On the sensitivity of Granger Causality to errors-in-variables, linear transformations and subsampling [a]

Brian D. O. Anderson [b]
Australian National University

Manfred Deistler [c]
Technische Universität Wien

Jean-Marie Dufour [d]
McGill University

First version: March 2017
Revised: January 2018, April 2018, August 2018, September 2018
This version: September 2018
Compiled: September 12, 2018, 1:00pm

[a]This work is supported by the Australian Research Council's Discovery Project DP-160104500, CSIRO-Data61, Austrian FWF grant project number P24198-N18, the William Dow Chair of Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec).

[b]Hangzhou Dianzi University, the Research School of Engineering, ANU College of Engineering and Computer Science, Australian National University, 115 North Road, Acton ACT 2601, Australia and Data61-CSIRO. e-mail: `brian.anderson@anu.edu.au`

[c]MTechnische Universität Wien Institut für Stochastik und Wirtschaftsmathematik Forschungsgruppe Ökonometrie und Systemtheorie, Wiedner Hauptstr. 8, 4.OG gelb, A 1040 Wien Österreich.
e-mail: `manfred.deistler@tuwien.ac.at`

[d]William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill Uniersity, Leacock Building, Room 414, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 6071; FAX: (1) 514 398 4800.
e-mail: `jean-marie.dufour@mcgill.ca`. Web page: http://www.jeanmariedufour.com

Abstract

This paper studies the sensitivity of Granger causality to the addition of noise, the introduction of subsampling, and the application of causal invertible filters to weakly stationary processes. Using canonical spectral factors and Wold decompositions, we give general conditions under which additive noise or filtering distorts Granger-causal properties by inducing (spurious) Granger causality, as well as conditions under which it does not. For the errors-in-variables case, we give a continuity result, which implies that: a "small" noise-to-signal ratio entails "small" distortions in Granger causality. On filtering, we give general necessary and sufficient conditions under which "spurious" causal relations between (vector) time series are not induced by linear transformations of the variables involved. This also yields transformations (or filters) which can eliminate Granger causality from one vector to another one. In a number of cases, we clarify results in the existing literature, with a number of calculations streamlining some existing approaches.

Keywords: Granger causality; Sensitivity; Signal-to-noise ratio; Errors-in-variables; Measurement errors; Filtering, Subsampling.
MSC code: 62M10

Contents

1 Introduction

Granger causality is one of the most important concepts for the analysis of the structure of multivariate time series. Accordingly, the original paper of Granger (1969) triggered a substantial number of publications, see *e.g.* Sims (1972), Pierce and Haugh (1977), Granger (1980, 1988), Geweke (1982, 1984*a*, 1984*b*), Boudjellaba, Dufour and Roy (1992), Dufour and Tessier (1993), Dufour and Renault (1998), Al-Sadoon (2014) and the references therein. Here we deal with an aspect of Granger causality, namely the sensitivity of Granger causality relations with respect to measurement errors (or errors-in-variables) in the observations. In particular, we study the effect of additive noise on Granger causality in the context of a general weakly stationary multivariate model, especially in view of finding when spurious causality could appear, and when properties of non-causality are unaffected by measurement errors.

The problem of measurement errors is a classical issue in statistical theory; see *e.g.* the reviews of Fuller (1987), Wansbeek and Meijer (2000), Carroll, Ruppert, Stefanski and Crainiceanu (2006), Gustafson (2003), and Buonaccorsi (2010). However, except for the early article by Newbold (1978), there is surprisingly little work on the effect of errors-in-variables on Granger causality. In this work, Newbold showed that measurement errors can produce artificial feedback in the noisy series, even though no such feedback is present before noise is superimposed. No general characterization of cases where such spurious causality could appear was however provided. From a wider perspective, several authors have emphasized that the addition of noise to time series (errors-in-variables) can substantially modify the structure of the series, leading to distortions and identification problems; see *e.g.* the literature reviewed by Maravall (1979), Anderson and Deistler (1984), Anderson (1985), Deistler and Anderson (1989), and Scherrer and Deistler (1998). Note also that measurement errors may give rise to additive "outliers" which may strongly influence the results of estimation and testing procedures.

The question of the sensitivity to measurement errors is quite distinct for the effect of aggregation and subsampling, for these transformations typically considerably reduce the effective sample size. For work on the latter problems, the reader may consult Tiao and Wei (1976), Wallis (1974), Sims (1974), Wei (1982), Hylleberg (1986), Marcellino (1999), Kaiser and Maravall (2001), Breitung and Swanson (2002), McCrorie and Chambers (2006) Barnett and Seth (2011, 2015, 2017), Smirnov and Bezruchko (2012), Gong, Zhang, Schlökopf, Tao and Geiger (2015), Ghysels, Hill and Motegi (2016), and the references in the survey of Silvestrini and Veredas (2008).

Errors-in-variables can be interpreted as missing variables : if the noise were observable, it could be included as an additional variable, and different conclusions can emerge. As previously observed by several authors [see Hsiao (1982), Lütktepohl (1982), Dufour and Renault (1998), Triacca (1998, 2000)], causality properties in the sense of Wiener-Granger depend crucially on the information set considered, which can affect both the sheer presence of causality (or non-causality) and causality measures [Geweke (1982), Dufour and Taamouti (2010), Dufour, Garcia and Taamouti (2012)]. Of course, the central difficulty remains that noise is typically unobserved. In this paper, we revisit the questions of the effect of (unobserved) additive noise on Granger (non-)causality, and using the same tools, rapidly traverse also issues of the effects of filtering and subsampling.

Let $X = (X(t) \,|\, t \in \mathbb{Z}),\ X(t) : \Omega \to \mathbb{R}^d$, be a vector process of dimension d with finite second moments, where $\mathbb{Z}$ represents the integers and $\mathbb{R}$ the real numbers. We assume that X is weakly stationary, centered (*i.e.* $\mathsf{E}[X(t)] = 0$) and Gaussian, with a full-rank rational spectral density.[1] We postulate that the process X can be regarded as a juxtaposition of two subprocesses $X = (X_A^\top \ X_B^\top)^\top$. The broad question we study is whether the past values of X_A improve the prediction of X_B. To be more precise, one says that X_A does not Granger cause X_B if

$$\mathsf{E}[X_B(t) \,|\, X_A(s), X_B(s) : s < t] = \mathsf{E}[X_B(t) \,|\, X_B(s) : s < t] \tag{1}$$

or equivalently

$$\mathrm{Var}[X_B(t) \,|\, X_A(s), X_B(s) : s < t] = \mathrm{Var}[X_B(t) \,|\, X_B(s) : s < t]\,. \tag{2}$$

Here $\mathsf{E}[X_B(t) \,|\, X_A(s), X_B(s) : s < t]$ denotes the conditional expectation of $X_B(t)$ [given the variables $X_A(s), X_B(s)$ such that $s < t$ (and similarly elsewhere)], and Var the variance of the one-step-ahead forecast error. If inequality holds in (1) and (2), one says that X_A (Granger) causes X_B. Granger (1969) in addition introduced the notion of "instantaneous causality", meaning that the approximation of $X_B(t)$ can be more accurately achieved if $X_A(t)$ is known:

$$\mathsf{E}[X_B(t) \,|\, X_A(t), X_A(s), X_B(s) : s < t] \neq \mathsf{E}[X_B(t) \,|\, X_A(s), X_B(s) : s < t]\,; \tag{3}$$

for further discussion of this notion, see Pierce and Haugh (1977) and Granger (1988). The assumption of second-order stationarity is clearly restrictive, but is standard in the Granger-causality literature. Further, general characterizations of non-causality are typically little affected when common forms of forms of non-stationarity – such deterministic time trends and integration) – are allowed; see, for example, Dufour and Renault (1998) and Dufour, Pelletier and Renault (2006).

[1] Without the Gaussian assumption, the results presented in this paper continue to hold provided conditional expectations are replaced by projections onto the Hilbert space spanned by components of the respective stationary processes.

It is clear from the above definitions that Granger causality depends on the vector X considered and on the way X is split into subvectors X_A and X_B. Such choices (which are of course finite in number) depend on the context: which variables are of interest, and the objectives of the analysis. For example, X_A can represent policy instruments (*e.g.*, fiscal and monetary variables) or leading indicators of economic activity, and X_B economic outcomes (*e.g.*, national income, unemployment, etc.): the nature of the variables often provides a natural criterion for splitting X into subvectors. Clearly, the causal structure of a time series should in general depend on such choices. However, the question remains whether apparently less fundamental features, such as contamination by noise and various linear transformations, including filtering and subsampling, can affect the causal properties of a time series.

This paper studies the sensitivity of Granger causality to the addition of noise, the application of causal invertible filters, and subsampling in weakly stationary processes. We give general conditions under which additive noise or filtering creates distortions by inducing (spurious) Granger causality, as well as conditions under which it does not. Even though additive noise and filtering can in general produce spurious Granger causality, there is a remarkably wide range of cases where it does not. For example, if the "caused variable" X_B is not noisy, noise added to the "causal variable" X_A cannot induce spurious Granger causality from X_A to X_B. This covers cases where lagged values of X_A are contaminated by noise, and X_B does Granger-cause X_A. We also give a continuity result which entails a "small" noise-to-signal ratio in measurement errors entails "small" distortions in Granger causality. In a number of cases, we clarify results in the existing literature, with a number of calculations streamlining some existing approaches.

We also consider the effects of linear transformations, filtering and subsampling. In particular, we give general necessary and sufficient conditions under which "spurious" causal relations between (vector) time series are not be induced by linear transformations of the variables involved. This also yields linear transformations (or filters) which can eliminate Granger causality from one vector to another one.

Section 2 summarizes a collection of known results available for the characterization of Granger causality, using canonical spectral factors, Wold decompositions and spectra. In Section 3, we establish some connections not clearly stated in the earlier literature, which are useful for studying causality in the presence of measurement errors. These include : a general lower bound on the conditional variance of the sum of two processes, and some general relations between Granger causality and instantaneous causality. In Section 4, we study the effect of measurement errors on Granger non-causality. Section 5 provides the continuity result in terms of signal-to-noise ratio. The effects of linear transformations, filtering and subsampling are studied in Sections 6 and 7. Section 8 offers some concluding remarks. Proofs appear in the Appendix.

2 Characterizations of Granger causality

In this section, we review some classical characterizations of Granger causality which will be useful for studying the effect of errors-in-variables. We first record some notational conventions associated with rational (matrix) transfer functions [see *e.g.* Rozanov (1967), Hannan and Deistler (2012)]. We emphasize the use of spectral methods, for which Geweke (1982, 1984*a*, 1984*b*) was an early promoter in the context of analyzing Granger-Wiener causality.

A rational transfer function is called *stable* if its poles are outside the unit circle, and it is called *miniphase* or *minimum phase* if its zeros are outside the unit circle. If we commence from a rational spectral density $\Phi_{XX}(z)$, $z \in \mathbb{C}$, which is positive definite everywhere on the unit circle, there is a spectral factorization

$$\Phi_{XX}(z) = W(z)\,Q\,W^{\top}(z^{-1}) \tag{4}$$

in which the spectral factor $W(z)$ is a square real rational, stable and miniphase, transfer function and Q is positive definite symmetric; see Rozanov (1967), Hannan and Deistler (2012). $W(z)$ defines a linear filter upon replacing z by the backshift operator L [i.e., $LX(t) := X(t-1)$]. The notation $W(z)$ allows one to study the properties of lag operators in terms of the analytical properties of functions of a complex variable $z \in \mathbb{C}$. Under the normalization $W(0) = I_d$, $W(z)$ and Q are unique. We also consider the following assumption.

