Mitigating the effects of measurement noise on Granger causality

Hariharan Nalatore* and Mingzhou Ding†

The J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, Florida 32611, USA

Govindan Rangarajan‡

Department of Mathematics, Indian Institute of Science, Bangalore 560 012, India

Computing Granger causal relations among bivariate experimentally observed time series has received increasing attention over the past few years. Such causal relations, if correctly estimated, can yield significant insights into the dynamical organization of the system being investigated. Since experimental measurements are inevitably contaminated by noise, it is thus important to understand the effects of such noise on Granger causality estimation. The first goal of this paper is to provide an analytical and numerical analysis of this problem. Specifically, we show that, due to noise contamination, (1) spurious causality between two measured variables can arise and (2) true causality can be suppressed. The second goal of the paper is to provide a denoising strategy to mitigate this problem. Specifically, we propose a denoising algorithm based on the combined use of the Kalman filter theory and the expectation-maximization algorithm. Numerical examples are used to demonstrate the effectiveness of the denoising approach.

PACS number(s): 05.40.Ca, 87.19.La, 84.35.+i, 02.50.Sk

I. INTRODUCTION

Granger causality [1] has become the method of choice to determine whether and how two time series exert causal influences on each other. In this method one starts by modeling simultaneously acquired time series as coming from a multivariate or vector autoregressive (VAR) stochastic process. One time series is said to have a causal influence on the other if the residual error in the autoregressive model of the second time series (at a given point of time) is reduced by incorporating past measurements from the first. This method and related methods have found applications in a wide variety of fields including physics [2–7], economics [1,10–12], and neuroscience [8,9]. Its nonlinear extension has recently appeared in [13] and has been applied to study problems in condensed matter physics [14].

The statistical basis of Granger causality estimation is linear regression. It is known that regression analysis is sensitive to the impact of measurement noise [15]. Given the inevitable occurrence of such noise in experimental time series, it is imperative that we determine whether and how such added noise can adversely affect Granger causality estimation. Previous studies [16] have suggested that such adverse effects can indeed occur. In this paper, we make further progress by obtaining analytical expressions that explicitly demonstrate how the interplay between measurement noise and system parameters affects Granger causality estimation. Moreover, we show how this deleterious effect of noise can be reduced by a denoising method, which is based on the Kalman filter theory and the expectation-maximization (EM) algorithm. We refer to our denoising algorithm as the Kalman EM (KEM) denoising algorithm.

The organization of this paper is as follows. In Sec. II, we start by introducing an alternative formulation of Granger causality [17] and proceed to outline a framework within which the effects of added (measurement) noise on the estimation of directional influences in bivariate autoregressive processes can be addressed. To simplify matters, we then consider a bivariate first-order autoregressive [AR(1)] process in Sec. III. Here explicit expressions for the effect of noise on Granger causality are derived. These expressions allow us to show that, for two time series that are unidirectionally coupled, spurious causality can arise when noise is added to the driving time series and true causality can be suppressed by the presence of noise in either time series. The theoretical results are illustrated by numerical simulations. In Sec. IV, we briefly introduce the KEM denoising algorithm and apply it to the example considered in Sec. III. Our results show that the KEM algorithm can mitigate the effects of noise and restore the true causal relations between the two time series. In Sec. V, we consider a coupled neuron model which produces time series that closely resemble that recorded in neural systems. The effect of noise on Granger causality and the effectiveness of the KEM algorithm in mitigating the noise effect are illustrated numerically. Our conclusions are given in Sec. VI.

II. THEORETICAL FRAMEWORK

Consider two time series $X(t)$ and $Y(t)$. To compute Granger causality, we model them as a combined bivariate autoregressive process of order p. In what follows, the model order p is assumed to be known, since this aspect is not central to our analysis. The bivariate autoregressive model can then be represented as

$$\sum_{k=0}^{p} [a_k X(t-k) + b_k Y(t-k)] = E_1(t), \tag{1}$$

*Electronic address: hnalatore@bme.ufl.edu
†Electronic address: mding@bme.ufl.edu
‡Electronic address: rangaraj@math.iisc.ernet.in

$$\sum_{k=0}^{p} [c_k X(t-k) + d_k Y(t-k)] = E_2(t), \qquad (2)$$

where a_k, b_k, c_k, and d_k are the AR coefficients and $E_i(t)$ are the temporally uncorrelated residual errors.

For our purposes, it is more convenient to rewrite the above bivariate process as two univariate processes (this can always be done according to [17])

$$P_1(B)X(t) = \xi(t) . \qquad P_2(B)Y(t) = \eta(t), \qquad (3)$$

where B is the lag operator defined as $B^k X(t) = X(t-k)$ and P_1 and P_2 are polynomials (of possibly infinite order) in the lag operator B. It should be noted that the new noise terms $\xi(t)$ and $\eta(t)$ are no longer uncorrelated. Let $\gamma_{12}(k)$ denote the covariance at lag k between these two noises:

$$\gamma_{12}(k) \equiv cov(\xi(t), \eta(t-k)) \quad k = \dots, -1, 0, 1, \dots . \quad (4)$$

A theorem by Pierce and Haugh [17] states that $Y(t)$ causes $X(t)$ in the Granger sense if and only if

$$\gamma_{12}(k) \neq 0 \quad \text{for some } k > 0. \qquad (5)$$

Similarly $X(t)$ causes $Y(t)$ if and only if $\gamma_{12}(k) \neq 0$ for some $k < 0$.

