamma^{2(n-t+1)}$.

Taking these together, we have

$$n_K = \frac{S_t^2}{\sum_s \gamma^{2|t-s|}} = \frac{(1+\gamma)^2(1-\gamma^2)}{(1-\gamma)^2(1+\gamma^2)}\big(1+o(1)\big) = \frac{(1+\gamma)^3}{(1-\gamma)(1+\gamma^2)}\big(1+o(1)\big).$$

Since $\gamma \to 1$, this equals $4(1+o(1))/(1-\gamma)$.

Part (b). Let $\theta_{\max} := \max_s \theta_s$. The largest weight is the weight of observation t itself, so $\theta_{\max} = \gamma^0/S_t = 1/S_t$. Multiplying by the expression for n_K in part (a) gives

$$n_K\, \theta_{\max} = \frac{S_t}{\sum_s \gamma^{2|t-s|}} = \frac{(1+\gamma)^2}{1+\gamma^2}\big(1+o(1)\big) \to 2.$$

Hence $\theta_{\max} \le 3/n_K$ for n large. For $m \ge 2$, we write $\theta_s^m = \theta_s^{m-2}\,\theta_s^2$ and bound the first factor by the largest weight, $\theta_s^{m-2} \le \theta_{\max}^{m-2}$. Summing over s gives

$$p_m = \sum_s \theta_s^m \le \theta_{\max}^{m-2} \sum_s \theta_s^2 = \theta_{\max}^{m-2}\, p_2 \le \Big(\frac{3}{n_K}\Big)^{m-2} \frac{1}{n_K},$$

where the last step uses $p_2 = 1/n_K$ and $\theta_{\max} \le 3/n_K$.

Part (c). When we derive the U-statistic later, we have three nested sums over i, j and k, from which we exclude the terms where $i = j$, $i = k$ or $j = k$. We therefore want to know

$$c_t = \sum_{i,j,k \text{ all different}} \theta_i \theta_j \theta_k.$$

We compute this sum in three steps. First, we fix i and $j \ne i$, and let k run over all indices except i and j. Since the weights sum to one, these weights sum to exactly $1 - \theta_i - \theta_j$. Second,

we sum over all $j \neq i$:

$$\sum_{j \neq i} \theta_j \, (1 - \theta_i - \theta_j) = (1 - \theta_i) \sum_{j \neq i} \theta_j - \sum_{j \neq i} \theta_j^2 = (1 - \theta_i)^2 - (p_2 - \theta_i^2) = 1 - 2\theta_i + 2\theta_i^2 - p_2.$$

Here we use that the weights other than θ_i sum to $1 - \theta_i$, and that their squares sum to $p_2 - \theta_i^2$. Third, we multiply by θ_i and sum over all i:

$$c_t = \sum_i \theta_i \big(1 - 2\theta_i + 2\theta_i^2 - p_2\big) = 1 - 2p_2 + 2p_3 - p_2 = 1 - 3p_2 + 2p_3.$$

By part (b), $p_2 = 1/n_K$ and $p_3 \leq 3/n_K^2$. Hence $c_t = 1 + O(n_K^{-1})$.

Part (d). When we derive the variance of the U-statistic later, the same sums appear with squared weights. We therefore want to know

$$\sum_{i \neq j} \theta_i^2 \theta_j^2 \qquad \text{and} \qquad \sum_{i,j,k \text{ all different}} \theta_i^2 \theta_j^2 \theta_k^2.$$

We follow the steps of part (c) with the squared weights θ_s^2. The difference is that the squared weights sum to p_2 instead of one.

For the first sum, we fix i and let j run over all indices except i. The squared weights of these j sum to $p_2 - \theta_i^2$. Summing over i gives

$$\sum_{i \neq j} \theta_i^2 \theta_j^2 = \sum_i \theta_i^2 \, (p_2 - \theta_i^2) = p_2^2 - p_4.$$

For the second sum, we use three steps. First, we fix i and $j \neq i$, and let k run over all indices except i and j. The squared weights of these k sum to $p_2 - \theta_i^2 - \theta_j^2$. Second, we sum over all $j \neq i$:

$$\sum_{j \neq i} \theta_j^2 \, (p_2 - \theta_i^2 - \theta_j^2) = (p_2 - \theta_i^2) \sum_{j \neq i} \theta_j^2 - \sum_{j \neq i} \theta_j^4 = (p_2 - \theta_i^2)^2 - (p_4 - \theta_i^4) = p_2^2 - 2p_2\theta_i^2 + 2\theta_i^4 - p_4.$$

Here we use that the squared weights other than θ_i^2 sum to $p_2 - \theta_i^2$, and that their squares sum to $p_4 - \theta_i^4$. Third, we multiply by θ_i^2 and sum over all i:

$$\sum_{i,j,k \text{ all different}} \theta_i^2 \theta_j^2 \theta_k^2 = \sum_i \theta_i^2 \big(p_2^2 - 2p_2\theta_i^2 + 2\theta_i^4 - p_4\big) = p_2^3 - 2p_2p_4 + 2p_6 - p_2p_4 = p_2^3 - 3p_2p_4 + 2p_6.$$

The leading terms are $p_2^2 = n_K^{-2}$ and $p_2^3 = n_K^{-3}$. The other terms are small compared with these.