Assumption 1. (***Full rank stationary process with no spectral zero on the unit circle***) *$X = (X_A^{\top}\ X_B^{\top})^{\top}$ is a real full-rank stationary stochastic process in $\mathbb{R}^d$, with rational spectrum $\Phi_{XX}(z)$ having no zero on the unit circle, such that* (4) *is satisfied, $W(0) = I_d$, and*

$$W(z) = \begin{bmatrix} W_{11}(z) & W_{12}(z) \\ W_{21}(z) & W_{22}(z) \end{bmatrix}, \quad Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \tag{5}$$

are partitioned conformably with $X = (X_A^{\top}\ X_B^{\top})^{\top}$.

The above assumption entails that $X(t)$ has both a moving average (Wold) representation

$$X(t) = W(L)\varepsilon(t) \tag{6}$$

and an autoregressive representation

$$\Pi(L)X(t) = \varepsilon(t) \tag{7}$$

where $\det[W(z)] \neq 0$ and $\det[\Pi(z)] \neq 0$ for $|z| < 1$, $\Pi(z) = W(z)^{-1}$, and $\varepsilon(t) = [\varepsilon_A(t)^\top\ \varepsilon_B(t)^\top]^\top$ represents the innovations of the process, partitioned conformably with $X = (X_A^\top\ X_B^\top)^\top$. The following theorems provide characterizations of Granger causality; see Sims (1972), Pierce and Haugh (1977), Geweke (1982, 1984*a*, 1984*b*), Boudjellaba et al. (1992), Dufour and Tessier (1993), Dufour and Renault (1998). The first one is based on the structure of the spectral factor matrix $W(z)$.

Theorem 1. (***Canonical spectral factor characterization of Granger causality***) *Suppose Assumption 1 holds. Then the following two conditions are equivalent:*

1. *X_A does not Granger cause X_B;*
2. *$W_{21}(z) = 0$.*

The following conditions are also equivalent:

1. *X_A neither Granger causes X_B, nor does it cause X_B instantaneously;*
2. *$W_{21}(z) = 0$ and Q is block diagonal (i.e. $Q_{12} = Q_{21}^\top = 0$).*

The intuition behind the above claim is the following. Let the innovation process be denoted by $\varepsilon(t) = [\varepsilon_A(t)^\top\ \varepsilon_B(t)^\top]^\top$ with ε_A and ε_B two independent white noise processes. When $W_{21}(z) = 0$, we have :

$$\begin{aligned} X_A(t) &= W_{11}(L)\varepsilon_A(t) + W_{12}(L)\varepsilon_B(t)\,, \\ X_B(t) &= W_{22}(L)\varepsilon_B(t)\,. \end{aligned} \tag{8}$$

It is intuitively reasonable to conclude from these equations that knowledge of the X_A process up till time $t-1$ will not be of help in determining the ε_B process and thus the X_B process. Spectral approaches for Granger causality analysis were emphasized in the seminal work of Geweke (1982, 1984*a*, 1984*b*).

For completeness, we note a further characterization of Granger causality, which follows from the above.

Theorem 2. (***AR characterization of Granger causality***) *Suppose Assumption 1 holds, and $X(t)$ has the (possibly infinite) autoregressive representation*

$$X(t) = \sum_{i=1}^{\infty} A_i X(t-i) + \varepsilon(t)\,, \quad A_i = \begin{bmatrix} A_{i11} & A_{i12} \\ A_{i21} & A_{i22} \end{bmatrix}, \quad \mathrm{Var}[\varepsilon(t)] = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{9}$$

where the A_i and the covariance matrix $\mathrm{Var}[\varepsilon(t)]$ of the innovations sequence $\varepsilon(t)$ are partitioned conformably with $X = (X_A^\top\ X_B^\top)^\top$. Then X_A does not Granger cause X_B if and only if $A_{i21} = 0$ for all $i \geq 1$. In addition, X_A neither Granger causes X_B, nor does it cause X_B instantaneously if and only if $A_{i21} = 0$ for all $i \geq 1$ and $\Sigma_{12} = \Sigma_{21}^\top = 0$.

Theorems 1 and 2 give characterizations of the absence of causality based on the spectral factor and infinite AR representations (the latter is obtained from the inverse of the spectral factor). Sims (1972) gave an additional characterization (for $d = 2$), based on Wiener filtering ideas, where no factorization is required. Let the spectral density Φ_{XX} be partitioned conformably with $X = (X_A^\top\ X_B^\top)^\top$ as

$$\Phi_{XX} = \begin{bmatrix} \Phi_{AA} & \Phi_{AB} \\ \Phi_{BA} & \Phi_{BB} \end{bmatrix}. \tag{10}$$

Then we have the following spectral characterization of non-causality.

Theorem 3. (***Transfer function characterization of Granger causality***) *Suppose Assumption 1 holds, and let Φ_{XX} be partitioned as in* (10)*. Then, the following conditions are equivalent :*

1. *X_A does not Granger cause X_B ;*
2. *$\Phi_{AB}(z)\Phi_{BB}^{-1}(z)$ is a stable transfer function.*

The following conditions are also equivalent:

1. *X_A neither Granger causes X_B nor does it cause X_B instantaneously;*
2. *$\Phi_{AB}(z)\Phi_{BB}^{-1}(z)$ is a stable transfer function assuming the value 0 at $z = 0$.*

Remark 1 The above theorem can be viewed as an extension of the corresponding theorem given by Sims (1972, Theorem 2) in the special case where $d = 2$. Theorem 3 allows for $d \geq 2$, and covers instantaneous causality as well.[2] We are not contending that the characterization of this theorem is necessarily attractive from a computational point of view. As later parts of the paper show though, the result is of theoretical interest, in that it can be applied to give rapid derivations of the sensitivity properties associated with Granger causality.
Remark 2 The transfer function $\Phi_{AB}(z)\Phi_{BB}^{-1}(z)$ is the transfer function of the optimum two-sided Wiener filter for approximating the process X_A from the process X_B; the two-sided aspect refers both to the fact that the transfer function has a Laurent series expansion with both negative and positive powers of z, and to the related fact that $X_A(t)$ is being approximated from $X_B(s), -\infty < s < \infty$, *i.e.* from the past and future of X_B. If the two-sided transfer function in a particular case is causally one-sided, then future values of X_B are irrelevant in approximating current values of X_A. This will be the case if past values of X_A do not affect present or future values of X_B.
Remark 3 It is important to note that the characterizations given in this section hold for series in discrete time observed at a given frequency. They are directly applicable to continuous time series, and modifications arise typically when the series are transformed or filtered. The effect of such transformations will be considered in sections 6 and 7 below.

3 Directions of Granger causality

In the literature, one finds remarkable similarity between conditions said to capture "X_A does not cause X_B" and "X_B causes X_A" and similar pairings. To study the effect of errors-in-variables on causality, we establish in this section some connections not clearly stated in the earlier literature. We start with the following preliminary result.

Lemma 1. *Let X and Y be two independent stationary stochastic processes with spectral densities. Let $Z = X + Y$. Then the covariance matrix of the one step prediction error in approximating $Z(t+1)$ from $Z(s),\ s \leq t$ is bounded from below by the sum of the covariance matrices of the one step prediction error in approximating $X(t+1)$ from $X(s),\ s \leq t$ and in approximating $Y(t+1)$ from $Y(s),\ s \leq t$:*

$$\mathrm{Var}[Z(t)\,|\,Z(s) : s < t] \geq \mathrm{Var}[X(t)\,|\,X(s) : s < t] + \mathrm{Var}[Y(t)\,|\,Y(s) : s < t]\,. \tag{11}$$

Now we spell out the following relations between Granger causality and instantaneous causality.

Theorem 4. *Adopt the same hypothesis as in Theorem 1. Suppose X_A does not Granger cause X_B nor does it cause X_B instantaneously. Then either the two processes are independent, or X_B Granger causes X_A. Further, suppose alternatively that X_A does not cause X_B. Then, either the two processes are independent, or X_B Granger causes X_A, or X_B causes X_A instantaneously.*

Note that neither claim of the theorem goes in the reverse direction. This is because it is possible that both X_A Granger causes X_B and simultaneously X_B Granger causes X_A. Such a situation will generally arise when the canonical spectral factor W is not triangular (or diagonal), as in the following example:

$$X_A(t) = \varepsilon_A(t) + X_B(t-1)\,, \quad X_B(t) = \frac{1}{2}X_A(t-1) + \varepsilon_B(t)\,. \tag{12}$$

Here, $\varepsilon_A, \varepsilon_B$ are independent white noise processes with variances Q_A, Q_B. One can verify that

$$\begin{bmatrix} X_A(t) \\ X_B(t) \end{bmatrix} = \frac{1}{1+(1/2)L^2} \begin{bmatrix} 1 & L \\ (1/2)L & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_A(t) \\ \varepsilon_B(t) \end{bmatrix} \tag{13}$$

and the transfer function matrix is easily verified to be stable and minimum phase, assuming the value I when $z = 0$. It is easily checked that $\mathrm{Var}[X_A(t)\,|\,X_A(s), X_B(s), s < t] = Q_A$, $\mathrm{Var}[X_B(t)\,|\,X_B(s), X_A(s), s < t] = Q_B$ while $\mathrm{Var}[X_A(t)\,|\,X_A(s), s < t] > Q_A$, $\mathrm{Var}[X_B(t)\,|\,X_B(s), s < t] > Q_B$ by a similar argument to that used in the proof of Theorem 4.

4 Additive noise and Granger causality

In this section, we consider the effect of additive noise on Granger causality [compare with Anderson and Deistler (1984) and Anderson (1985)]. Our starting point, again, is the full-rank stationary process $X = [X_A^\top\ X_B^\top]^\top$ with rational spectral density.

Suppose that X_A does not Granger cause X_B. Suppose further that the processes X_A, X_B are both contaminated by stationary colored additive noise processes N_A, N_B with rational spectral densities, which are independent of each other and of the

[2] There may be a proof in the literature for $d \geq 2$, but we are not aware of it. For completeness, a proof appears in the appendix.

processes X_A, X_B. Then one can ask whether it is now true that the process $\bar{X}_A = X_A + N_A$ does not Granger cause the process $\bar{X}_B = X_B + N_B$. Perhaps of equal if not greater interest is the associated question: suppose that $\bar{X}_A, \bar{X}_B$ are regarded as noisy measurements of underlying processes X_A, X_B and that analysis of measurement data reveals that $\bar{X}_A$ does not cause $\bar{X}_B$. Can one conclude then that X_A does not Granger cause X_B?

In the next subsection, we will construct an example showing that the answer to the first question is generally no, a conclusion that is perhaps not counterintuitive since non-causality corresponds to zero restrictions. In the following subsection, we show how the Sims (1972) characterization of the absence of Granger causality summarized in Theorem 3 reveals that the claim remains valid if the contaminating noise N_B is zero, and this is generically a necessary condition for the claim to hold. There is no similar requirement on the noise N_A. In a paper of Solo (2007), several important questions are raised about the sensitivity of Granger causality (or its absence) to changes in the underlying assumptions. In this section, we consider one of these, namely the effect of additive noise. Our results differ from those obtained in Solo (2007).[3] We first study the stationary full-rank vector process $X = \{X(t) : t \in \mathbb{Z}\}$ such that $X(t) = [X_A(t)^\top \; X_B(t)^\top]^\top$ can be regarded as the juxtaposition of two subprocesses X_A and X_B. Suppose that X_A does not Granger cause X_B nor does it cause X_B instantaneously.