Now we add measurement noises $\xi'(t)$ and $\eta'(t)$ to $X(t)$ and $Y(t)$, respectively:

$$X^{(c)}(t) = X(t) + \xi'(t), \qquad (6)$$

$$Y^{(c)}(t) = Y(t) + \eta'(t). \qquad (7)$$

Here $\xi'(t)$, $\eta'(t)$ are uncorrelated white noises that are uncorrelated with $X(t)$, $Y(t)$, $\xi(t)$, and $\eta(t)$. Following Newbold [16], the above equations can be rewritten as

$$P_1(B)X^{(c)}(t) = P_1(B)X(t) + P_1(B)\xi'(t), \qquad (8)$$

$$P_2(B)Y^{(c)}(t) = P_2(B)Y(t) + P_2(B)\eta'(t). \qquad (9)$$

Using Eq. (3) we get

$$P_1(B)X^{(c)}(t) = \xi(t) + P_1(B)\xi'(t), \qquad (10)$$

$$P_2(B)Y^{(c)}(t) = \eta(t) + P_2(B)\eta'(t). \qquad (11)$$

Following the procedure in Granger and Morris [18], the linear combination of white noise processes on the right hand sides can be rewritten in terms of invertible moving average processes [19]:

$$\xi(t) + P_1(B)\xi'(t) = P_3(B)\xi^{(c)}(t), \qquad (12)$$

$$\eta(t) + P_2(B)\eta'(t) = P_4(B)\eta^{(c)}(t), \qquad (13)$$

where $\xi^{(c)}$ and $\eta^{(c)}$ are again uncorrelated white noise processes. Thus we get

$$P_3^{-1}(B)P_1(B)X^{(c)}(t) = \xi^{(c)}(t),$$

$$P_4^{-1}(B)P_2(B)Y^{(c)}(t) = \eta^{(c)}(t). \qquad (14)$$

This is again in the form of two univariate AR processes. Therefore the theorem of Pierce and Haugh can be applied to

yield the result that the noisy signal $Y^{(c)}(t)$ causes $X^{(c)}(t)$ in the Granger sense if and only if

$$\gamma_{12}^{(c)}(k) \equiv cov(\xi^{(c)}(t), \eta^{(c)}(t-k)) \neq 0, \qquad (15)$$

for some $k > 0$. Similarly $X^{(c)}(t)$ cause $Y^{(c)}(t)$ if and only if

$$\gamma_{12}^{(c)}(k) \neq 0, \qquad (16)$$

for some $k < 0$.

We can relate $\gamma_{12}^{(c)}$ to γ_{12} as follows. Consider the corresponding covariance generating functions (which are nothing but the z transforms of the cross covariances)

$$\bar{\gamma}_{12}(z) = \sum_{k=-\infty}^{\infty} \gamma_{12}(k)z^k,$$

$$\bar{\gamma}_{12}^{(c)}(z) = \sum_{k=-\infty}^{\infty} \gamma_{12}^{(c)}(k)z^k. \qquad (17)$$

We can show that [16]

$$\bar{\gamma}_{12}^{(c)}(z) = P_3^{-1}(z)P_4^{-1}(z^{-1})\bar{\gamma}_{12}(z). \qquad (18)$$

Even if $\gamma_{12}(k) = 0$ for all $k < 0$ (i.e., X does not cause Y) it is possible that $\gamma_{12}^{(c)}(k) \neq 0$ for some negative k because of the additional term $P_3^{-1}(z)P_4^{-1}(z^{-1})$ that has been introduced by the measurement noise. This gives rise to spurious Granger causality ($X^{(c)}$ causes $Y^{(c)}$), which is a consequence of the added measurement noise.

III. A BIVARIATE AR(1) PROCESS

In the previous section, we demonstrated that measurement noise can affect Granger causality. But the treatment given was quite general in nature. In this section we specialize to a simple bivariate AR(1) process and obtain explicit expressions for the effect of noise on Granger causality.

Consider the following bivariate AR(1) process:

$$X(t) = aX(t-1) + bY(t-1) + E_1(t),$$

$$Y(t) = dY(t-1) + E_2(t). \qquad (19)$$

From the above expressions, it is clear that Y drives X for nonzero values of b and X does not drive Y in this model. More specifically, we see that Y at an earlier time $t-1$ affects X at the current time t. There is no such corresponding influence of X on Y.

When noises $\xi'(t)$ and $\eta'(t)$ with variances $\sigma_{\xi'}^2$ and $\sigma_{\eta'}^2$, respectively, are added to the data generated by Eq. (19), after some algebra (see the Appendix for details), we find the following expressions for $P_3(B)$ and $P_4(B)$:

$$P_3(B) = 1 + a_1'B + a_2'B^2, \quad P_4(B) = 1 - d'B. \qquad (20)$$

Here

$$d' = \frac{s \pm \sqrt{s^2 - 4}}{2}, \qquad (21)$$

where

$$s \equiv \left(\frac{1}{d} + d\right) + \frac{1}{d}\frac{\sigma_\eta^2}{\sigma_{\eta'}^2}. \qquad (22)$$

The expressions for a_1' and a_2' are very long and for our purposes it is sufficient to note that they go to zero as the added noise goes to zero (as expected). We see that $|s| > 2$ for any values of d, σ_η^2, and $\sigma_{\eta'}^2$. Therefore $\sqrt{S^2 - 4}$ and hence d' are well defined. We also have the following results:

(a) as $|d| \rightarrow 0$, $|d'| < |d| \rightarrow 0$;

(b) as $d \rightarrow 1$, $d' \rightarrow 1 + \frac{\sigma_\eta^2}{2\sigma_{\eta'}^2} - \frac{\sigma_\eta^2}{2\sigma_{\eta'}^2}\sqrt{1 + 4/\frac{\sigma_\eta^2}{\sigma_{\eta'}^2}}$;

(c) as the ratio $\frac{\sigma_\eta^2}{\sigma_{\eta'}^2} \rightarrow 0$, $d' \rightarrow d$;

(d) as the ratio $\frac{\sigma_\eta^2}{\sigma_{\eta'}^2} \rightarrow \infty$, $d' \rightarrow 0$.