By part (b), $p_4 \le 9n_K^{-3}$ and $p_6 \le 81n_K^{-5}$, so that $p_4/p_2^2 \le 9/n_K$ and $p_6/p_2^3 \le 81/n_K^2$. Hence

$$\sum_{i\neq j}\theta_i^2\theta_j^2 = n_K^{-2}\big(1+O(n_K^{-1})\big), \qquad \sum_{i,j,k \text{ all different}}\theta_i^2\theta_j^2\theta_k^2 = n_K^{-3}\big(1+O(n_K^{-1})\big).$$

Part (e). The middle step of part (c) computes exactly a_s, with s in the role of i. So $a_s = 1 - 2\theta_s + 2\theta_s^2 - p_2$. Every weight is at most $\theta_{\max} \le 3/n_K$ by part (b), and $p_2 = 1/n_K$. Hence, for every s,

$$|a_s - 1| \le 2\theta_{\max} + 2\theta_{\max}^2 + p_2 \le \frac{6}{n_K} + \frac{18}{n_K^2} + \frac{1}{n_K} \le \frac{C}{n_K}.$$

□

A.2 The weighted U-statistic

Following DP, the statistic $T_t(h)$ can be written in terms of a U-statistic by symmetrization with respect to the three indices. Lemma A.2 below gives the kernel κ of this U-statistic in Equation (22), and shows that, for a given bandwidth h, $T_t(h)$ is a weighted third-order U-statistic with kernel κ and weights $\theta_i\theta_j\theta_k$, which sum to c_t (Lemma A.1 (c)).

Following DP, we define the expected values of the kernel under f_u, the density of W at rescaled time u (Section 4). Let $q_h(u) := q_h(f_u)$ be the expected value of $\kappa(W_1, W_2, W_3)$ when W_1, W_2, W_3 are independent with density f_u. Let $r_u(w,h)$ be the expected value of $\kappa(w, W_2, W_3)$ when W_2, W_3 are independent with density f_u, and let $r_{0,u}(w) := \lim_{h\to 0} r_u(w,h)$. DP define $r(w,h)$ and $r_0(w)$ in the same way, but for one density that holds for the full sample. We define them for the local density f_u at time u.

Lemma A.2 (U-statistic representation): *For every $h > 0$ and $\gamma \in (0,1]$,*

$$T_t(h) = \sum_{i,j,k \ \textit{all different}} \theta_i\theta_j\theta_k\,\kappa(W_i, W_j, W_k),$$

where the weights $\theta_i\theta_j\theta_k$ sum to c_t (Lemma A.1 (c)). For $\gamma = 1$, $c_t = (n-1)(n-2)/n^2$ and $T_t(h) = c_t\,T_n(h)$.

Proof. At time t, the density estimates in Equation (10) are $\hat{f}_{t|n}(Z_i) = (2h)^{-d_Z}\sum_{j\neq i}\theta_j I_{ij}^Z$. Substituting them gives

$$T_t(h) = (2h)^{-4}\sum_i\sum_{j\neq i}\sum_{k\neq i}\theta_i\theta_j\theta_k\Big(I_{ij}^W\,I_{ik}^{Y_{-1}} - I_{ij}^{Y_{-1}X_{-1}}I_{ik}^{Y_{-1}Y}\Big).$$

As in DP, the terms with $j = k$ are zero, so the sum runs over i, j, k that are all different. The weight $\theta_i\theta_j\theta_k$ is symmetric in i, j and k. The statistic can therefore be written as a weighted U-statistic by symmetrization with respect to the three indices, as in DP. This gives

$$T_t(h) = \sum_{i,j,k \text{ all different}} \theta_i\theta_j\theta_k \, \kappa(W_i, W_j, W_k),$$

where

$$\kappa(W_i, W_j, W_k) = \frac{(2h)^{-4}}{6} \begin{pmatrix} (I_{ij}^W I_{ik}^{Y_{-1}} - I_{ij}^{Y_{-1}X_{-1}} I_{ik}^{Y_{-1}Y}) + (I_{ik}^W I_{ij}^{Y_{-1}} - I_{ik}^{Y_{-1}X_{-1}} I_{ij}^{Y_{-1}Y}) + \\ (I_{ji}^W I_{jk}^{Y_{-1}} - I_{ji}^{Y_{-1}X_{-1}} I_{jk}^{Y_{-1}Y}) + (I_{jk}^W I_{ji}^{Y_{-1}} - I_{jk}^{Y_{-1}X_{-1}} I_{ji}^{Y_{-1}Y}) + \\ (I_{ki}^W I_{kj}^{Y_{-1}} - I_{ki}^{Y_{-1}X_{-1}} I_{kj}^{Y_{-1}Y}) + (I_{kj}^W I_{ki}^{Y_{-1}} - I_{kj}^{Y_{-1}X_{-1}} I_{ki}^{Y_{-1}Y}) \end{pmatrix}. \tag{22}$$

For a given bandwidth h, $T_t(h)$ is therefore a weighted third-order U-statistic. With the equal weight $1/(n(n-1)(n-2))$ instead of $\theta_i\theta_j\theta_k$, this is the representation of DP. The weights sum to c_t by Lemma A.1 (c). For $\gamma = 1$, all weights equal $1/n$, so that each term has weight n^{-3} instead of $1/(n(n-1)(n-2))$. Hence $T_t(h) = c_t \, T_n(h)$ with $c_t = (n-1)(n-2)/n^2$. □

A.3 Hoeffding decomposition

By Assumption 4.1, $W_1, \ldots, W_n$ are independent with density f_u. Let W' be an independent copy. The Hoeffding decomposition writes the kernel as a sum of uncorrelated terms, as in DP. Let $\kappa_2(w_1, w_2) := \mathbb{E}\, \kappa(w_1, w_2, W')$ and

$$\begin{aligned} g_1(w) &:= r_u(w, h) - q_h(u), \\ g_2(w_1, w_2) &:= \kappa_2(w_1, w_2) - r_u(w_1, h) - r_u(w_2, h) + q_h(u), \\ g_3(w_1, w_2, w_3) &:= \kappa(w_1, w_2, w_3) - \sum_{\{a,b\}} \kappa_2(w_a, w_b) + \sum_a r_u(w_a, h) - q_h(u), \end{aligned}$$

where the sums run over the three pairs and the three indices of $(1, 2, 3)$. Then

$$\kappa(w_1, w_2, w_3) = q_h(u) + \sum_a g_1(w_a) + \sum_{\{a,b\}} g_2(w_a, w_b) + g_3(w_1, w_2, w_3) \tag{23}$$

for all (w_1, w_2, w_3). The functions g_1, g_2 and g_3 are symmetric and degenerate: their expectation over any one argument, with the others fixed, is zero. As a consequence, two terms $g_\ell(W_A)$ and $g_{\ell'}(W_{A'})$, with A and A' sets of indices, are uncorrelated unless $A = A'$.