4.1 Noise-induced Granger causality

We will now introduce the promised example. To define the X_A, X_B processes where X_A does not Granger cause X_B nor does it cause X_B instantaneously, following Theorem 1 we shall choose an upper triangular canonical spectral factor. The two processes are scalar, and we assume

$$W(z) = \begin{bmatrix} 1+\frac{1}{2}z & z \\ 0 & 1+\frac{1}{2}z \end{bmatrix} \tag{14}$$

and we further assume the innovations covariance Q is the identity matrix. An easy calculation delivers

$$\Phi_{XX} = \begin{bmatrix} \Phi_{AA} & \Phi_{AB} \\ \Phi_{BA} & \Phi_{BB} \end{bmatrix} = \begin{bmatrix} \frac{9}{4}+\frac{1}{2}z+\frac{1}{2}z^{-1} & \frac{1}{2}+z \\ \frac{1}{2}+z^{-1} & \frac{5}{4}+\frac{1}{2}z+\frac{1}{2}z^{-1} \end{bmatrix}. \tag{15}$$

Now assume that additive noise with a white spectrum of intensity $\frac{3}{4}$ is added to X_B, to produce a new process $\bar{X}_B$, while no noise is added to X_A. The cross spectrum between X_A and X_B is unaffected. So the new joint spectral matrix is

$$\Phi_{\bar{X}\bar{X}} = \begin{bmatrix} \Phi_{AA} & \Phi_{A\bar{B}} \\ \Phi_{\bar{B}A} & \Phi_{\bar{B}\bar{B}} \end{bmatrix} = \begin{bmatrix} \frac{9}{4}+\frac{1}{2}z+\frac{1}{2}z^{-1} & \frac{1}{2}+z \\ \frac{1}{2}+z^{-1} & 2+\frac{1}{2}z+\frac{1}{2}z^{-1} \end{bmatrix}. \tag{16}$$

If it were true that $\bar{X}_A$ does not Granger cause $\bar{X}_B$, nor cause $\bar{X}_B$ instantaneously, then this matrix would need to have a canonical spectral factor $\bar{W}(z)$ say, which like $W(z)$ is upper triangular with $\bar{W}(0) = I$, and an associated innovations covariance matrix which is diagonal. In order to derive a contradiction, let us assume this to be the case and find $\bar{W}(z)$. The upper triangularity implies that the $(2,2)$ term $\bar{W}_{22}$ of $\bar{W}(z)$ must satisfy $\bar{W}_{22}(0) = I$ and

$$\Phi_{\bar{B}\bar{B}}(z) = \bar{W}_{22}(z)\bar{Q}_2\bar{W}_{22}(z^{-1}), \tag{17}$$

which means that $\bar{W}_{22}(z)$ itself is a canonical spectral factor, for $\Phi_{\bar{B}\bar{B}}(z)$. One can easily verify that

$$2+\frac{1}{2}z+\frac{1}{2}z^{-1} = (1+\frac{\sqrt{3}}{2})(1+\frac{z}{2+\sqrt{3}})(1+\frac{z^{-1}}{2+\sqrt{3}}), \tag{18}$$

so we see that

$$\bar{W}_{22}(z) = 1+\frac{z}{2+\sqrt{3}}, \quad \bar{Q}_2 = 1+\frac{\sqrt{3}}{2} \tag{19}$$

Now consider the $(1,2)$ entry $\Phi_{A\bar{B}}(z)$ of the spectrum. From the fact that when $\bar{W}(z)$ is triangular, we have that

$$\Phi_{A\bar{B}}(z) = \bar{W}_{12}(z)\bar{Q}_2\bar{W}_{22}(z^{-1}) \tag{20}$$

from which we obtain

$$\frac{1}{2}+z = \bar{W}_{12}(z)(1+\frac{\sqrt{3}}{2})(1+\frac{z^{-1}}{2+\sqrt{3}}). \tag{21}$$

It is easy to see that $\bar{W}_{12}(z)$ has a pole at $-1/(2+\sqrt{3})$, which is inside the unit circle. This is a contradiction to the requirement on the poles of a canonical spectral factor that they should all lie outside the unit circle.

[3] Solo in a private communication has indicated that an erroneous step in his proof leads to the discrepancy between his and our results.

4.2 Spectral characterization of noise-induced Granger causality

It is now straightforward to understand the effect of adding noise to the processes X_A, X_B on the property that X_A does not Granger cause X_B. Suppose as before that N_A, N_B are two processes, independent of X_A, X_B and each other, and added to X_A, X_B to yield new processes $\bar{X}_A = X_A + N_A$, $\bar{X}_B = X_B + N_B$. The outcome is that

$$\Phi_{\bar{A}\bar{B}} = \Phi_{AB}\,, \quad \Phi_{\bar{B}\bar{B}} = \Phi_{BB} + \Phi_{N_B N_B}\,. \tag{22}$$

The absence of Granger causality will carry over, *i.e.* $\bar{X}_A$ will not Granger cause $\bar{X}_B$ if and only if (by Theorem 3), $\Phi_{\bar{A}\bar{B}}\Phi_{\bar{B}\bar{B}}^{-1}$ is a stable transfer function. If there is noise on the process X_A but not the process X_B, the result is immediate that absence of causality continues to hold; the same transfer function fraction in fact arises, for $\Phi_{AB}\Phi_{BB}^{-1} = \Phi_{\bar{A}\bar{B}}\Phi_{\bar{B}\bar{B}}^{-1}$. On the other hand, if there is noise on the process X_B, for "almost all" spectra of $\Phi_{N_B N_B}$, including certainly a white spectrum, unless Φ_{BB} is itself white, the zeros of $\Phi_{BB} + \Phi_{N_B N_B}$ will differ from those of Φ_{BB} and not be the same as the poles of $\Phi_{\bar{A}\bar{B}} = \Phi_{AB}$. So *the cancellation of unstable pole-zero pairs in forming the fraction will no longer occur* and the absence of Granger causality will then be lost.

Now let us postulate that processes $\bar{X}_A, \bar{X}_B$ are measured and found to have the property that $\bar{X}_A$ does not Granger cause $\bar{X}_B$; these processes are assumed to be noisy versions of underlying processes X_A, X_B, with the additive noise processes being independent of each other and the underlying X_A, X_B processes. Ultimate interest lies in saying whether or not X_A Granger causes X_B. Then the above argument shows that if we knew that there was no noise perturbing X_B, processing of the noisy measurements would allow answering of the question. On the other hand, if there is noise perturbing X_B, one could not infer from the presence or absence of a causality property involving $\bar{X}_A, \bar{X}_B$ the corresponding property for X_A, X_B. The noise process N_B would need to have a specialized spectrum for absence of causality in the noisy case to imply it in the noiseless case. Note that there is no adjustment to the conclusions which arises in the special case of the noise process N_B being white.

The results above are summed up in the following theorem.

Theorem 5. *Adopt the same hypothesis as in Theorem 1. Let N_A, N_B be two stationary processes with rational spectra, with the same dimensions as X_A, X_B respectively, where X, N_A, N_B mutually independent, and set $\bar{X}_A = X_A + N_A$, $\bar{X}_B = X_B + N_B$.*

1. *If $N_B = 0$, then*

   X_A does not Granger cause X_B if and only if $\bar{X}_A$ does not Granger cause $\bar{X}_B$.

2. *If $N_B \neq 0$ and not all the unstable zeros of $\Phi_{BB} + \Phi_{N_B N_B}$ cancel (unstable) zeros of Φ_{AB}, we have the following implications:*
   (a) if X_A does not Granger cause X_B, then $\bar{X}_A$ Granger causes $\bar{X}_B$;
   (b) if $\bar{X}_A$ does not Granger cause $\bar{X}_B$, then X_A Granger causes X_B.

Remark If X_B is not noisy [$N_B = 0$], noise associated with the "causal variable" X_A cannot induce spurious Granger causality from X_A to X_B, despite possibly complicated dynamics on both X_A and X_B. Another special case of interest is provided by the situation where the two processes are actually independent. Then $\Phi_{AB} = 0$, and so the relevant transfer function $\Phi_{AB}\Phi_{BB}^{-1}$ with or without noise added remains zero and there is no causality introduced through the addition of noise.

We comment that our conclusions are at variance with those of Solo (2007), who asserts that addition of both noise sequences N_A, N_B to X_A, X_B where X_A does not Granger cause X_B means that $\bar{X}_A$ does not Granger cause $\bar{X}_B$. There appears to be an unjustified assumption in his work (as confirmed in private communication) where he constructs a triangular spectral factor for the $\bar{X}$ process but does not ensure that the off diagonal term is guaranteed to be stable–stability is simply assumed automatically. Such stability would be a necessary condition for asserting that $\bar{X}_A$ does not Granger cause $\bar{X}_B$.

5 Signal-to-noise ratio and Granger causality

In this section, we argue a form of continuity result. If there is additive noise perturbing an arrangement where there is absence of causality, then although generically absence of causality will be lost, we shall show that in a certain sense made more precise below, the introduced degree of non-causality is small. The practical effect of this result is that small amounts of noise in a particular situation may well be tolerable.

Our starting point is the following observation.

Lemma 2. *Consider a complex matrix function $M(z)$, analytic in $\rho < |z| < \rho^{-1}$, $0 < \rho < 1$ with $M(z) = M^\top(z^{-1})$, and positive definite on $|z| = 1$. Suppose*

$$M(z) = \sum_{i=-\infty}^{\infty} m_i z^i, \quad m_i = m_{-i}^\top \in \mathbb{R}^{d\times d} \tag{23}$$

and define the causal and anticausal parts by

$$M_+(z) = \frac{1}{2}m_0 + \sum_{i=1}^{\infty} m_i z^i \text{ and } M_-(z) = \frac{1}{2}m_0 + \sum_{i=-\infty}^{-1} m_i z^i . \tag{24}$$

Then the matrix function $L(z) := I + \varepsilon M(z)$ is analytic in $\rho < |z| < \rho^{-1}$, with $L(z) = L^\top(z^{-1})$, and positive definite on $|z| = 1$. Further to first order in $\varepsilon > 0$, there holds

$$L = I + \varepsilon M \approx (I + \varepsilon M_+)(I + \varepsilon M_-) \tag{25}$$

with $I + \varepsilon M_+$ stable and miniphase.

We remark that the terminology "to first order in ε" is shorthand for saying that the L_2 norm of the error between L above and the approximation of it on the right hand side of (25), call it $\Delta(z)$, is of order ε^2. The square of this L_2 norm can be computed with the aid of an integration of around the unit circle, as trace$\frac{1}{2\pi}\int[\Delta(\exp(j\omega))]^2 d\omega$ or by taking the squared sum of the coefficients in the Laurent series of the error, *i.e.* $\sum_{-\infty}^{\infty} \mathrm{tr}[\delta_i \delta_i^\top]$.

We will use this result to show that small perturbations in a spectrum give small perturbations in the associated spectral factors, and thence conclude that Granger causality is in a sense continuously dependent on the noise spectrum, it being absent when there is no noise. Accordingly we consider the arrangement studied in the previous section, with the introduction of a scaling parameter on the noise : thus $X = [X_A^\top \; X_B^\top]^\top$ and X_A does not Granger cause X_B nor does it cause X_B instantaneously. The canonical factor $W(z)$ for the noise-free spectrum $\Phi_{XX}(z)$ is upper block triangular and the innovations covariance matrix Q is block diagonal, and they obey the fundamental spectral factorization equation (4). Assume that $\varepsilon^{1/2} N_B$ for some $\varepsilon > 0$ is a noise process additively perturbing X_B, thus

$$\bar{X}_B = X_B + \varepsilon^{1/2} N_B, \quad \Phi_{\bar{B}\bar{B}} = \Phi_{BB} + \varepsilon \Phi_{N_B N_B} .$$

(We have effectively previously dealt with the effect of having a noise process N_A perturbing X_A–the noisy process $X_A + N_A$ is known to inherit the property of not Granger causing X_B, and so no further consideration is given to N_A and for convenience we take it as zero).