Substituting the expressions for $P_3(B)$ and $P_4(B)$ in Eq. (18) we get

$$\overline{\gamma}_{12}^{(c)}(z) = (1 + a_1'z + a_2'z^2)^{-1}(1 - d'z^{-1})^{-1}\overline{\gamma}_{12}(z). \qquad (23)$$

We now expand both sides in powers of z:

$$\cdots + \gamma_{12}^{(c)}(-1)z^{-1} + \gamma_{12}^{(c)}(0) + \gamma_{12}^{(c)}(1)z + \cdots$$
$$= [1 - a_1'z + (a_1^2 - a_2')z^2 + \cdots]$$
$$\times (1 + d'z^{-1} + d'^2z^{-2} + \cdots)$$
$$\times [\cdots + \gamma_{12}(-1)z^{-1} + \gamma_{12}(0) + \gamma_{12}(1)z + \cdots].$$
$$(24)$$

Collecting terms proportional to z^{-1}, z^0, z^1, etc., we obtain the following expressions for the cross covariances at lag -1, 0, and 1:

$$\gamma_{12}^{(c)}(-1) = d'(1 - a_1'd' + \cdots)[\gamma_{12}(0) + d'\gamma_{12}(1) + \cdots), \qquad (25)$$

$$\gamma_{12}^{(c)}(0) = (1 - a_1'd' + \cdots)[\gamma_{12}(0) + d'\gamma_{12}(1) + \cdots], \qquad (26)$$

$$\gamma_{12}^{(c)}(1) = \gamma_{12}(1) - a_1'\gamma_{12}(1) - a_1'd'\gamma_{12}(1) + \cdots. \qquad (27)$$

We observe that $\gamma_{12}^{(c)}(k)$ for $k < 0$ [and in particular $\gamma_{12}^{(c)}(-1)$] is no longer zero, implying that the $X^{(c)}$ drives $Y^{(c)}$, thus giving rise to a spurious causal direction. The spurious causality term $\gamma_{12}^{(c)}(-1)$ is proportional to d'. This can be shown to be true for all the other spurious terms $\gamma_{12}^{(c)}(k)$, $k < -1$, as well. Hence they all go to zero if $d' \rightarrow 0$ (i.e., if Y has no measurement noise). This happens even if a_1' and a_2' are nonzero (i.e., even if the X measurement is contaminated by noise). Hence we arrive at an important conclusion that, if Y is driving X, only measurement noise in Y can cause spurious causality. If Y has no measurement noise, no amount of measurement noise in X can lead to spurious causality. Further, using the asymptotic properties of d' listed earlier, we can easily see that the magnitude of the spurious causality increases as $d \rightarrow 1$ and as the ratio $\sigma_\eta^2/\sigma_{\eta'}^2 \rightarrow 0$.

The foregoing demonstrates that noise can lead to spurious causal influences that are not part of the underlying processes. Here we show that the true causality terms [$\gamma_{12}(k)$ for $k > 0$] are also modified by the presence of noise. They un-dergo a change even if $d' = 0$. For example, $\gamma_{12}(1)$ is changed to $\gamma_{12}(1) - a_1'\gamma_{12}(0)$ even if $d' = 0$. Therefore, it is quite possible that even a true causal direction can be masked by added noise and the measurement noises in both time series contribute to this suppression. As the ratios $\sigma_\xi^2/\sigma_{\xi'}^2$ and $\sigma_\eta^2/\sigma_{\eta'}^2 \rightarrow \infty$, a_1', a_2', d' all go to zero and $\gamma_{12}^{(c)} \rightarrow \gamma_{12}$, as expected.

We make one final observation. If we replace z by $e^{i2\pi f}$ (where f is the frequency) in the covariance generating function [cf. Eq. (17)] we obtain the cross spectrum. Hence all the above results carry over to the spectral or frequency domain.

To illustrate the above theoretical results, we estimate the Granger causality spectrum (in the frequency domain) for a bivariate AR process numerically. First, we briefly summarize the theory behind this computation [9]. The bivariate AR process given in Eq. (1) can be written as:

$$\sum_{k=0}^p A(k)Z(t - k) = E(t), \qquad (28)$$

where $Z(t) = [X(t), Y(t)]^T$, $E(t) = [E_1(t), E_2(t)]^T$ and

$$A(k) = \begin{pmatrix} -a_k & -b_k \\ -c_k & -d_k \end{pmatrix}, \qquad (29)$$

for $1 \leq k \leq p$. $A(0)$ is the 2×2 identity matrix. Here, $E(t)$ is a temporally uncorrelated residual error with covariance matrix Σ. We obtain estimates of the coefficient matrices $A(k)$ by solving the multivariate Yule-Walker equations [20] using the Levinson-Wiggins-Robinson (LWR) algorithm [21]. From $A(k)$ and Σ we estimate the spectral matrix $S(f)$ by the relation

$$S(f) = H(f)\Sigma H^*(f), \qquad (30)$$

where $H(f) = [\sum_{k=0}^p A(k)e^{-2\pi ikf}]^{-1}$ is the transfer function of the system.

The Granger causality spectrum from Y to X is given by [9,22] (see also [23])

$$I_{Y \rightarrow X}(f) = -\ln\left(1 - \frac{(\Sigma_{22} - \Sigma_{12}^2/\Sigma_{11})|H_{12}(f)|^2}{S_{11}(f)}\right). \qquad (31)$$

Here, Σ_{11}, Σ_{22}, and Σ_{12} are the elements of Σ and $S_{11}(f)$ is the power spectrum of X at frequency f. $H_{ij}(f)$ is the $\{ij\}$th element of the transfer function matrix $H(f)$. Similarly, the Granger causality spectrum from X to Y is defined by

$$I_{X \rightarrow Y}(f) = -\ln\left[1 - \frac{(\Sigma_{11} - \Sigma_{12}^2/\Sigma_{22})|H_{21}(f)|^2}{S_{22}(f)}\right], \qquad (32)$$

and $S_{22}(f)$ is the power spectrum of Y at frequency f.