LEMMA A.3 (Weighted Hoeffding decomposition): *Let $T_t := T_t(h)$ and let a_s be as in Lemma A.1 (e).*

Then $T_t = c_t\, q_h(u) + \Lambda_t + R_{2,t} + R_{3,t}$, where the three parts are uncorrelated, with

$$\Lambda_t := 3\sum_s \theta_s a_s\, g_1(W_s), \qquad \mathbb{V}(\Lambda_t) = 9\sum_s \theta_s^2 a_s^2\ \mathbb{E}\ g_1^2,$$

$$R_{2,t} := 3\sum_{i\neq j} \theta_i\theta_j(1-\theta_i-\theta_j)\, g_2(W_i,W_j), \qquad \mathbb{V}(R_{2,t}) = 18\sum_{i\neq j} \theta_i^2\theta_j^2(1-\theta_i-\theta_j)^2\ \mathbb{E}\ g_2^2,$$

$$R_{3,t} := \sum_{i,j,k\ \textit{all different}} \theta_i\theta_j\theta_k\, g_3(W_i,W_j,W_k), \qquad \mathbb{V}(R_{3,t}) = 6\sum_{i,j,k\ \textit{all different}} \theta_i^2\theta_j^2\theta_k^2\ \mathbb{E}\ g_3^2.$$

Proof. By Lemma A.2, $T_t = \sum_{i,j,k \text{ all different}} \theta_i\theta_j\theta_k\, \kappa(W_i,W_j,W_k)$. We substitute Equation (23) for $\kappa(W_i,W_j,W_k)$ and collect the weights group by group. The constant $q_h(u)$ is multiplied by the sum of the weights, which is c_t by Lemma A.1 (c). The weight $\theta_i\theta_j\theta_k$ is symmetric in i, j and k, so the three terms $g_1(W_i)$, $g_1(W_j)$ and $g_1(W_k)$ give the same total, and so do the three terms $g_2(W_i,W_j)$, $g_2(W_i,W_k)$ and $g_2(W_j,W_k)$. For a fixed $i = s$, the weight $\theta_j\theta_k$ summed over all j and k that are different from each other and from s is a_s, by Lemma A.1 (e). For a fixed pair (i,j), the weight θ_k summed over all $k \neq i,j$ is $1-\theta_i-\theta_j$, since the weights sum to one. Taking these together gives

$$3\sum_{i,j,k \text{ all different}} \theta_i\theta_j\theta_k\, g_1(W_i) = 3\sum_s \theta_s\Big(\sum_{j,k\neq s,\ j\neq k} \theta_j\theta_k\Big) g_1(W_s) = \Lambda_t,$$

$$3\sum_{i,j,k \text{ all different}} \theta_i\theta_j\theta_k\, g_2(W_i,W_j) = 3\sum_{i\neq j} \theta_i\theta_j\Big(\sum_{k\neq i,j} \theta_k\Big) g_2(W_i,W_j) = R_{2,t}.$$

The terms $g_3(W_i,W_j,W_k)$ with their weights form $R_{3,t}$. Since the indices within each term are all different, the observations in each term are independent, and since g_1, g_2 and g_3 are degenerate, the three parts are uncorrelated. □

A.4 Mean and variance

Lemma A.4 (Mean and variance): *Let Assumptions 4.1–4.3 hold. Let s_u, $q_{2,u}$ and $q_{3,u}$ be the function s and the constants q_2 and q_3 of DP under f_u. Then $\mathbb{E}\ T_t(h) = c_t\, q_h(u)$, with $q_h(u) - q(u) = \mathbb{E}_u[s_u(W)]\, h^2 + o(h^2)$, and*

$$\mathbb{V}\big(T_t(h)\big) = \frac{9\mathbb{V}_u(r_u(W,h))}{n_K} + \frac{18\, q_{2,u}}{n_K^2 h^3} + \frac{6\, q_{3,u}}{n_K^3 h^4},$$

where each term holds up to a factor $1 + o(1)$.

With n in place of n_K, this is the variance of $T_n(h)$ of DP.

Proof. Mean. By Lemma A.3, $\mathbb{E}\ T_t(h) = c_t\, q_h(u)$, because g_1, g_2 and g_3 have mean zero. DP

show that $r(w,h) - r_0(w) = s(w)h^2 + o(h^2)$ for a density with bounded second derivatives, with an explicit function s. By Assumption 4.2, this applies to f_u. Taking expectations gives $q_h(u) - q(u) = \mathbb{E}_u[s_u(W)]h^2 + o(h^2)$.

Variance. We use the three variances of Lemma A.3. Since $c_t\, q_h(u)$ is a constant and Λ_t, $R_{2,t}$ and $R_{3,t}$ are uncorrelated,

$$\mathbb{V}\big(T_t(h)\big) = \mathbb{V}\big(\Lambda_t + R_{2,t} + R_{3,t}\big) = \mathbb{V}(\Lambda_t) + \mathbb{V}(R_{2,t}) + \mathbb{V}(R_{3,t}),$$

so we can treat the three parts separately.