Now note that

$$\Phi_{\bar{X}\bar{X}} = \Phi_{XX} + \varepsilon \Phi_{NN} . \tag{26}$$

The spectrum $\Phi_{\bar{X}\bar{X}}$ gives rise to a canonical spectral factor, call it $\bar{W}(z)$ and an associated covariance matrix, call it $\bar{Q}$, satisfying

$$\Phi_{\bar{X}\bar{X}}(z) = \bar{W}(z)\bar{Q}\bar{W}^\top(z^{-1}) . \tag{27}$$

Our first result follows.

Theorem 6. *Adopt the same hypothesis as in Theorem 1 and let N_B be a stationary process with rational spectrum, with the same dimension as X_B , and with X, N_B independent. For fixed positive ε, define $\bar{X}_B = X_B + \varepsilon^{1/2} N_B$ so that $\Phi_{\bar{X}\bar{X}} = \Phi_{XX} + \varepsilon \Phi_{NN}$ where the $(1,1)$, $(1,2)$, $(2,1)$ blocks of Φ_{NN} are zero, and the $(2,2)$ block is $\Phi_{N_B N_B}$. Let $W(z), Q$ with $W(z)$ upper block triangular and Q block diagonal and $\bar{W}(z), \bar{Q}$ define canonical spectral factorizations of $\Phi_{XX}(z)$ and $\Phi_{\bar{X}\bar{X}}(z)$ as in* (4) *and* (27) *respectively. Then*

1. *$\bar{W}(z) - W(z)$ is $O(\varepsilon)$ on $|z| = 1$;*
2. *$\bar{Q} - Q$ is $O(\varepsilon)$;*
3. *$\Phi_{A\bar{B}}\Phi_{\bar{B}\bar{B}}^{-1} - \Phi_{AB}\Phi_{BB}^{-1}$ is $O(\varepsilon)$ on $|z| = 1$, and the anticausal part of $\Phi_{A\bar{B}}\Phi_{\bar{B}\bar{B}}^{-1}$ is $O(\varepsilon)$ on $|z| = 1$;*
4. *for suitably small ε, $\bar{W}_{22}(z)$ is minimum phase.*

We remark that the first and third bounds imply bounds on the L_2 norms of the quantities which are also $O(\varepsilon)$. Evidently, the $\bar{X}$ process is "close to" a process in which X_A does not cause $\bar{X}_B$ in two senses : the canonical spectral factor is close to upper block triangular with the innovations covariance matrix being block diagonal, and (separately), the anti-causal part of the two-sided Wiener filter associated with predicting X_A from $\bar{X}_B$ has small magnitude on $|z| = 1$ and in L_2 norm.

In the above theorem, we focused on the changes to transfer functions and to the innovations covariance caused by the introduction of noise. It is also relevant to compare the prediction error variances when $X_A(s), s \le t, X_B(s), s < t$ and $X_A(s), s \le t, X_{\bar{B}}(s), s < t$ are used to predict X_B and $\bar{X}_B$ respectively. The results are summarized in the following theorem. It shows that the prediction error "measure" of Granger causality is $O(\varepsilon^2)$.

Theorem 7. *Adopt the same hypothesis as in Theorem 6 and assume that $\varepsilon > 0$ is sufficiently small that $\bar{W}_{22}$ is minimum phase. Then there exist positive R, R' of $O(\varepsilon^2)$ for which there hold the upper and lower bounds:*

$$\begin{aligned} \mathrm{Var}\,[\bar{X}_B(t) - \mathsf{E}[\bar{X}_B(t)\,|\,\bar{X}_B(s) : s < t]] &\geq R + \bar{Q}_{22} - \bar{Q}_{12}^\top \bar{Q}_{11}^{-1} \bar{Q}_{12} \\ &= R + \mathrm{Var}\,[\bar{X}_B(t) - \mathsf{E}[\bar{X}_B(t)\,|\,X_A(t), X_A(s), \bar{X}_B(s) : s < t]] \end{aligned} \tag{28}$$

and

$$\mathrm{Var}\,[\bar{X}_B(t) - \mathsf{E}[\bar{X}_B(t)\,|\,\bar{X}_B(s) : s < t]] \leq (1+R')\mathrm{Var}\,[\bar{X}_B(t) - \mathsf{E}[\bar{X}_B(t)\,|\,X_A(t), X_A(s), \bar{X}_B(s) : s < t]]\,. \tag{29}$$

6 Effect of filtering on Granger causality

Consider a stationary full-rank process $X = [X_A^\top \; X_B^\top]^\top$. Instead of observing processes X_A, X_B, we observe the process

$$\begin{bmatrix} \bar{X}_A(t) \\ \bar{X}_B(t) \end{bmatrix} = T(L) \begin{bmatrix} X_A(t) \\ X_B(t) \end{bmatrix}, \quad T(L) := \begin{bmatrix} T_A(L) & T_{AB}(L) \\ T_{BA}(L) & T_B(L) \end{bmatrix}, \tag{30}$$

where $T(L)$ is a causal transfer function, partitioned conformably with $[X_A^\top \; X_B^\top]^\top$. We consider the question: if X_A does not Granger cause X_B, will $\bar{X}_A$ not Granger cause $\bar{X}_B$? Conversely, and on occasions more importantly, if one observes that $\bar{X}_A$ does not Granger cause $\bar{X}_B$, can one conclude that X_A does not Granger cause X_B? Questions of this type go back some time, see *e.g.* Pierce and Haugh (1977), Solo (2007, 2016), Florin, Gross, Pfeifer, Fink and Timmermann (2013), Barnett and Seth (2011), and Seth, Chorley and Barnett (2013).

In the following theorem, we give a general necessary and sufficient condition for $\bar{X}_A$ not to cause $\bar{X}_B$ in the sense of Granger.

Theorem 8. *Suppose Assumption 1 holds, and let $[\bar{X}_A^\top(t)\, \bar{X}_B^\top(t)]^\top$ be defined by* (30) *where $T(L)$ is a causal stable miniphase transfer function such that $T(0) = I_d$. Then, $\bar{X}_A$ does not Granger cause $\bar{X}_B$ if and only if*

$$\Pi_{21}(L)T^A(L) + \Pi_{22}(L)T^{BA}(L) = 0 \tag{31}$$

where

$$\Pi(L) = W(L)^{-1} = \begin{bmatrix} \Pi_{11}(L) & \Pi_{12}(L) \\ \Pi_{21}(L) & \Pi_{22}(L) \end{bmatrix} \text{ and } T(L)^{-1} = \begin{bmatrix} T^A(L) & T^{AB}(L) \\ T^{BA}(L) & T^B(L) \end{bmatrix} \tag{32}$$

are partitioned conformably with $[X_A^\top \; X_B^\top]^\top$.

Provided $\Pi_{22}(z)$ is miniphase, condition (31) shows it is generally possible to choose the filter $T(L)$ so that $\bar{X}_A$ does not Granger cause $\bar{X}_B$, irrespective of the causal relation between X_A and X_B. Filtering can make Granger causality invisible. In general, Granger causality from X_A to X_B is **not invariant** to the application of a multivariate causal miniphase filter to $X(t)$. When condition (31) holds, $\bar{X}_A$ does not Granger cause $\bar{X}_B$ even if X_A does Granger cause X_B. Conversely, if $\Pi_{12}(L) = 0$, so X_A does not Granger cause X_B, (31) does not hold if we have $\Pi_{22}(L)T^{BA}(L) \neq 0$: X_A does not Granger cause X_B, and $\bar{X}_A$ does Granger cause $\bar{X}_B$. For example, for a bivariate system ($d = 2$), the latter situation happens when

$$\Pi(L) = I_2 \quad \text{and} \quad T(L) = \begin{bmatrix} 1 & 0 \\ 0.5L & 1 \end{bmatrix}; \tag{33}$$

here $X(t) = \varepsilon(t)$ and

$$\bar{X}_A(t) = X_A(t) = \varepsilon_A(t)\,, \quad \bar{X}_B(t) = (0.5)X_A(t-1) + X_B(t) = (0.5)\bar{X}_A(t-1) + \varepsilon_B(t)\,. \tag{34}$$

Barnett and Seth (2011, Abstract, Sections 1 and 3) claim that "G-causality for a stationary vector autoregressive (VAR) process is fully invariant under the application of an arbitrary invertible filter". The above counterexample shows clearly this claim should be qualified. Conditions under which this type of invariance hold are given by Theorem 8.

In the following corollary, we consider the important case where $T(L)$ has a block triangular form.

Corollary 1. *Under the assumptions of Theorem 8, the following equivalences hold. If $T_{BA}(L) = 0$,*

$$X_A \text{ does not Granger cause } X_B \text{ if and only if } \bar{X}_A \text{ does not Granger cause } \bar{X}_B. \tag{35}$$

If $T_{AB}(L) = 0$,

$$\bar{X}_A \text{ does not Granger cause } \bar{X}_B \text{ if and only if } \Pi_{21}(L) = \Pi_{22}(L)T_B(L)^{-1}T_{BA}(L)\,. \tag{36}$$

Remark 1: (35) in Theorem 8 means that Granger causality from X_A to X_B is unaffected by linear causal filtering as long as $\bar{X}_B(t)$ only depends on current and past values of $X_B(t)$. The filtered series $\bar{X}_A(t)$ can involve lagged values of both $X_A(t)$ and $X_B(t)$. Further, the fact that (35) is an equivalence means that non-causality between the filtered series does allow one to conclude that the unfiltered series have the same property. Solo (2016) recently considered the special case where $T_{AB}(L) = 0$ and $T_{BA}(L) = 0$, and showed only sufficiency (X_A does not Granger cause X_B implies that $\bar{X}_A$ does not Granger cause $\bar{X}_B$); see also Florin et al. (2013), Barnett and Seth (2011) and Seth et al. (2013) for related results.
Remark 2: (36) of Theorem 8 gives a general condition under which Granger causality from X_A to X_B can be suppressed through filtering. In particular, if $\Pi_{21}(z)$ is miniphase, we can always choose $T(L)$ so that $\bar{X}_A$ does not Granger cause $\bar{X}_B$. For example, this is achieved by taking $T_B(L)$ invertible and

$$T_{BA}(L) = T_B(L)\Pi_{22}(L)^{-1}\Pi_{21}(L). \tag{37}$$

In particular, by taking $T_B(L) = I$ and $T_{BA}(L) = \Pi_{22}(L)^{-1}\Pi_{21}(L)$, we find that the filter

$$T(L) = \begin{bmatrix} I & 0 \\ \Pi_{22}(L)^{-1}\Pi_{21}(L) & I \end{bmatrix} \tag{38}$$

eliminates Granger causality from X_A to $\bar{X}_B$: $\bar{X}_B$ is "orthogonal" to X_A in the Granger sense.
Remark 3: The assumption that $W(L)$ is miniphase entails that both $W_A(L)$ and $W_B(L)$ are miniphase when $W_{BA}(L) = 0$, for then $\det[W(z)] = \det[W_A(z)]\det[W_B(z)]$. There is a heuristic explanation of why the introduction of a non miniphase but stable $W_B(z)$ might produce causality where previously it did not hold. Suppose X_A does not cause X_B, but X_B does cause X_A. For a specific example, suppose that $X_A(t) = X_B(t-2) + \varepsilon_A(t)$, $X_B(t) = \varepsilon_B(t)$, with $\varepsilon_A, \varepsilon_B$ independent white noise sequences. Now suppose that X_B is subject to processing by a filter which delays it. It is well understood in signal processing theory that any non-minimum phase filter introduces a delay; a particularly evident example is the filter with transfer function z^p, which produces a delay of p units. Suppose p is say 3. Then $X_A(t) = \bar{X}_B(t+1) + \varepsilon_A(t)$, so future values of $\bar{X}_B$ are correlated with past values of X_A, which means precisely that X_A does cause $\bar{X}_B$.
Remark 4:. Consider the case where X_A defines a filter which is simply a delay. Then the theorem says that if X_A does not Granger cause X_B, the same will be true if X_A is replaced by a delayed version of itself. This can of course be argued from first principles also.