We now estimate the Granger causality spectrum for the specific AR(1) process given in Eq. (19) where Y drives X and X does not drive Y. The parameter values used are $a = 0.4$, $b = 0.6$, $d = 0.9$, $\sigma_\xi = 0.2$, and $\sigma_\eta = 1.0$. We obtain two time series X and Y by numerically simulating the VAR model and then adding Gaussian measurement noise with $\sigma_{\xi'} = 0.2$ and $\sigma_{\eta'} = 2.5$. For concreteness we assume that each time unit corresponds to 5 ms. In other words, the sampling
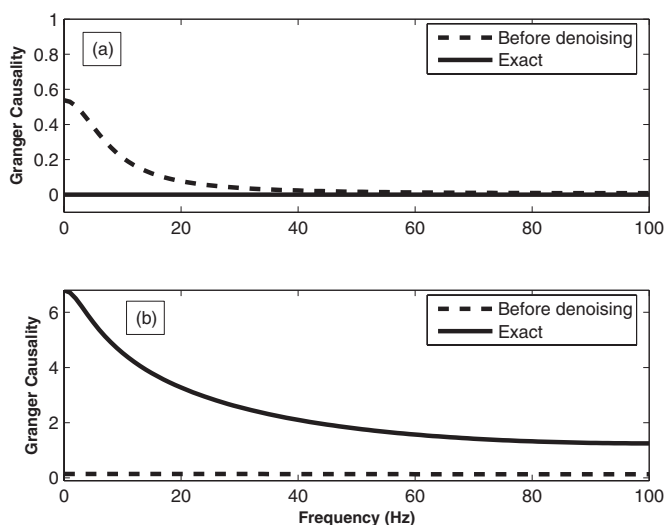
FIG. 1. Granger causality spectra for a bivariate AR(1) process. (a) Causality of $X \rightarrow Y$. (b) Causality of $Y \rightarrow X$. The solid lines represent true causality spectra and the dashed lines represent spectra from noisy data.
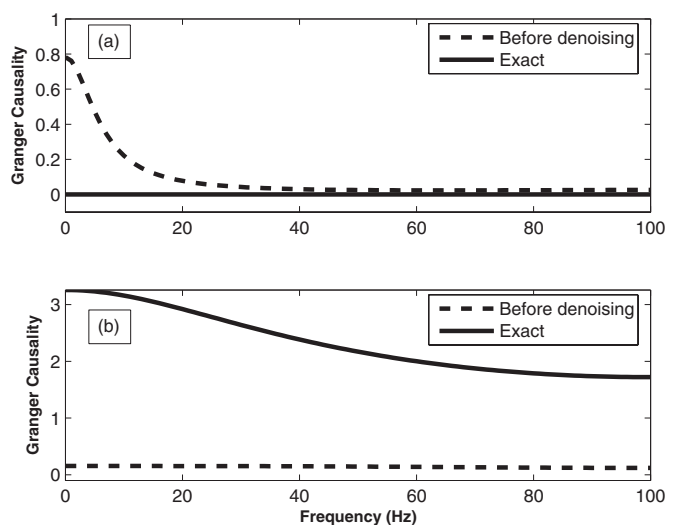

FIG. 2. Granger causality spectra for a bivariate AR(2) process. (a) Causality of $X \rightarrow Y$. (b) Causality of $Y \rightarrow X$. The solid lines represent true causality spectra and the dashed lines represent spectra from noisy data.

rate is 200 Hz, and thus the Nyquist frequency is 100 Hz. The data set consists of 100 realizations, each of length 250 ms (50 points). These 100 realizations are used to obtain expected values of the covariance matrices in the LWR and KEM algorithms (see the next section). The Granger causality spectra $I_{X \rightarrow Y}(f)$ and $I_{Y \rightarrow X}(f)$ are plotted in Fig. 1. The solid lines represents the true causality spectra while the dashed lines represent the noisy causality spectra.

Similarly, we also simulated the following bivariate AR(2) process:

$$X(t) = aX(t-1) + bY(t-1) + E_1(t),$$

$$Y(t) = d_1 Y(t-1) + d_2 Y(t-2) + E_2(t). \tag{33}$$

The values of the parameters a and b used were the same as in the previous AR(1) process example [Eq. (19)] except for the values of the new parameters d_1 and d_2 which were chosen to be 0.4 and 0.5, respectively. We again obtained two time series X and Y and then added Gaussian measurement noise with $\sigma_{\xi'} = 0.2$ and $\sigma_{\eta'} = 2.5$ to X and Y, respectively. The Granger causality spectra $I_{X \rightarrow Y}(f)$ and $I_{Y \rightarrow X}(f)$ are plotted in Fig. 2. As before, the solid lines and dashed lines represent the true causality spectra and noisy causality spectra, respectively.

We observe that the measurement noise has a dramatic effect in both of these cases: It completely reverses the true causal directions. For the noisy data, X appears to drive Y and Y does not appear to drive X.

The above theoretical and numerical results bring out clearly the adverse effect that noise can have on correctly determining directional influences. The same is also true for other quantities like the power spectrum and coherence. Therefore it is imperative that the effect of noise be mitigated to the extent possible.

IV. THE KEM DENOISING ALGORITHM

In the previous section we have seen that noisy data can lead to grave misinterpretation of directional influences. We now provide a practical solution to this problem by combining the Kalman smoother with the expectation-maximization algorithm [24]. The detailed algorithm is long and tedious. We outline the main logical steps below.