Part 1: $\mathbb{V}(\Lambda_t)$. By Lemma A.1 (e), $a_s^2 = 1 + O(n_K^{-1})$ uniformly in s, and $\sum_s \theta_s^2 = n_K^{-1}$. Since $\mathbb{E}\; g_1^2 = \mathbb{V}_u(r_u(W,h))$,

$$\mathbb{V}(\Lambda_t) = 9\sum_s \theta_s^2 a_s^2\; \mathbb{E}\; g_1^2 = 9\,\mathbb{V}_u(r_u(W,h)) \sum_s \theta_s^2\big(1 + O(n_K^{-1})\big) = \frac{9\mathbb{V}_u(r_u(W,h))}{n_K}\big(1 + O(n_K^{-1})\big).$$

Part 2: $\mathbb{V}(R_{2,t})$. By Lemma A.3, $\mathbb{V}(R_{2,t}) = 18\sum_{i\neq j} \theta_i^2\theta_j^2(1-\theta_i-\theta_j)^2\; \mathbb{E}\; g_2^2$. We can treat both the weights and the expectation $\mathbb{E}\; g_2^2$ separately. By Lemma A.1 (b), every weight is at most $3/n_K$, so $(1-\theta_i-\theta_j)^2 = 1 + O(n_K^{-1})$ for all i and j. Together with Lemma A.1 (d), this gives

$$\sum_{i\neq j} \theta_i^2\theta_j^2(1-\theta_i-\theta_j)^2 = \big(1 + O(n_K^{-1})\big) \sum_{i\neq j} \theta_i^2\theta_j^2 = n_K^{-2}\big(1 + O(n_K^{-1})\big).$$

For $\mathbb{E}\; g_2^2$, we can use that $\mathbb{E}\; g_2^2 = q_{2,u}h^{-3}(1+o(1))$, which follows from DP's Condition 2. Multiplying the two gives $\mathbb{V}(R_{2,t}) = 18\, q_{2,u} n_K^{-2} h^{-3}(1+o(1))$. For the uniform kernel, DP find $q_2 = \frac{4}{36}\, 2^{-3}\; \mathbb{E}\big[f_u(W)\, f_u(Y_{-1})^2\big]$.

Part 3: $\mathbb{V}(R_{3,t})$. By Lemma A.3, $\mathbb{V}(R_{3,t}) = 6\sum_{i,j,k \text{ all different}} \theta_i^2\theta_j^2\theta_k^2\; \mathbb{E}\; g_3^2$. We can again treat the weights and the expectation $\mathbb{E}\; g_3^2$ separately. By Lemma A.1 (d),

$$\sum_{i,j,k \text{ all different}} \theta_i^2\theta_j^2\theta_k^2 = n_K^{-3}\big(1 + O(n_K^{-1})\big).$$

For $\mathbb{E}\; g_3^2$, we can use that $\mathbb{E}\; g_3^2 = q_{3,u}h^{-4}(1+o(1))$, which follows from DP's Condition 3. Multiplying the two gives $\mathbb{V}(R_{3,t}) = 6\, q_{3,u} n_K^{-3} h^{-4}(1+o(1))$. □

A.5 Proof of Theorem 4.1

Proof. Let $T_t := T_t(h)$. By Lemma A.1 (a) and Assumption 4.3 (ii), $n_K \to \infty$, and by Assumption 4.3 (iii), $h \to 0$. By Lemma A.3,

$$\sqrt{n_K}\big(T_t - q(u)\big) = \sqrt{n_K}\,\Lambda_t + \sqrt{n_K}\,(R_{2,t} + R_{3,t}) + \sqrt{n_K}\,c_t\big(q_h(u) - q(u)\big) + \sqrt{n_K}\,(c_t - 1)\,q(u).$$

We show that the first term converges to $\mathrm{N}(0, \sigma^2(u))$ and that the other three terms converge to zero.

Step 1: asymptotic normality of Λ_t. We cannot directly apply the same CLT as DP, because the weights θ_s are not equal, so that the terms of Λ_t are not identically distributed. Instead, we use Lyapunov's central limit theorem (Billingsley, 1995). It applies to a sum of independent terms with mean zero and finite third moments, if the Lyapunov ratio goes to zero.

Let us write $\Lambda_t = \sum_s \xi_s$ with $\xi_s := 3\theta_s a_s g_1(W_s)$. By Assumption 4.1, the ξ_s are independent, and they have mean zero because g_1 has mean zero under f_u. Since g_1 is bounded (Assumption 4.2), Lemma A.1 (b) and (e) give $|\xi_s| \le C\theta_s \le C/n_K$. Additionally, let $v_n^2 := \mathbb{V}(\Lambda_t)$. By Lemma A.4 and DP, $\mathbb{V}_u(r_u(W,h)) \to \mathbb{V}_u(r_{0,u}(W))$ as $h \to 0$, so $n_K v_n^2 \to \sigma^2(u) > 0$, where, as in DP,

$$\sigma^2(u) := 9\mathbb{V}_u\big(r_{0,u}(W)\big). \tag{24}$$

The Lyapunov ratio satisfies

$$\frac{\sum_s \mathbb{E}\,|\xi_s|^3}{v_n^3} \le \frac{\max_s |\xi_s| \sum_s \mathbb{E}\,\xi_s^2}{v_n^3} = \frac{\max_s |\xi_s|}{v_n} \le \frac{C/n_K}{\sigma(u) n_K^{-1/2}(1 + o(1))} \to 0.$$

Then, by Lyapunov's central limit theorem, $\Lambda_t / v_n \xrightarrow{d} \mathrm{N}(0,1)$. Since $\sqrt{n_K}\, v_n \to \sigma(u)$, $\sqrt{n_K}\,\Lambda_t \xrightarrow{d} \mathrm{N}(0, \sigma^2(u))$.