7 Effect of subsampling on Granger causality

A further question raised in the work of Solo (2007) deals with subsampling. Suppose that a process X_A does not Granger cause X_B nor does it cause X_B instantaneously. Suppose the two processes are subsampled. Will the absence of Granger causality continue to hold for the subprocesses?

As we show here, again appealing to Theorem 3, absence of Granger causality may be lost, and Granger causality may arise in the subsampled processes.

By way of brief comment, we note a refinement of the question. Subsampling of two processes at the same rate may not occur synchronously. Thus, for example, samples of X_A with even time index might be considered along with samples of X_B with odd time index. Also, subsampling may occur at different rates. In fact, it could be that X_A is not subsampled, while X_B is. We will not consider these variants in this section.

We first note a very easily established theorem on the transformation of spectra under subsampling [see Hannan (1970)].[4]

Theorem 9. *Let $\Phi_{XX}(z)$ denote the spectrum of a process X, and suppose the process is sampled (synchronously in the case of a multivariate process) every M time units for some positive integer M, to generate a process $\bar{X}$, defined for time instants $\ldots, -M, 0, M, 2M, \ldots$. Then the spectrum of $\bar{X}$, defined using covariance data with lags which are integer multiples of M as*

$$\Phi_{\bar{X}\bar{X}}(z^M) = \sum_{k=-\infty}^{\infty} z^{Mk}\mathsf{E}[X(0)X^\top(-kM)] \tag{39}$$

is expressible as

$$\Phi_{\bar{X}\bar{X}}(z^M) = \frac{1}{M}\sum_{i=0}^{M-1}\Phi_{XX}(\omega^i z) \tag{40}$$

where $\omega = \exp(\sqrt{-1}2\pi/M)$.

[4]The relation between Granger causality and subsampling could also be studied by postulating a deeper structural model, such as a continuous-time model; for some earlier work on continuous-time Granger causality, see Florens and Fougère (1996) and Comte and Renault (1996). We focus here on the standard discrete-time second-order stationary model described in sections 1 - 2.

The proof is an immediate consequence of the fact that

$$\Phi_{XX}(\omega^i z) = \sum_{k=-\infty}^{\infty} (\omega^i z)^k \mathsf{E}[X(0)X^\top(-k)] \tag{41}$$

and the addition of these equations for $i = 0, \ldots, M-1$.

In our counterexample, we shall take $M = 2$. This means that $\Phi_{\bar{X}\bar{X}}(z) = \Phi_{XX}(z) + \Phi_{XX}(-z)$. In detail, consider a process X defined by a unit matrix for the innovations covariance, and a canonical spectral factor given by

$$W(z) = \begin{bmatrix} *** & \frac{z}{1-\frac{z}{2}} \\ 0 & \frac{1-\frac{z}{\sqrt{2}}}{1-\frac{z}{2}} \end{bmatrix} \tag{42}$$

with the particular expression for the $(1,1)$ entry not being provided, since it turns out to be irrelevant to the calculation. From this expression, it follows that the $(1,2)$ and $(2,2)$ entries of Φ_{XX} are :

$$\Phi_{AB}(z) = \Big(\frac{z}{1-\frac{z}{2}}\Big)\Big(\frac{1-\frac{z^{-1}}{\sqrt{2}}}{1-\frac{z^{-1}}{2}}\Big) = \frac{z-\frac{1}{\sqrt{2}}}{\frac{5}{4}-\frac{1}{2}(z+z^{-1})}\,, \tag{43}$$

$$\Phi_{BB}(z) = \Big(\frac{1-\frac{z}{\sqrt{2}}}{(1-\frac{z}{2})}\Big)\Big(\frac{1-\frac{z^{-1}}{\sqrt{2}}}{(1-\frac{z^{-1}}{2})}\Big) = \frac{\frac{3}{2}-\frac{1}{\sqrt{2}}(z+z^{-1})}{\frac{5}{4}-\frac{1}{2}(z+z^{-1})}\,. \tag{44}$$

Based on the formula for subsampled spectra, we now have:

$$\Phi_{\bar{A}\bar{B}}(z^2) = \frac{z-\frac{1}{\sqrt{2}}}{\frac{5}{4}-\frac{1}{2}(z+z^{-1})} + \frac{-z-\frac{1}{\sqrt{2}}}{\frac{5}{4}+\frac{1}{2}(z+z^{-1})} = \frac{(1-\frac{5}{2\sqrt{2}})+z^2}{[\frac{5}{4}-\frac{1}{2}(z+z^{-1})][\frac{5}{4}+\frac{1}{2}(z+z^{-1})]}\,, \tag{45}$$

$$\Phi_{\bar{B}\bar{B}}(z^2) = \frac{\frac{3}{2}-\frac{1}{\sqrt{2}}(z+z^{-1})}{\frac{5}{4}-\frac{1}{2}(z+z^{-1})} + \frac{\frac{3}{2}+\frac{1}{\sqrt{2}}(z+z^{-1})}{\frac{5}{4}+\frac{1}{2}(z+z^{-1})} = \frac{[\frac{15}{8}-\sqrt{2}-\frac{1}{\sqrt{2}}(z^2+z^{-2})]}{[\frac{5}{4}-\frac{1}{2}(z+z^{-1})][\frac{5}{4}+\frac{1}{2}(z+z^{-1})]}\,, \tag{46}$$

hence

$$\Phi_{\bar{A}\bar{B}}(z^2)\Phi_{\bar{B}\bar{B}}^{-1}(z^2) = \frac{(1-\frac{5}{2\sqrt{2}})+z^2}{\frac{15}{8}-\sqrt{2}-\frac{1}{\sqrt{2}}(z^2+z^{-2})} = \frac{z^4+(1-\frac{5}{2\sqrt{2}})z^2}{-\frac{1}{\sqrt{2}}z^4+(\frac{15}{8}-\sqrt{2})z^2-\frac{1}{\sqrt{2}}}\,. \tag{47}$$

The denominator polynomial has four zeros on the unit circle (at $\pm 0.81 \pm 0.58\sqrt{-1}$) and there are no cancellations with the zeros of the numerator polynomial (at $0, 0, \pm 0.876\sqrt{-1}$). (The presence of unit circle zeros rather than zeros inside the unit circle is not of particular note. One can also verify that if the $(2,2)$ term of the spectral factor in the example is changed to be $(1-\frac{z}{3})(1-\frac{z}{2})^{-1}$, then the denominator polynomial has two zeros inside and two outside the unit circle, at ± 1.382 and ± 0.869.) So in both cases the transfer function is not causal. *This means that the downsampling destroys the absence of Granger causality*.

This result is not consistent with the claim of Solo (2007). It would appear that in seeking to construct a canonical spectral factorization for the subsampled process, he forces triangularity (which is needed to conclude absence of Granger causality) but in the process cannot assure that the proposed spectral factor remains stable, *i.e.* the simultaneous requirements on the spectral factor for it to be triangular and stable cannot both be met. The fact that an underlying process may have unidirectional Granger causality, while a time aggregated version of it (a form of subsampling) has bidirectional Granger causality, has also been observed in Chambers and McCrorie (2004).

8 Conclusion

In this paper, we have sought to explain how intuition can be misleading in determining the effect of certain changes made to an underlying model on the Granger causality properties associated with that model. For the important case of additive noise, we have established general conditions under which additive noise does not affect non-causality properties, as well as conditions under which noise induces spurious Granger causality. It is clear from these that additive noise generally distorts causal relations, even though there are interesting cases where it does not. For example, in the case of measurement errors, if X_B is not noisy, noise on X_A does not induce spurious Granger causality from X_A to X_B. In order to derive our results, we mostly rely on generating function methods, canonical spectral factors, Wold decompositions and spectra. These approaches considerably simplify many proofs. Besides rigorous demonstrations, our findings extend and qualify the early results of Newbold (1978), which focused on special cases. Further, in our discussion of noise, we have drawn attention to the usefulness of thinking of an "amount" of causality : if noise is "low", the distortion to the predictability and Granger causal properties of the process is also "low".

We have also revisited the effects of linear transformations, filtering and subsampling on Granger causality properties. These results include general necessary and sufficient conditions under which "spurious" causal relations between time series are not induced by linear transformations of the variables involved. This characterization also yields linear transformations (or filters) which can eliminate Granger causality from one vector to another one. The various results presented allow one to correct some erroneous statements in the earlier literature [as in Solo (2007) and Barnett and Seth (2011)] on Granger causality in the presence of measurement errors and variable transformations.

The fact that additive noise can distort Granger causality relations means that empirical findings on such properties become more delicate to interpret. Indeed, inferences on "noisy variables" remain perfectly valid as long the causal properties are ascribed to observed variables. If we have no theory (such as an economic or a physical model) that can lead one to distinguish between a "true" latent variable and a "measured" variable, there is no error-in-variables problem. Otherwise, one must be aware that the causal structure of the original "uncontaminated" variables can be different. In this context, it is quite interesting to observe that no "spurious" causality can arise when noise only affects the output variable (see Theorem 5). Further, when the signal-to-noise ratio is large, distortions on apparent causality will also be small (see Theorem 6).

Similarly, the fact that aggregation and subsampling can distort dynamic relations has been widely discussed; see Tiao and Wei (1976), Wallis (1974), Sims (1974), Wei (1982), Hylleberg (1986), Marcellino (1999), Kaiser and Maravall (2001), Breitung and Swanson (2002), Gong et al. (2015), and the references in the survey of Silvestrini and Veredas (2008). For a specific example showing that causal relations are modified by changing observation frequency, see Dufour et al. (2012). However, Theorem 8 and Corollary 1 give general necessary and sufficient conditions under which "spurious" causal relations between (vector) time series will not be induced by linear transformations of the variables involved. This also yields linear transformations (or filters) which can eliminate Granger causality from one vector to another one.

Unaddressed to this point is a corresponding result on subsampling. It is well known that a continuous-time band-limited process and a sampled version of it with sampling frequency in excess of twice the approximate "cut-off" frequency of the continuous-time process have more or less the same information. This would suggest that Granger causality properties of continuous-time band-limited processes would be "almost" preserved if they were sampled frequently enough, *i.e.* in these circumstances the loss of Granger causality would be small. On a possible avenue in this direction, see Pollock (2012).

Due to dependence on an information set, Granger causality properties are not generally invariant to observation frequency. In contrast with contamination by "noise" – which is typically problematic – non-invariance to observation frequency may not be a "problem". Indeed, it can have considerable practical meaning and usefulness: different mechanisms may matter for short and long-run predictability as well as decisions, in the same way that different "laws" apply to micro-phenomena and macro-phenomena in physics. For economic and financial decisions, short-run and long-decisions may depend on different factors and require different rules. High-frequency decisions require prediction for data observed at high frequency, while longer-run decisions require predictability for data observed at low frequencies. Granger causality analysis at different observation frequencies can provide information on this. It would certainly be of interest to develop a systematic framework for exploring and exploiting such features. The use of mixed frequency data [as proposed in Ghysels et al. (2016)] could also be relevant in this context. But these problems clearly go beyond the scope of the present paper.

A different area worthy of examination is that of networked systems. The underlying structure could be viewed using a directed graph, An edge from vertex i to vertex j would mean that process i is caused by process j. One could begin by examining whether, in a path graph with vertex set $\{v_1, v_2, \ldots, v_N\}$ and edge set $\{e_{12}, e_{23}, \ldots, e_{(N-1)N}\}$, Granger causality properties of nonadjacent nodes could be predicted from the corresponding properties of adjacent nodes.