The Kalman filter [25] is a standard algorithm for denoising noisy data. To apply this, we first need to recast a VAR process with measurement noise in the so-called state-space form. This is nothing but the difference equation analog of converting a higher-order differential equation to a system of first-order differential equations. Once this is done, our VAR model takes on the following form:

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + \mathbf{w}_{t+1}, \tag{34}$$

$$\mathbf{y}_t = C\mathbf{x}_t + \mathbf{v}_t. \tag{35}$$

Here $\mathbf{x}_t$ is an $M \times 1$ ("true") state vector at time t. A is an $M \times M$ state matrix. $\mathbf{w}_t$ is a zero-mean Gaussian independent and identically distributed random variable with covariance matrix Q. Bivariate AR(p) models can be put in the form $\mathbf{x}_{t+1} = A\mathbf{x}_t + \mathbf{w}_{t+1}$ by defining $M = 2p$ auxiliary variables $x_{i,t}$. The $N \times 1$ vector $\mathbf{y}_t$ is the observed or measured value of $\mathbf{x}_t$ in N channels. C is an $N \times M$ observation matrix and is a fixed, known matrix for VAR models. Hence we will ignore this in future discussions. The $N \times 1$ vector $\mathbf{v}_t$ is the measurement noise which has zero mean, and is Gaussian and independent and identically distributed with covariance matrix R.

Kalman filter, however, cannot be directly applied to denoise experimental or observed data since it assumes the knowledge of the model describing the state-space dynamics. In practice, such knowledge is often not available. To get around this problem, we apply the Kalman smoother in conjunction with the expectation and maximization algorithm

[24,26–28]. Thus, this denoising algorithm will henceforth be called the KEM algorithm. In this algorithm, one follows the standard procedure for estimating state-space parameters from data using the maximum likelihood method. The appropriate likelihood function in our case is the joint log likelihood log $P(\{\mathbf{x}\},\{\mathbf{y}\})$ where $\{\mathbf{x}\}$ denotes $\{\mathbf{x}_t\}$ (for all t) and similarly for $\{\mathbf{y}\}$. In the usual maximum likelihood method, P would not depend on $\mathbf{x}$ and we would therefore maximize the above quantity directly (conditioned on the observed $\mathbf{y}_t$ values) and obtain the unknown state-space parameters. But in our case, P depends on $\mathbf{x}$ which is also unknown. To get rid of $\mathbf{x}$, we take the expected value of the log likelihood

$$O = E[\log P(\{\mathbf{x}\},\{\mathbf{y}\})|\{\mathbf{y}\}].$$

As usual, we have conditioned the expectation on the known observations $\{\mathbf{y}\}$.

To compute O, it turns out we need the expectations of $\mathbf{x}$ and $\mathbf{x}\mathbf{x}^T$ (where T denotes the transpose) conditioned on $\mathbf{y}$. These expectations are obtained by applying the Kalman smoother on the noisy data. We use the Kalman smoother and not the Kalman filter since we are utilizing all the observations $\mathbf{y}$ instead of only the past observations. This is the appropriate thing to do in our case since we are performing an off-line analysis where all observations are known. In other words, in the Kalman smoother, we perform both a forward pass and a backward pass on the data in order to make use of all observations.

To apply the Kalman smoother, however, we still need the state-space model parameters (just as in the Kalman filter case). To circumvent this problem, we start with initial estimates for these parameters (A, Q, and R) as follows. From the noisy data, using the LWR algorithm, we obtain the VAR model coefficient matrices [8]. Then a standard transformation [25] is used to put these matrices in the state-space form giving the initial estimate for A. The initial estimate of Q is taken to be the identity matrix following the standard procedure [25]. The initial estimate of R is taken to be half the covariance matrix at lag zero of the noisy data. The approximate model order can be determined by applying the akaike information criterion (AIC) [29] in the LWR algorithm. This step is admittedly rather *ad hoc*. Further studies to optimize the above initial estimates and the VAR model order p are currently being carried out. Once we have initial estimates of the model parameters, we can apply the Kalman smoother to obtain the various conditional expectations and evaluate the expected log likelihood O. This is called the expectation (E) step.

Next, we go to the maximization (M) step. Each of the parameters A, Q, R, etc. is reestimated by maximizing O. Using these improved estimates, we can apply the E step again followed by the M step. This iterative process is continued untill the value of log likelihood function converges to a maximum. We could now directly use the VAR parameters estimated from the KEM algorithm for further analysis as is usually done. But here we prefer to use the following procedure which was found to yield better performance. The final denoised data (that is, the estimate of $\mathbf{x}$ obtained from the KEM algorithm) are treated as the new experimental time series and subjected to parametric spectral analysis from
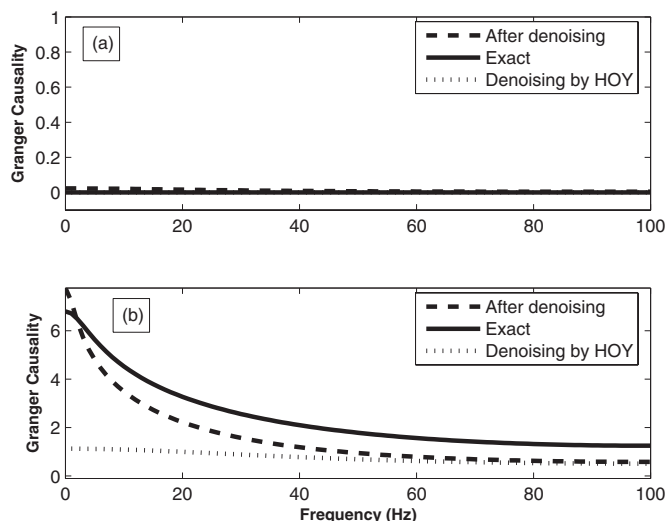

FIG. 3. Granger causality spectra for the bivariate AR(1) process in Fig. 1. (a) Causality of $X \rightarrow Y$. (b) Causality of $Y \rightarrow X$. The solid lines represent true causality spectra and the dashed lines represent spectra obtained from the denoised data using the KEM algorithm. The dotted lines represent spectra obtained using the HOY algorithm.

which Granger causality measures can be derived. The Matlab code implementing this algorithm for our applications is available from the authors upon request.