Step 2: the degenerate parts are negligible. $R_{2,t}$ and $R_{3,t}$ have mean zero, and by Lemma A.4 and Assumption 4.3 (iii),

$$n_K\,\mathbb{E}\big(R_{2,t} + R_{3,t}\big)^2 \le C\big(n_K^{-1}h^{-3} + n_K^{-2}h^{-4}\big) = C\big(n_K^{3\beta-1} + n_K^{4\beta-2}\big) \to 0,$$

because $\beta < \frac{1}{3}$. By Chebyshev's inequality, $\sqrt{n_K}(R_{2,t} + R_{3,t}) \xrightarrow{p} 0$.

Step 3: the bias is negligible. By Lemma A.4 and $c_t \le 1$, $\sqrt{n_K}\, c_t |q_h(u) - q(u)| \le C\sqrt{n_K}\, h^2 = C n_K^{1/2-2\beta} \to 0$, because $\beta > \frac{1}{4}$.

Step 4: the factor c_t is negligible. By Lemma A.1 (c), $|c_t - 1| \le C/n_K$, and $q(u)$ is bounded by Assumption 4.2. Hence $\sqrt{n_K}\,|c_t - 1|\,|q(u)| \le C n_K^{-1/2} \to 0$.

By Slutsky's theorem, Steps 1–4 give $\sqrt{n_K}(T_t - q(u)) \xrightarrow{d} \mathrm{N}(0, \sigma^2(u))$. Under $\mathcal{H}_0(u)$, $q(u) = 0$. This proves the theorem. □

A.6 The optimal bandwidth

COROLLARY A.1 (Bandwidth): *Let Assumptions 4.1, 4.2 and 4.3 (i)–(ii) hold, let $h \to 0$ with $n_K h \to \infty$, and let s_u and $q_{2,u}$ be the function s and the constant q_2 of DP under f_u, with $\mathbb{E}_u[s_u(W)] \neq 0$. Then*

$$\mathbb{E}\big(T_t(h) - q(u)\big)^2 = \big(\mathbb{E}_u[s_u(W)]\big)^2 h^4 + \frac{\sigma^2(u)}{n_K} + \frac{18\, q_{2,u}}{n_K^2 h^3} + o\Big(h^4 + \frac{1}{n_K} + \frac{1}{n_K^2 h^3}\Big). \tag{25}$$

The bandwidth that minimizes the leading terms is $h^* = C^* n_K^{-2/7}$, with the constant C^* of DP.

Proof. The mean squared error is the squared bias plus the variance, and we take both from Lemma A.4. The derivation follows that of DP for the unweighted statistic.

Bias. By Lemma A.4, the bias is $\mathbb{E}\, T_t(h) - q(u) = c_t\big(q_h(u) - q(u)\big) + (c_t - 1)\, q(u)$. By the same lemma, $q_h(u) - q(u) = \mathbb{E}_u[s_u(W)]h^2 + o(h^2)$, and $c_t = 1 + O(n_K^{-1})$ by Lemma A.1 (c). Hence the first term is $\mathbb{E}_u[s_u(W)]h^2 + o(h^2)$ and the second term is $O(n_K^{-1})$. Squaring gives $(\mathbb{E}_u[s_u(W)])^2 h^4 + o(h^4 + n_K^{-1})$.

Variance. Lemma A.4 gives

$$\mathbb{V}\big(T_t(h)\big) = \frac{9\mathbb{V}_u(r_u(W,h))}{n_K} + \frac{18\, q_{2,u}}{n_K^2 h^3} + \frac{6\, q_{3,u}}{n_K^3 h^4},$$

where each term holds up to a factor $1 + o(1)$. In the first term, $9\mathbb{V}_u(r_u(W,h)) \to \sigma^2(u)$ as $h \to 0$, so this term is $\sigma^2(u)/n_K\,(1 + o(1))$. The second term stays as it is. The third term is smaller than the second by a factor $1/(n_K h) \to 0$, so it is $o(n_K^{-2} h^{-3})$. Adding the squared bias and the variance gives Equation (25).

Optimal bandwidth. Of the three terms in Equation (25), the term $\sigma^2(u)/n_K$ does not depend on h. We therefore minimize the sum of the other two, $(\mathbb{E}_u[s_u(W)])^2 h^4 + 18 q_{2,u} n_K^{-2} h^{-3}$, over h. Setting its derivative to zero gives

$$4\big(\mathbb{E}_u[s_u(W)]\big)^2 h^3 = \frac{3 \cdot 18\, q_{2,u}}{n_K^2 h^4}, \qquad \text{so} \qquad h^* = \Big(\frac{18 \cdot 3\, q_{2,u}}{4(\mathbb{E}_u[s_u(W)])^2}\Big)^{1/7} n_K^{-2/7} = C^* n_K^{-2/7}.$$

This corresponds to the optimal bandwidth of DP, with n replaced by n_K. The constant C^* is that of DP in Appendix B, evaluated under f_u, with $a = q_{2,u}$ and $b = \mathbb{E}_u[s_u(W)]$. □