The results given in this paper rely on the common assumption of (second-order) stationarity. Even though there is no reason to think distortions induced by measurement errors would suddenly stop to exist in nonstationary setups, it would be of interest to extend our results to nonstationarity setups, *e.g.* through Hilbert space techniques, especially to evaluate whether non-stationarity tends to increase or decrease the distortions. Note that nonstationary models are explicitly allowed in Dufour and Renault (1998) and Dufour et al. (2006); for other papers which apply Hilbert space methods to Granger causality, see Hosoya (1977), Florens and Mouchart (1982, 1985), Florens, Mouchart and Rolin (1993), Florens and Fougère (1996), Triacca (1998, 2000), Al-Sadoon (2014). In order to study measurement errors, careful consideration of the type of nonstationarities is needed, and different proof methods may be required. Another possibly more general approach to incorporate non-stationarity would consist in working with the underlying probability measures, along the lines suggested by Mykland (1986). We leave such extensions to future work.

A Appendix

Proof of Lemma 1 Notice first that

$$\begin{aligned}\mathsf{E}[Z(t)\,|\,X(s),Y(s):s<t] &= \mathsf{E}[X(t)+Y(t)\,|\,X(s),Y(s):s<t]\\ &= \mathsf{E}[X(t)\,|\,X(s),Y(s):s<t]+\mathsf{E}[Y(t)\,|\,X(s),Y(s):s<t]\\ &= \mathsf{E}[X(t)\,|\,X(s):s<t]+\mathsf{E}[Y(t)\,|\,Y(s):s<t] \qquad (48)\end{aligned}$$

Accordingly, using the fact that the space spanned by $Z(s)=X(s)+Y(s)$ for $s<t$ is a subspace of the space spanned by $X(s),Y(s)$ for $s<t$, we have

$$\begin{aligned}\mathrm{Var}[Z(t)\,|\,Z(s):s<t] &\geq \mathrm{Var}[X(t)+Y(t)\,|\,X(s),Y(s):s<t]\\ &= \mathrm{Var}[X(t)\,|\,X(s):s<t]+\mathrm{Var}[Y(t)\,|\,Y(s):s<t] \qquad (49)\end{aligned}$$

as claimed. Of course the independence of the processes X,Y is used at the second last step.

Proof of Theorem 3 First, consider the question of absence of Granger causality and instantaneous causality and assume that condition 2 of Theorem 1 holds. Using (4), (10) and the fact that $W_{21}=0$, it is easily checked that

$$\Phi_{AB}\Phi_{BB}^{-1}=W_{12}W_{22}^{-1}\,. \qquad (50)$$

Also, the triangular structure of W and its minimum phase character imply that W_{22}^{-1} is stable. Hence $\Phi_{AB}\Phi_{BB}^{-1}$ is also stable. It also assumes the value 0 at $z=0$ because of the normalization condition $W(0)=I$.

To prove the converse, consider a canonical spectral factorization of the type of (4) (with no a priori restriction on the block triangular structure of W). Suppose that $\Phi_{AB}\Phi_{BB}^{-1}:=T(z)$ is stable and zero at $z=0$. We first assert that this will imply the condition $W_{21}=0$ in the minimum phase, stable spectral factor normalized to be I at $z=0$. To see this, suppose first that $V_B(z)$ is a minimum phase, stable transfer function with $V_B(\infty)=I$ and R_B is a positive definite matrix such that

$$\Phi_{BB}(z)=V_B(z)R_BV_B^{\top}(z^{-1})\,. \qquad (51)$$

Define

$$V_{AB}(z)=\Phi_{AB}(z)\Phi_{BB}^{-1}(z)V_B(z)\,. \qquad (52)$$

Note that $V_{AB}(z)$ is then stable and zero at $z=0$. Note also that the last two equations yield

$$\Phi_{AB}(z)=V_{AB}(z)R_BV_B^{\top}(z^{-1})\,. \qquad (53)$$

By the fact that Φ_{XX} is positive definite for all $z=e^{j\omega}$, it follows that $\Phi_{AA}-\Phi_{AB}\Phi_{BB}^{-1}\Phi_{BA}$ is also positive definite for all $z=e^{j\omega}$, and accordingly, we can define a minimum phase, stable transfer function $V_A(z)$ with $V_A(0)=I$ and a positive definite R_A such that

$$\Phi_{AA}(z)-\Phi_{AB}(z)\Phi_{BB}^{-1}(z)\Phi_{BA}(z)=V_A(z)R_AV_A^{\top}(z^{-1})\,. \qquad (54)$$

Observe that

$$\Phi_{AB}(z)\Phi_{BB}^{-1}(z)\Phi_{BA}(z)=\Phi_{AB}(z)\Phi_{BB}^{-1}(z)[V_B(z)R_BV_B^{\top}(z^{-1})]\Phi_{BB}^{-1}(z)\Phi_{BA}(z)=V_{AB}(z)R_BV_{AB}^{\top}(z^{-1}) \qquad (55)$$

or equivalently

$$\Phi_{AA}(z)=V_{AB}(z)R_BV_{AB}^{\top}(z^{-1})+V_A(z)R_AV_A^{\top}(z^{-1})\,. \qquad (56)$$

Now from equations (56), (53) and (51), we have

$$\begin{bmatrix}\Phi_{AA}(z) & \Phi_{AB}(z)\\ \Phi_{BA}(z) & \Phi_{BB}(z)\end{bmatrix}=\begin{bmatrix}V_A(z) & V_{AB}(z)\\ 0 & V_B(z)\end{bmatrix}\begin{bmatrix}R_A & 0\\ 0 & R_B\end{bmatrix}\begin{bmatrix}V_A^{\top}(z^{-1}) & 0\\ V_{AB}^{\top}(z^{-1}) & V_B^{\top}(z^{-1})\end{bmatrix}. \qquad (57)$$

Now make the definitions

$$V(z)=\begin{bmatrix}V_{AA}(z) & V_{AB}(z)\\ 0 & V_{BB}(z)\end{bmatrix},\ R=\begin{bmatrix}R_A & 0\\ 0 & R_B\end{bmatrix}, \qquad (58)$$

and observe that $V(z)$ is stable, minimum phase and has $V(0)=I$. Accordingly, $V(z)$ is the unique miniphase stable spectral factor with $V(0)=I$, and condition 2 of Theorem 1 is evidently fulfilled.

The argument for Granger causality only is similar but not identical, and so we present the details. Assume then there is an absence of Granger causality, which means that there is a spectral factorization with canonical spectral factor $W(z)$ and not necessarily block diagonal innovations covariance matrix of the form

$$\Phi_{XX}(z) = \begin{bmatrix} W_{11}(z) & W_{12}(z) \\ 0 & W_{22}(z) \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^\top & Q_{22} \end{bmatrix} \begin{bmatrix} W_{11}^\top(z^{-1}) & 0 \\ W_{12}^\top(z^{-1}) & W_{22}^\top(z^{-1}) \end{bmatrix}. \tag{59}$$

It then follows that

$$\Phi_{AB}\Phi_{BB}^{-1} = (W_{11}Q_{12} + W_{12}Q_{22})W_{22}^{-1} \tag{60}$$

and this is evidently stable.

For the converse, we follow the proof applying for the case where the aim was to conclude an absence of Granger causality and instantaneous causality. The proof applies with the first change that $V_{AB}(z)$ is no longer guaranteed to be zero at $z = 0$. Equation (57) holds, including the fact that R is block diagonal, but $V(\infty) \neq I$, due to the generally nonzero nature of $V_{AB}(0)$. Now by setting

$$W(z) = \begin{bmatrix} I & -V_{AB}(0) \\ 0 & I \end{bmatrix} V(z) \tag{61}$$

equation (59) arises, and Granger causality is then proved.

Proof of Theorem 4 For the first claim, let $W_{11}(z), W_{12}(z), W_{22}(z), Q_{11}, Q_{12}$ and Q_{22} be the matrices of the canonical spectral factor description of the joint spectrum of X_A, X_B. If $W_{12} = 0$, then it is evident that the two processes are independent. So we must prove that if $W_{12} \neq 0$, then X_B Granger causes X_A. This is equivalent to showing that

$$\mathrm{Var}[X_A(t) \,|\, X_A(s), X_B(s) : s < t] \leq \mathrm{Var}[X_A(t) \,|\, X_A(s) : s < t] \tag{62}$$

and the two conditional covariance matrices are unequal. Now observe that from the canonical factorization of the joint process X_A, X_B, we have immediately

$$\mathrm{Var}[X_A(t) \,|\, X_B(t), X_A(s), X_B(s) : s < t] = Q_{11}\,. \tag{63}$$

Further, we know that $X_A(t) = W_{11}(L)\varepsilon_A(t) + W_{12}(L)\varepsilon_B(t)$. This expresses X_A as a sum of two independent processes and the lemma above applies. Note that the variance of the one step ahead prediction for the process $W_{11}(L)\varepsilon_A(t)$ is precisely Q_1, since $W_{11}(z)$ is a canonical spectral factor. Hence the claim (62) is established.
For the second part, we are required to show that when X_A does not cause X_B, either the processes are independent or

$$\mathrm{Var}[X_A(t) \,|\, X_B(t), X_A(s), X_B(s) : s < t] \leq \mathrm{Var}[X_A(t) \,|\, X_A(s) : s < t] \tag{64}$$

where equality does not hold. Let Q_{12} be the $(1,2)$ block of the matrix Q in the canonical spectral factorization; in general, it is not zero. Then it is easily seen that

$$\mathrm{Var}[X_A(t) \,|\, X_B(t), X_A(s), X_B(s) : s < t] = Q_{11} - Q_{12}Q_{22}^{-1}Q_{12}^\top\,. \tag{65}$$

Next observe that

$$X_A(t) = W_{11}(L)[\varepsilon_A(t) - Q_{12}Q_{22}^{-1}\varepsilon_B(t)] + [W_{11}(L)Q_{12}Q_{22}^{-1} + W_{12}(L)]\varepsilon_B(t)\,. \tag{66}$$

This expresses X_A as the sum of two independent processes, since $\varepsilon_A(t) - Q_{12}Q_{22}^{-1}\varepsilon_B(t)$ and $\varepsilon_B(t)$ are independent. In the additive decomposition, the second process will evanesce if and only if $W_{12}(z) = 0, Q_{12} = 0$, *i.e.* the two processes X_A, X_B are independent. The first process is in fact the process $X_A(t) - \mathsf{E}[X_A(t) \,|\, X_B(t), X_A(s), X_B(s) : s < t]$, as is easily checked. Hence once again the lemma applies to yield the result (64) as required.