We have compared the denoising capabilities of the KEM algorithm with two widely used algorithms, the higher-order Yule-Walker (HOY) method [30] and the overdetermined higher-order Yule-Walker method [31]. We find that the denoising capabilities of the KEM algorithm are superior. Detailed results will be presented elsewhere. In Fig. 3, we explicitly show that the KEM algorithm performs better than the HOY method (see below).

The KEM algorithm is applied to denoise the data shown in Figs. 1 and 2. Figure 3 displays the same exact Granger causality spectra (solid lines) as those in Fig. 1 and the Granger causality spectra (dashed lines) obtained from the denoised data using the KEM algorithm. Causality spectra obtained using the HOY method are also shown (as dotted lines). It is clear that the KEM method performs better. In Fig. 4, the solid lines again represent the same exact Granger causality as that in Fig. 2 and the dashed lines represent the Granger causality spectra obtained from the denoised data of a bivariate AR(2) process. We see that the correct causal directions are recovered and that the denoised spectra are reasonably close to the true causality spectra for both AR(1) and AR(2) processes. We stress that these results are achieved without assuming any knowledge of the VAR models [Eqs. (19) and (33)] that generated the original time series data.

V. CAUSAL RELATIONS IN A NEURAL NETWORK MODEL

In this section, we analyze the effect of noise on time series generated by a neural network model. We first demon-
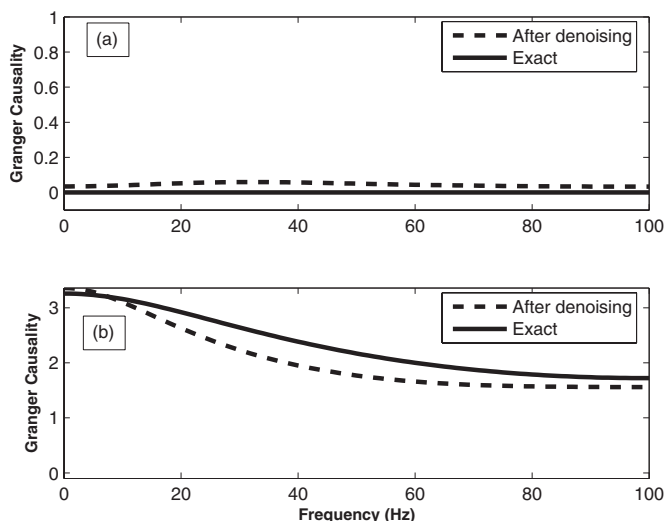
FIG. 4. Granger causality spectra for the bivariate AR(2) process in Fig. 2. (a) Causality of $X \to Y$. (b) Causality of $Y \to X$. The solid lines represent true causality spectra and the dashed lines represent spectra obtained from the denoised data using the KEM algorithm.

strate the effect of measurement noise on causality directions and then the effect of applying the KEM algorithm on the noisy data.

Our simulation model comprises two coupled cortical columns where each column is made up of an excitatory and an inhibitory neuronal population [32]. The equations governing the dynamics of the two columns are given by

$$\ddot{x}_i + (a+b)\dot{x}_i + abx_i = -k_{ei}Q(y_i(t), Q_{m0}) + k_{ij}Q(x_j(t), Q_{m0})$$
$$+ \xi_{x_i}(t), \qquad (36)$$

$$\ddot{y}_i + (a+b)\dot{y}_i + aby_i = k_{ie}Q(x_i(t), Q_{m0}) + \xi_{y_i}(t), \quad (37)$$

where $i \neq j = 1, 2$. Here x and y represent local field potentials of the excitatory and inhibitory populations, respectively, $k_{ie} > 0$ gives the coupling gain from the excitatory (x) to the inhibitory (y) population, and $k_{ei} > 0$ is the strength of the reciprocal coupling. The neuronal populations are coupled through a sigmoidal function $Q(x, Q_{m0})$ which represents the pulse densities converted from x with Q_{m0} a modulatory parameter. The function $Q(x, Q_{m0})$ is defined by

$$Q(x, Q_{m0}) = \begin{cases} Q_{m0}(1 - e^{-(e^x - 1)/Q_{m0}}) & \text{if } x > -u_0, \\ -1 & \text{if } x \leq -u_0, \end{cases}$$
$$(38)$$

where $u_0 = -\ln[1 + \ln(1 + 1/Q_{m0})]$. The coupling strength k_{ij} is the gain from the excitatory population of column j to the excitatory population of column i, with $k_{ij} = 0$ for $i = j$. The terms $\xi(t)$ represent independent Gaussian white noise inputs given to each neuronal population.

The parameter values used were $a = 0.22/\text{ms}$, $b = 0.72/\text{ms}$, $k_{ie} = 0.1$, $k_{ei} = 0.4$, $k_{12} = 0$, $k_{21} = 0.25$, and $Q_{m0} = 5$. The standard deviation for the Gaussian white noise was chosen as 0.2. Assuming a sampling rate of 200 Hz, 200
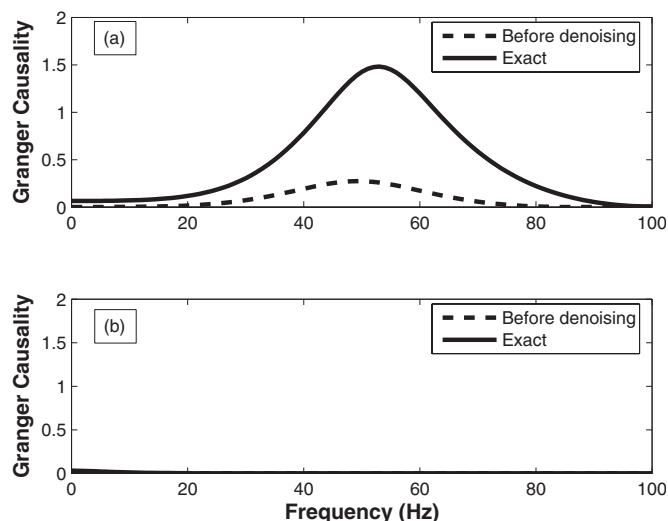

FIG. 5. Granger causality spectra for noisy data from a neural network model. (a) Causality of $x_1 \to x_2$. (b) Causality of $x_2 \to x_1$. The solid lines represent true causality spectra (noise-free data) and the dashed lines represent spectra from noisy data.

realizations of the signals were generated, each of length 30 s (6000 points).