B Optimal bandwidth for finance data

DP develop a rule of thumb for finance data. For the derivation, we refer to their paper. We only state the result here. The optimal bandwidth is defined as

$$h_n^* := C^* n^{-2/7}$$

with

$$C^* := \left(\frac{18 \cdot 3a}{4b^2} \right)^{1/7}$$

In our notation, DP assume that $Y_{-1} \sim \mathcal{N}(0,1)$ and that, conditional on Y_{-1}, the variables X_{-1} and Y are independent and $\mathcal{N}(0, 1 + \delta Y_{-1}^2)$ distributed. This is a reasonable approximation for small δ. Under these assumptions, they obtain

$$a = \frac{1}{1152\sqrt{2}\pi^{3/2}} + o(\delta) \quad \text{and} \quad b = \delta^2 \left(\frac{1}{288\sqrt{3}\pi^2} + o(\delta) \right).$$

C Additional results for the S&P 500 application

This appendix reports the S&P 500 results for the smoothing parameters $\gamma = 0.998$, 0.996 and 0.992, with bandwidths $h = 0.98$, 1.20 and 1.46. The matched moving windows have 2,001, 1,001 and 501 observations, which give the same number of lagged pairs as the interior effective sample sizes of the local DP test. Table 5 reports the rejection percentages for every filter and benchmark, and Table 6 shows how often the matched local and moving-window DP tests reach the same decision. Both tables use, for each γ, the moving-window centers that are interior dates of the local DP test. Figures 8 and 9 show the local and moving-window p-values before and after volatility filtering. Figure 10 shows the parametric benchmark.

Table 5: Percentage of dates with a rejection of GNC, all smoothing settings.

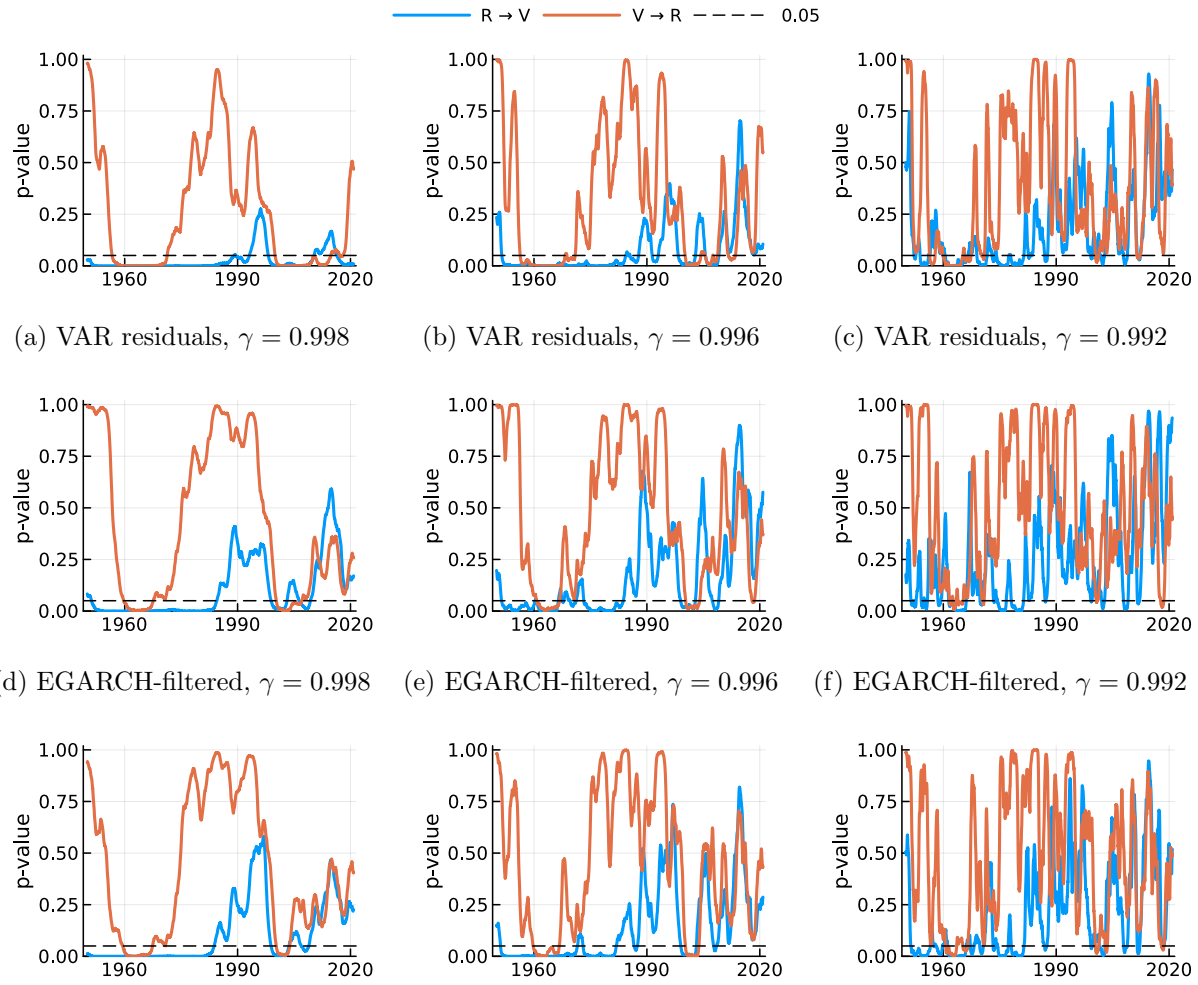
	VAR residuals			EGARCH-filtered			EWMA-standardized		
Specification	R → V	V → R	Joint	R → V	V → R	Joint	R → V	V → R	Joint
$\gamma = 0.998$, $h = 0.98$	81.6	46.7	39.6	59.8	23.3	18.2	58.0	18.3	18.0
$\gamma = 0.996$, $h = 1.20$	66.5	34.9	29.8	48.4	15.6	12.5	50.9	15.8	14.7
$\gamma = 0.992$, $h = 1.46$	34.9	15.4	9.9	25.1	6.5	4.3	36.1	9.1	4.8
Moving-window DP, $m = 2{,}001$	84.7	48.1	44.2	60.4	26.8	24.2	60.6	24.0	23.3
Moving-window DP, $m = 1{,}001$	68.2	30.0	26.0	52.1	12.5	10.6	51.4	13.2	12.7
Moving-window DP, $m = 501$	43.9	17.9	13.0	25.0	7.9	3.8	35.8	9.8	6.4
Moving-window Cheung–Ng	18.9	4.3	2.5	—	—	—	—	—	—