Proof of Theorem 6 The fact that $\Phi_{\tilde{X}\tilde{X}}$ satisfies (26) implies that

$$\Phi_{\tilde{X}\tilde{X}} = WQ^{1/2}[I + \varepsilon Q^{-1/2}W^{-1}\Phi_{NN}W^{-*}Q^{-1/2}]Q^{1/2}W^*\,. \tag{67}$$

Now identify M in Lemma 2 with $Q^{-1/2}W^{-1}\Phi_{NN}W^{-*}Q^{-1/2}$ to conclude that a stable and miniphase spectral factor of the noise perturbed spectrum $\Phi_{\tilde{X}\tilde{X}}$ is (to first order in ε, corresponding to a low noise situation),

$$\tilde{W} = WQ^{1/2}[I + \varepsilon(Q^{-1/2}W^{-1}\Phi_{NN}W^{-*}Q^{-1/2})_+] \tag{68}$$

in the sense that

$$\Phi_{\tilde{X}\tilde{X}}(z) = \tilde{W}(z)\tilde{W}^\top(z^{-1})\,. \tag{69}$$

Note that $\tilde{W}(z)$ is not canonical because the requirement $\tilde{W}(0)=I$ is not fulfilled. Define J to be the value of

$$[I+\varepsilon(Q^{-1/2}W^{-1}\Phi_{NN}W^{-*}Q^{-1/2})_{+}] \tag{70}$$

at $z=0$. Note that this is precisely $I+\varepsilon K$, where $2K$ is necessarily nonnegative being $(1/2\pi)$ times the integral around the unit circle of $Q^{-1/2}W^{-1}\Phi_{NN}W^{-*}Q^{-1/2}$ and note that $J-I$ (which is εK) and thus $J^{-1}-I$ are $O(\varepsilon)$. Then the canonical spectral factor will be

$$\bar{W}(z)=\tilde{W}(z)J^{-1}Q^{-1/2} \tag{71}$$

since this assures that (27) will hold where we can identify

$$\bar{Q}=\begin{bmatrix} \bar{Q}_{11} & \bar{Q}_{12} \\ \bar{Q}_{12}^{\top} & \bar{Q}_{22} \end{bmatrix}=Q^{1/2}J^{2}Q^{1/2}. \tag{72}$$

From the calculations immediately preceding the last two equations, it is easy to see that the difference $\bar{W}(z)-W(z)$ will be $O(\varepsilon)$, as will $\bar{Q}_{12}$ and $Q-\bar{Q}$.
Next we consider the claim concerning the two-sided Wiener filter for estimating X_A from $\bar{X}_B$. It is easily seen using (67) that $\Phi_{BA}(z)\Phi_{BB}^{-1}(z)$ differs for each fixed $z=\exp(j\omega)$ from $\Phi_{\bar{B}A}(z)\Phi_{\bar{B}\bar{B}}^{-1}(z)$ by an amount bounded by $O(\varepsilon)$ as ε goes to zero, and hence the anticausal part will have the same property. Since the anticausal part of $\Phi_{BA}(z)\Phi_{BB}^{-1}(z)$ is actually zero, this means the anticausal part of $\Phi_{BA}(z)\Phi_{BB}^{-1}(z)$ will have an L_2 norm that is $O(\varepsilon)$.
To prove the final claim, observe that because $W(z)$ is block upper triangular and canonical, the submatrix $W_{22}(z)$ is minimum phase. Now the right side of $W_{22}^{-1}(z)\bar{W}_{22}(z)=I+W_{22}^{-1}(z)[\bar{W}_{22}(z)-W_{22}(z)]$ represents a perturbation of I by a stable matrix whose norm is bounded by $O(\varepsilon)$ on the unit circle, and accordingly it is minimum phase. Hence the product with $W_{22}(z)$ is also minimum phase, *i.e.* $\bar{W}_{22}(z)$ is minimum phase. This completes the proof of the theorem.

Proof of Theorem 7 We start by expressing $\bar{X}_B$ as a sum of two independent processes. Thus observe that we can express $\bar{X}_B$ as

$$\bar{X}_B(t)=[\bar{W}_{22}(L)\bar{Q}_{12}^{\top}\bar{Q}_{11}^{-1}+\bar{W}_{21}(L)]\bar{\varepsilon}_A(t)+\bar{W}_{22}(L)[\bar{\varepsilon}_B(t)-\bar{Q}_{12}^{\top}\bar{Q}_{11}^{-1}\bar{\varepsilon}_A(t)] \tag{73}$$

in obvious notation. Note that the two processes $\bar{\varepsilon}_A(t)$ and $\tilde{\varepsilon}_B(t)=\bar{\varepsilon}_B(t)-\bar{Q}_{12}^{\top}\bar{Q}_{11}^{-1}\bar{\varepsilon}_A(t)$ are orthogonal, with covariance matrix

$$\mathsf{E}\{\begin{bmatrix} \bar{\varepsilon}_A(t) \\ \tilde{\varepsilon}_B(t) \end{bmatrix}\begin{bmatrix}\bar{\varepsilon}_A^{\top}(t) \; \tilde{\varepsilon}_B^{\top}(t)\end{bmatrix}\}=\begin{bmatrix} \bar{Q}_{11} & 0 \\ 0 & \bar{Q}_{22}-\bar{Q}_{12}^{\top}\bar{Q}_{11}^{-1}\bar{Q}_{12} \end{bmatrix}. \tag{74}$$

Let R denotes the error variance in estimating the value at time t of $[\bar{W}_{22}(L)\bar{Q}_{12}^{\top}\bar{Q}_{11}^{-1}+\bar{W}_{21}(L)]\bar{\varepsilon}_A(t)$ from values for $s<t$. Recall also that $\bar{W}_{22}(z)$ is minimum phase, and must satisfy $\bar{W}_{22}(0)=I$. Accordingly, it is a canonical spectral factor of the spectrum of $\bar{W}_{22}(L)[\bar{\varepsilon}_B(t)-\bar{W}_{12}^{\top}\bar{Q}_{11}^{-1}\bar{\varepsilon}_A(t)]$. Hence the prediction error covariance associated with estimating the value at time t of $\bar{W}_{22}(L)\tilde{\varepsilon}_B(t)$ from values for $s<t$ is precisely $\bar{Q}_{22}-\bar{Q}_{12}^{\top}\bar{Q}_{11}^{-1}\bar{Q}_{12}$. We can now use Lemma 1 as we did in the proof of Theorem 4 to conclude that

$$\mathrm{Var}\big[\bar{X}_B(t)-\mathsf{E}[\bar{X}_B(t)\,|\,\bar{X}_B(s):s<t]\big]\geq R+\bar{Q}_{22}-\bar{Q}_{12}^{\top}\bar{Q}_{11}^{-1}\bar{Q}_{12}\,. \tag{75}$$

From the canonical factorization for $\Phi_{\bar{X}\bar{X}}$ we have that the second term on the right in the above equation is given by

$$\mathrm{Var}\big[\bar{X}_B(t)-\mathsf{E}[\bar{X}_B(t)\,|\,X_A(t),X_A(s),\bar{X}_B(s):s<t\big]=\bar{Q}_{22}-\bar{Q}_{12}^{\top}\bar{Q}_{11}^{-1}\bar{Q}_{12}\,. \tag{76}$$

We now turn to establishing the bound on R. $\bar{X}_B(t)$ from $\bar{X}_A(s),s\leq t,\bar{X}_B(s),s<t$. This is, as is well known, precisely $\bar{Q}_{22}-\bar{Q}_{12}^{\top}\bar{Q}_{11}^{-1}\bar{Q}_{12}$. That for predicting $\bar{X}_B(t)$ from $\bar{X}_B(s)$ is more complicated. To make progress observe that Consider the transfer function acting on $\bar{\varepsilon}_A$ in (73). From Theorem 6, $\bar{Q}_{12}$ and $\bar{W}_{21}(z)$ are $O(\varepsilon)$ and so the transfer function multiplying $\bar{\varepsilon}_A$ is of order ε. Hence we see that R, being the prediction error variance in estimating a variable whose spectrum is proportional to ε^2 must itself be proportional to ε^2. Hence the increase in prediction error covariance when $\bar{X}_A$ ceases to be available for estimating $\bar{X}_B$ is *bounded from below* by a quantity proportional to ε^2.
We now derive the overbound of the same order. Choose a constant R' so that for all $z=\exp(j\omega)$, there holds

$$R'\bar{W}_{22}(z)[\bar{Q}_{22}-\bar{Q}_{12}^{\top}\bar{Q}_{11}^{-1}\bar{Q}_{12}]\bar{W}_{22}^{\top}(z^{-1})\geq[\bar{W}_{22}(z)\bar{Q}_{12}^{\top}\bar{Q}_{11}^{-1}+\bar{W}_{21}(z)]\bar{Q}_{11}[\bar{Q}_{11}^{-1}\bar{Q}_{12}\bar{W}_{22}^{\top}(z^{-1})+\bar{W}_{21}^{\top}(z^{-1})]\,. \tag{77}$$

This is possible since for all $z=\exp(j\omega)$ the left side is positive definite. Since the right side is $O(\varepsilon^2)$, it is clear that R' can be taken also as $O(\varepsilon^2)$. Now $\bar{X}_B$ is written in (73) as the sum of two orthogonal processes. Hence the spectrum of $\bar{X}_B$ will be the sum of the spectra of these two processes, *i.e.*

$$\bar{W}_{22}(z)[\bar{Q}_{22}-\bar{Q}_{12}^{\top}\bar{Q}_{11}^{-1}\bar{Q}_{12}]\bar{W}_{22}^{\top}(z^{-1})+[\bar{W}_{22}(z)\bar{Q}_{12}^{\top}\bar{Q}_{11}^{-1}+\bar{W}_{21}(z)]\bar{Q}_{11}[\bar{Q}_{11}^{-1}\bar{Q}_{12}\bar{W}_{22}^{\top}(z^{-1})+\bar{W}_{21}^{\top}(z^{-1})]$$

for $z = \exp(j\omega)$ and this is overbounded by a spectrum, call it $\Phi_{CC}(z)$, with

$$\Phi_{CC}(z) = (R'+1)\bar{W}_{22}(z)[\bar{Q}_{22} - \bar{Q}_{12}^\top \bar{Q}_{11}^{-1}\bar{Q}_{12}]\bar{W}_{22}^\top(z^{-1}) \tag{78}$$

again for $z = \exp(j\omega)$. Hence there exists a process, call it X_D, which is independent of $\bar{X}_B$, for which $X_C = \bar{X}_B + X_D$ and whose spectrum is $\Phi_{CC}(z) - \Phi_{\bar{X}\bar{X}}(z)$. By Lemma 1, the variance of the one step prediction estimate using its own past of the process X_C with spectrum of Φ_{CC} overbounds the sum of the variances of the one step prediction estimates of each of $\bar{X}_B$ and X_D. *A fortiori* it overbounds the variance of the one step prediction estimate of $\bar{X}_B$: thus

$$(R'+1)[\bar{Q}_{22} - \bar{Q}_{12}^\top \bar{Q}_{11}^{-1}\bar{Q}_{12}] \geq \mathrm{Var}\,[\bar{X}_B(t) - \mathsf{E}[\bar{X}_B(t)\,|\,\bar{X}_B(s) : s < t]]\,. \tag{79}$$

Equivalently, we can write

$$\mathrm{Var}\,[\bar{X}_B(t) - \mathsf{E}[\bar{X}_B(t)\,|\,\bar{X}_B(s) : s < t]] \leq (R'+1)\mathrm{Var}\,[\bar{X}_B(t) - \mathsf{E}[\bar{X}_B(t)\,|\,X_A(t), X_A(s), \bar{X}_B(s) : s < t]]\,. \tag{80}$$

This completes the proof of the theorem.

Proof of Theorem 8 Since the process has an autoregressive representation, we can write :

$$\Pi(L)X(t) = \Pi(L)T(L)^{-1}T(L)X(t) = \bar{\Pi}(L)\bar{X}(t) = \varepsilon(t) \tag{81}$$

where

$$\bar{\Pi}(L) = \begin{bmatrix} \bar{\Pi}_{11}(L) & \bar{\Pi}_{12}(L) \\ \bar{\Pi}_{21}(L) & \bar{\Pi}_{22}(L) \end{bmatrix} = \begin{bmatrix} \Pi_{11}(L) & \Pi_{12}(L) \\ \Pi_{21}(L) & \Pi_{22}(L) \end{bmatrix} \begin{bmatrix} T^A(L) & T^{AB}(L) \\ T^{BA}(L) & T^B(L) \end{bmatrix}, \quad \varepsilon(t) = \begin{bmatrix} \varepsilon_A(t) \\ \varepsilon_B(t) \end{bmatrix}. \tag{82}$$

In particular,

$$\bar{\Pi}_{21}(L) = \Pi_{21}(L)T^A(L) + \Pi_{22}(L)T^{BA}(L)\,. \tag{83}$$

Since $\varepsilon(t)$ is uncorrelated with the past $X(t)$, hence also of the past of $\bar{X}(t)$, it follows from Proposition 1 in Boudjellaba et al. (1992) that : $\bar{X}_A$ does not Granger cause $\bar{X}_B$ if and only if $\bar{\Pi}_{21}(L) = 0$.