We now restrict our attention to the variables $x_1(t)$ and $x_2(t)$. Measurement noises (Gaussian white noises with standard deviations 2.0 and 3.0, respectively) were added to these variables. From the model it is clear that $x_1(t)$ should drive $x_2(t)$ since $k_{12} = 0$ while $k_{21} = 0.25$. The results of applying Granger causality analysis (using a VAR model of order 7) on these two variables are shown in Fig. 5. The solid lines represent the causality spectra for the noise-free data. The dashed lines represent the causality spectra for the noisy data. It is clear that the measurement noise has an effect on the causal relations by significantly reducing the true causality magnitude. In contrast to the example in Section III, however, no spurious causal direction is generated here, despite the fact that both time series are contaminated by measurement noise. Next, we applied the KEM algorithm to denoise the noisy data. When Granger causality analysis is performed on the denoised data, we obtain causality spectra that are closer to the true causality spectra (see Fig. 6). We note that the KEM algorithm is not able to completely remove the noise as the denoised spectra are still quite different from the true spectra.

To show that the denoised Granger spectrum is significantly different from that of the noisy data we use the bootstrap approach [33] to establish the significant difference between the two peaks observed in the Granger causality spectra of Figs. 5 and 6 (shown by dashed lines in these figures). One thousand resamples of noisy data and the denoised data were generated by randomly selecting trials with replacement. It should be noted that in any selected trial, the entire multichannel data are taken, thus preserving the auto- and cross-correlation structures. Thus, we employ a version of the block bootstrap method [33]. The peak values of Granger causality were computed for each resample using both noisy data and denoised data. Let us denote these peak values by the random variables Z_1 and Z_2, respectively. The
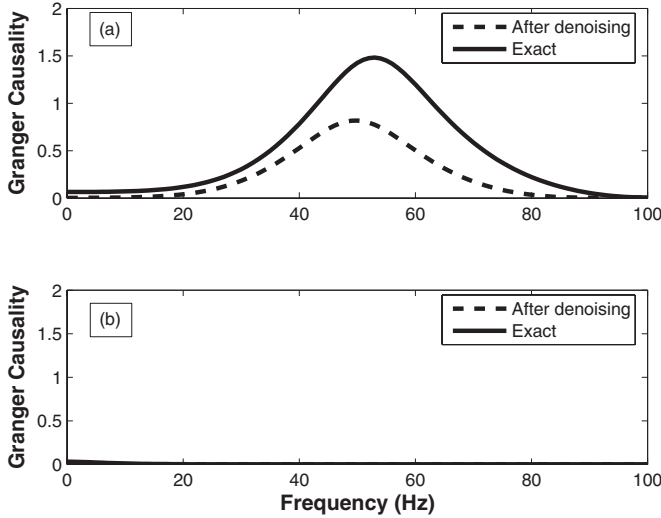
FIG. 6. Granger causality spectra of the neural network model. (a) Causality of $x_1 \rightarrow x_2$. (b) Causality of $x_2 \rightarrow x_1$. The solid lines represent true causality spectra (noise-free data) and the dashed lines represent spectra obtained from denoised data using the KEM algorithm.

two-population Student t test was performed to determine whether the means of Z_1 and Z_2 are different at a statistically significant level.

The null hypothesis was that the means of the two populations Z_1 and Z_2 are equal. The t value was found to be very large, 4.6446×10^3, and corresponds to a two-tailed p value less than 0.0001. Thus the null hypothesis that the two groups do not differ in mean is rejected. This establishes the fact that the peak of the Granger causality spectrum of the denoised data is significantly higher than that of the noisy data. Figure 7 shows the plot of Granger causality for the direction $x_1 \rightarrow x_2$ along with 95% confidence intervals. The
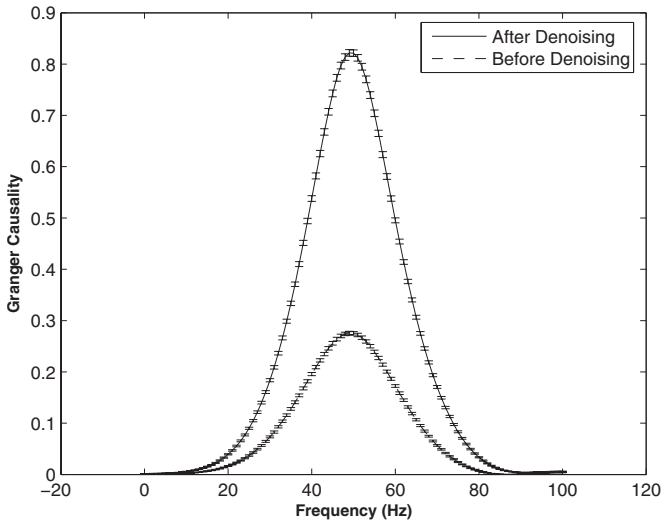

FIG. 7. Granger causality spectra of the neural network model for the direction $x_1 \rightarrow x_2$. The solid line represents the Granger causality for denoised data, while the dashed line represents the Granger causality for noisy data. 95% confidence intervals are also given.