Notes: Percentage of dates with $p < 0.05$, evaluated for each γ at the centers of the matched moving windows, every fifth trading day, that are interior dates of the local DP test. The moving-window Cheung–Ng test uses 1,001 observations per window and the grid for $\gamma = 0.996$. Joint means a rejection in both directions at the same date. EGARCH-filtered: return residuals divided by their EGARCH(1,1) volatility, volume residuals unchanged. EWMA-standardized: both residual series divided by their EWMA volatility.

Table 6: Agreement between the matched local and moving-window DP tests.

	R → V				V → R			
Setting	Local	Window	Both	Same	Local	Window	Both	Same
$\gamma = 0.998$, $m = 2{,}001$	81.6	84.7	78.9	91.5	46.7	48.1	43.8	92.9
$\gamma = 0.996$, $m = 1{,}001$	66.5	68.2	61.1	87.5	34.9	30.0	26.5	88.0
$\gamma = 0.992$, $m = 501$	34.9	43.9	28.8	78.7	15.4	17.9	12.0	90.6

Notes: Percentage of common dates on which the local DP test (Local), the moving-window test with m observations (Window), or both tests (Both) reject at the 5% level on the VAR residuals, and on which the two tests reach the same decision (Same). Common dates as in Table 5.



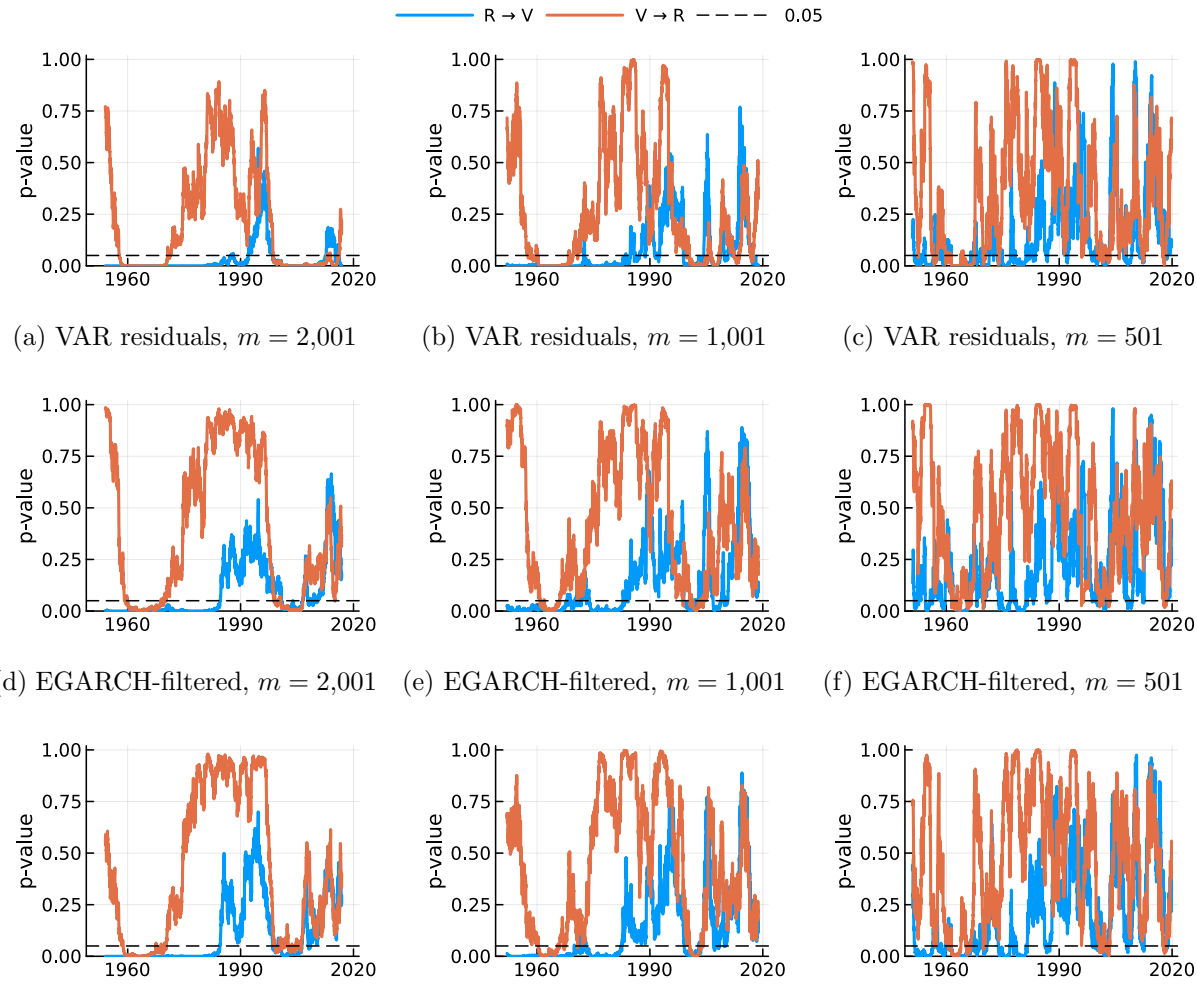
(a) VAR residuals, $\gamma = 0.998$ (b) VAR residuals, $\gamma = 0.996$ (c) VAR residuals, $\gamma = 0.992$

(d) EGARCH-filtered, $\gamma = 0.998$ (e) EGARCH-filtered, $\gamma = 0.996$ (f) EGARCH-filtered, $\gamma = 0.992$

(g) EWMA-standardized, $\gamma = 0.998$ (h) EWMA-standardized, $\gamma = 0.996$ (i) EWMA-standardized, $\gamma = 0.992$

FIGURE 8: Local DP test before and after volatility filtering, all smoothing settings.