Proof of Corollary 1 If $T_{BA}(L) = 0$, it follows from standard results on the inversion of partitioned matrices that

$$T(L)^{-1} = \begin{bmatrix} T_A(L) & T_{AB}(L) \\ 0 & T_B(L) \end{bmatrix}^{-1} = \begin{bmatrix} T_A(L)^{-1} & -T_A(L)^{-1}T_{AB}(L)T_B(L)^{-1} \\ 0 & T_B(L)^{-1} \end{bmatrix} = \begin{bmatrix} T^A(L) & T^{AB}(L) \\ T^{BA}(L) & T^B(L) \end{bmatrix}; \tag{84}$$

see Harville (2008, Chapter 8, Theorem 8.5.4). Then, the condition takes the form $\Pi_{21}(L)T_A(L)^{-1} = 0$, which in view of the invertibility of $T_A(L)$ is equivalent to $\Pi_{21}(L) = 0$. By Proposition 1 in Boudjellaba et al. (1992), X_A does not Granger cause X_B if and only if $\Pi_{21}(L) = 0$, so that

$$\bar{X}_A \text{ does not Granger cause } \bar{X}_B \text{ if and only if } X_A \text{ does not Granger cause } X_B\,. \tag{85}$$

If $T_{AB}(L) = 0$, we have

$$T(L)^{-1} = \begin{bmatrix} T_A(L) & 0 \\ T_{BA}(L) & T_B(L) \end{bmatrix}^{-1} = \begin{bmatrix} T_A(L)^{-1} & 0 \\ -T_B(L)^{-1}T_{BA}(L)T_A(L)^{-1} & T_B(L)^{-1} \end{bmatrix} = \begin{bmatrix} T^A(L) & T^{AB}(L) \\ T^{BA}(L) & T^B(L) \end{bmatrix} \tag{86}$$

so that $T^A(L) = T_A(L)^{-1}$ and $T^{BA}(L) = -T_B(L)^{-1}T_{BA}(L)T_A(L)^{-1}$, and condition (31) takes the form

$$\Pi_{21}(L)T^A(L) + \Pi_{22}(L)T^{BA}(L) = \Pi_{21}(L)T_A(L)^{-1} - \Pi_{22}(L)T_B(L)^{-1}T_{BA}(L)T_A(L)^{-1} = 0 \tag{87}$$

or, equivalently,

$$\Pi_{21}(L) = \Pi_{22}(L)T_B(L)^{-1}T_{BA}(L)\,. \tag{88}$$

Thus, $\bar{X}_A$ does not Granger cause $\bar{X}_B$ if and only if $\Pi_{21}(L) = \Pi_{22}(L)T_B(L)^{-1}T_{BA}(L)$.

References

Al-Sadoon, M. M. (2014), 'Geometric and long run aspects of Granger causality', *Journal of Econometrics* **178**, 558–568.

Anderson, B. D. O. (1985), 'Identification of scalar errors-in-variables models with dynamics', *Automatica* **21**(6), 709–716.

Anderson, B. D. O. and Deistler, M. (1984), 'Identifiability in dynamic errors-in-variables models', *Journal of Time Series Analysis* **5**, 1–13.

Barnett, L. C. and Seth, A. K. (2011), 'Behaviour of Granger causality under filtering: Theoretical invariance and practical application', *Journal of Neuroscience Methods* **211**, 404–419.

Barnett, L. C. and Seth, A. K. (2015), 'Granger causality for state-space models', *Physical Review E* **91**(4), 040101(5).

Barnett, L. C. and Seth, A. K. (2017), 'Detectability of Granger causality for subsampled continuous-time neurophysiological processes', *Journal of Neuroscience Methods* **275**, 93–121.

Boudjellaba, H., Dufour, J.-M. and Roy, R. (1992), 'Testing causality between two vectors in multivariate ARMA models', *Journal of the American Statistical Association* **87**(420), 1082–1090.

Breitung, J. and Swanson, N. R. (2002), 'Temporal aggregation and spurious instantaneous causality in multiple time series models', *Journal of Time Series Analysis* **23**, 651–665.

Buonaccorsi, J. P. (2010), *Measurement Error: Models, Methods, and Applications*, Chapman & Hall/CRC, Boca Raton, Florida.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006), *Measurement Error in Nonlinear Models: A Modern Perspective*, Chapman & Hall/CRC, Boca Raton, Florida.

Chambers, M. J. and McCrorie, R. (2004), Granger causality and the sampling of economic processes, Technical report, CentER Discussion Paper.

Comte, F. and Renault, E. (1996), 'Noncausality in continuous time models', *Econometric Theory* **12**(2), 215–256.

Deistler, M. and Anderson, B. D. O. (1989), 'Linear dynamic errors-in-variables models, some structure theory', *Journal of Econometrics* **41**, 39–63.

Dufour, J.-M., Garcia, R. and Taamouti, A. (2012), 'Measuring high-frequency causality between returns, realized volatility, and implied volatility', *Journal of Financial Econometrics* **10**(1), 1254–163.

Dufour, J.-M., Pelletier, D. and Renault, É. (2006), 'Short run and long run causality in time series: Inference', *Journal of Econometrics* **132**(2), 337–362.

Dufour, J.-M. and Renault, É. (1998), 'Short-run and long-run causality in time series: Theory', *Econometrica* **66**, 1099–1125.

Dufour, J.-M. and Taamouti, A. (2010), 'Short and long run causality measures: Theory and inference', *Journal of Econometrics* **154**, 42–58.

Dufour, J.-M. and Tessier, D. (1993), 'On the relationship between impulse response analysis, innovation accounting and Granger causality', *Economics Letters* **42**, 327–333.

Florens, J.-P. and Fougère, D. (1996), 'Noncausality in continuous time', *Econometrica* **64**(5), 1195–1212.

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

Florens, J.-P. and Mouchart, M. (1985), 'A linear theory for non-causality', *Econometrica* **53**, 157–175.

Florens, J.-P., Mouchart, M. and Rolin, J.-M. (1993), 'Noncausality and marginalization of Markov processes', *Econometric Theory* **9**(2), 241–262.

Florin, E., Gross, J., Pfeifer, J., Fink, G. R. and Timmermann, L. (2013), 'The effect of filtering on Granger causality based multivariate causality measures', *NeuroImage* **50**, 577–588.

Fuller, W. A. (1987), *Measurement Error Models*, John Wiley and Sons, New York.

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

Geweke, J. (1984*a*), Inference and causality in economic time series, *in* Z. Griliches and M. Intriligator, eds, 'Handbook of Econometrics', Vol. 2, North Holland, Amsterdam, Netherlands, pp. 1101–1144.

Geweke, J. (1984*b*), 'Measures of conditional linear dependence and feedback between time series', *Journal of the American Statistical Association* **79**(388), 907–915.

Ghysels, E., Hill, J. B. and Motegi, K. (2016), 'Testing for Granger causality with mixed frequency data', *Journal of Econometrics* **192**, 207–230.

Gong, M., Zhang, K., Schlökopf, B., Tao, D. and Geiger, P. (2015), Discovering temporal causal relations from subsampled data, *in* 'Proceedings of the 32*nd* International Conference on Machine Learning', Lille, France.

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

Granger, C. W. J. (1980), 'Testing for causality: A personal viewpoint', *Journal of Economic Dynamics and Control* **2**, 329–352.

Granger, C. W. J. (1988), 'Some recent developments in a concept of causality', *Journal of Economic Dynamics and Control* **4**, 243–259.

Gustafson, P. (2003), *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian adjustments*, Chapman & Hall/CRC, Boca Raton, Florida.

Hannan, E. J. (1970), *Multiple Time Series*, John Wiley & Sons, New York.

Hannan, E. J. and Deistler, M. (2012), *Statistical Theory of Linear Systems*, SIAM, Philadelphia, USA.

Harville, D. A. (2008), *Matrix Algebra from a Statistician's Perspective*, Springer Science, New York.

Hosoya, Y. (1977), 'On the Granger condition for non-causality', *Econometrica* **45**, 1735–1736.

Hsiao, C. (1982), 'Autoregressive modelling and causal ordering of economic variables', *Journal of Economic Dynamics and Control* **4**, 243–259.

Hylleberg, S. (1986), *Seasonality in Regression*, Academic Press, Orlando, Florida.

Kaiser, R. and Maravall, A. (2001), *Measuring Business Cycles in Economic Time Series*, number 154 *in* 'Lecture Notes in Statistics', Springer, Berlin.

Lütktepohl, H. (1982), 'Non-causality due to omitted variables', *Journal of Econometrics* **19**, 367–378.

Maravall, A. (1979), *Identification in Dynamic Shock-Error Models*, number 165 *in* 'Lecture Notes in Economics and Lecture Notes in Economics and Mathematical Systems', Springer, Berlin.

Marcellino, M. (1999), 'Some consequences of temporal aggregation in empirical analysis', *Journal of Business & Economic Statistics* **17**(1), 129–136.

McCrorie, J. R. and Chambers, M. J. (2006), 'Granger causality and the sampling of economic processes', *Journal of Econometrics* **132**, 311–336.

Mykland, P. (1986), Statistical causality, Technical Report 14, University of Bergen, Bergen, Norway.

Newbold, P. (1978), 'Feedback induced by mesurement errors', *International Economic Review* **19**(3), 787–791.

Pierce, D. A. and Haugh, L. D. (1977), 'Causality in temporal systems', *Journal of Econometrics* **5**, 265–293.

Pollock, D. S. G. (2012), 'Band-limited stochastic processes in discrete and and continuous time', *Studies in Nonlinear Dynamica and Econometrics* **16**(1), 1 – 27.

Rozanov, Y. (1967), *Stationary random processes*, Holden-Day, San Francisco, USA.

Scherrer, W. and Deistler, M. (1998), ‘A structure theory for linear dynamic errors-in-variables models’, *SIAM Journal of Control and Optimizations* **36**(6), 2148–2175.

Seth, A. K., Chorley, P. and Barnett, L. C. (2013), ‘Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling’, *NeuroImage* **65**, 540–555.

Silvestrini, A. and Veredas, D. (2008), ‘Temporal aggregation of univariate and multivariate time series models: A survey’, *Journal of Economic Surveys* **22**(3), 458–497.

Sims, C. A. (1972), ‘Money, income and causality’, *The American Economic Review* **62**(4), 540–552.

Sims, C. A. (1974), ‘Seasonality in regression’, *Journal of the American Statistical Association* **69**(347), 618–626.

Smirnov, D. and Bezruchko, B. (2012), ‘Spurious causalities due to low temporal resolution: Towards detection of bidirectional coupling from time series’, *EPL (Europhysics Letters)* **100**, 10005p1–10005p6.

Solo, V. (2007), On causality I: sampling and noise, *in* ‘Proc. 46th IEEE Conference on Decision and Control’, pp. 3634–3639.

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

Tiao, G. C. and Wei, W. S. (1976), ‘Effect of temporal aggregation on the dynamic relationship of two time series variables’, *Biometrika* **63**(3), 513–523.

Triacca, U. (1998), ‘Non-causality: The role of the omitted variables’, *Economics Letters* **60**, 317–320.

Triacca, U. (2000), ‘On the Hsiao definition of non-causality’, *Economics Letters* **66**, 261–264.

Wallis, K. F. (1974), ‘Seasonal adjustment and relations between variables’, *Journal of the American Statistical Association* **69**(345), 18–31.

Wansbeek, T. J. and Meijer, E. (2000), *Measurement Error and Latent Variables in Econometrics*, Vol. 37, Elsevier, Amsterdam.

Wei, W. (1982), ‘Measurement of linear dependence and feedback between multiple time series: comment’, *Journal of the American Statistical Association* **77**(378), 316–319.