Note: Pointwise p-values of the local DP test in both directions, on the VAR residuals (top row), with the return residuals divided by their EGARCH(1,1) volatility (middle row), and with both residual series divided by their EWMA volatility (bottom row), with $h = 0.98$, 1.20 and 1.46 for $\gamma = 0.998$, 0.996 and 0.992.



(a) VAR residuals, $m = 2{,}001$ (b) VAR residuals, $m = 1{,}001$ (c) VAR residuals, $m = 501$

(d) EGARCH-filtered, $m = 2{,}001$ (e) EGARCH-filtered, $m = 1{,}001$ (f) EGARCH-filtered, $m = 501$

(g) EWMA-standardized, $m = 2{,}001$ (h) EWMA-standardized, $m = 1{,}001$ (i) EWMA-standardized, $m = 501$

FIGURE 9: Moving-window DP test before and after volatility filtering.

Note: Pointwise p-values in both directions, with $h = 0.98$, 1.20 and 1.46 for $m = 2{,}001$, 1,001 and 501. Each value is plotted at the center of its m-observation window. Filters as in Figure 8.

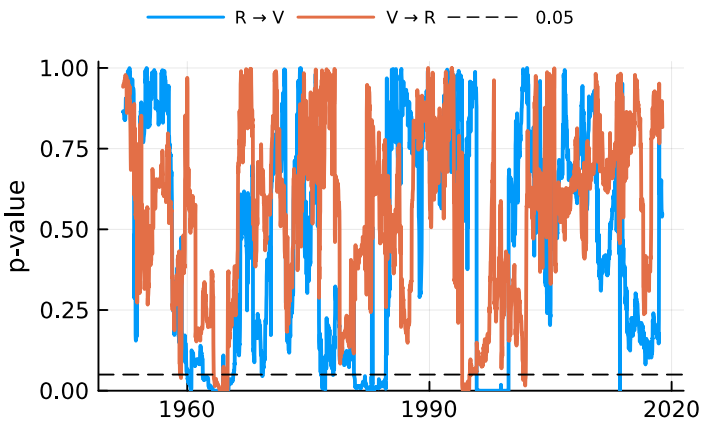


FIGURE 10: Moving-window Cheung–Ng causality-in-variance test.

Note: Pointwise p-values for the one-lag test on the VAR residuals. Each value is plotted at the center of its 1,001-observation window.

D Computation

This appendix describes the EWMA recursions behind the weights in Equation (11) and how we compute the statistic. The exponential weighting gives the filter recursion

$$\hat{f}_{t+1|t}(z) = \hat{f}_{t|t-1}(z) + (1-\gamma)\,\nu_t(z), \qquad t = 1, \ldots, n, \tag{26}$$

with the innovation

$$\nu_t(z) = (2h)^{-d_Z} I(\|z - Z_t\|_\infty \leq h) - \hat{f}_{t|t-1}(z). \tag{27}$$

Equivalently, $\hat{f}_{t+1|t}(z) = \gamma\,\hat{f}_{t|t-1}(z) + (1-\gamma)(2h)^{-d_Z} I(\|z - Z_t\|_\infty \leq h)$, so each step moves the estimate a fraction $1-\gamma$ towards the kernel of the newest observation. The estimate is a convex combination of nonnegative terms and therefore nonnegative. The closer γ is to 0, the more weight new observations get. For large t, observation Z_{t-k} receives the weight $(1-\gamma)\gamma^k$, $k = 0, 1, 2, \ldots$, in $\hat{f}_{t+1|t}$. The recursive structure is efficient to evaluate in real time. Evaluated at the n observations, each step of the recursion updates n density values with one new observation. This costs $\mathcal{O}(n)$ operations per step and $\mathcal{O}(n^2)$ for the full sample. Computing every density value directly as a weighted sum over all observations costs $\mathcal{O}(n)$ per value and time point, and therefore $\mathcal{O}(n^3)$ for the full sample.

When evaluating the densities in-sample, the filter is suboptimal, since it only uses past observations. The backward smoothing recursion also assigns weight to future observations. Let

$$r_{t-1}(z) = \gamma\, r_t(z) + \gamma\, \nu_t(z), \qquad t = n, \ldots, 2, \tag{28}$$

with $r_n(z) = 0$. The smoothed density estimate is

$$\hat{f}_{t|n}(z) = \hat{f}_{t|t-1}(z) + (1-\gamma)\big(r_t(z) + \nu_t(z)\big), \qquad t = 1, \ldots, n. \tag{29}$$

The filter and the smoother require an initial density $\hat{f}_{1|0}$, and the influence of this initialization fades out only slowly when γ is close to one. We therefore do not run the recursions. Instead, we compute the observation weights directly, which is possible for any linear filter (Koopman and Harvey, 2003). This gives the weights in Equation (11).

The direct weights have computational benefits in an offline analysis. The density estimates at different time points do not depend on each other, so they can be computed in parallel, which

is useful in simulation studies and other repeated experiments. Our implementation computes the weighted sums on a GPU and skips observations with $\gamma^{|t-s|} < 10^{-6}$. For $\gamma = 0.992$, each density value then sums over at most about 3,400 observations.