95% confidence intervals are calculated as $I_{x_1 \rightarrow x_2}(f) \pm 1.96\sigma_B$ (for each frequency f) where σ_B is the sample standard deviation of the 1000 bootstrap replications of $I_{x_1 \rightarrow x_2}(f)$.

VI. CONCLUSIONS

Our contributions in this paper are twofold. First, we demonstrate that measurement noise can significantly impact Granger causality analysis. Based on analytical expressions linking noise strengths and the VAR model parameters, it was shown that spurious causality can arise and that true causality can be suppressed due to noise contamination. Numerical simulations were performed to illustrate the theoretical results. Second, a practical solution to the measurement noise problem, called the KEM algorithm, was outlined, which combines the Kalman filter theory with the expectation and maximization algorithm. It was shown that the application of this algorithm to denoise the noisy data can significantly mitigate the deleterious effects of measurement noise on Granger causality estimation. It is worth noting that, despite the fact that the adverse effect of measurement noise on Granger causality has been known since 1978 [16], mitigation of this effect has received little attention. The KEM algorithm described in this paper is our attempt at addressing this shortcoming.

APPENDIX

In this appendix, we derive the expressions for $P_3(B)$ and $P_4(B)$ given in Eq. (20). We first determine $P_4(B)$. When a zero-mean white noise process $\eta'(t)$ with variance $\sigma^2_{\eta'}$ is added to $Y(t)$ we get

$$Y^{(c)}(t) = Y(t) + \eta'(t). \tag{A1}$$

Applying $(1 - dB)$ on both sides of the above equation we get

$$(1 - dB)Y^{(c)}(t) = (1 - dB)Y(t) + (1 - dB)\eta'(t)$$
$$= \eta(t) + (1 - dB)\eta'(t). \tag{A2}$$

We now determine a white noise process $\eta^{(c)}(t)$ such that

$$\eta(t) + (1 - dB)\eta'(t) = (1 - d'B)\eta^{(c)}(t). \tag{A3}$$

We need to determine d' and $\sigma^2_{\eta^{(c)}}$.

Taking variances on both sides of the above equation we get

$$\sigma^2_\eta + (1 + d^2)\sigma^2_{\eta'} = (1 + d'^2)\sigma^2_{\eta^{(c)}}. \tag{A4}$$

Taking the autocovariance at lag 1 on both sides we obtain

$$d\sigma^2_{\eta'} = d'\sigma^2_{\eta^{(c)}}. \tag{A5}$$

Since $\eta^{(c)}$ is a sum of η and $(1 - dB)\eta'$, we have $\sigma^2_{\eta^{(c)}} > \sigma^2_{\eta'}$. This implies that $|d'| < |d|$. Since stationarity of the

AR process requires $0 < |d| < 1$, we obtain the inequality $0 < |d'| < |d'| < 1$. Further, d' has the same sign as d.

We have

$$\sigma^2_{\eta^{(c)}} = \frac{d}{d'} \sigma^2_{\eta'}. \tag{A6}$$

Substituting in the variance equation we get

$$(1 + d'^2) \frac{d}{d'} \sigma^2_{\eta'} = (1 + d^2) \sigma^2_{\eta'} + \sigma^2_{\eta}, \tag{A7}$$

that is,

$$\left(\frac{1}{d'} + d'\right) = \left(\frac{1}{d} + d\right) + \frac{1}{d} \frac{\sigma^2_{\eta}}{\sigma^2_{\eta'}}. \tag{A8}$$

Let

$$s \equiv \left(\frac{1}{d} + d\right) + \frac{1}{d} \frac{\sigma^2_{\eta}}{\sigma^2_{\eta'}}.$$

This gives

$$\left(\frac{1}{d'} + d'\right) = s. \tag{A9}$$

Hence

$$d' = \frac{s \pm \sqrt{s^2 - 4}}{2}. \tag{A10}$$

Note that $|s| > 2$ for any values of d, σ^2_{η}, and $\sigma^2_{\eta'}$. Therefore $\sqrt{s^2 - 4}$ and hence d' are well defined. Further, since $|d'| < |d|$ if d is positive, $d' = (s - \sqrt{s^2 - 4})/2$ is the only valid solution. If d is negative, $d' = (s + \sqrt{s^2 - 4})/2$ is the only valid solution.

Next, we derive the expression for $P_3(B)$. First, we first need to rewrite $X(t)$ as an univariate process, i.e., we need to determine $P_1(B)$:

$$P_1(B)X(t) = \xi(t), \tag{A11}$$

where $\xi(t)$ is a zero-mean white noise process and

$$X(t) = aX(t-1) + bY(t-1) + E_1(t). \tag{A12}$$

Here $E_1(t)$ is a zero-mean white noise process with variance σ^2_{ϵ}. We have already seen that

$$(1 - dB)Y(t) = \eta(t). \tag{A13}$$

The equation for $X(t)$ can be written as

$$(1 - aB)X(t) = bY(t-1) + E_1(t). \tag{A14}$$

Substituting the expression for $Y(t-1)$ we obtain

$$(1 - aB)X(t) = b(1 - dB)^{-1}\eta(t-1) + E_1(t). \tag{A15}$$

We now find a white noise process $\xi(t)$ with variance σ^2_{ξ} such that

$$b(1 - dB)^{-1}\eta(t-1) + E_1(t) = (1 - rB)^{-1}\xi(t). \tag{A16}$$

To determine r and σ^2_{ξ}, we take the variance and autocovariance at lag 1 on both sides. Taking the variance we obtain

$$\frac{b^2\sigma^2_{\eta}}{(1 - d^2)} + \sigma^2_{\epsilon} = \frac{\sigma^2_{\xi}}{(1 - r^2)}. \tag{A17}$$

Taking the autocovariance at lag 1 and assuming that $\sigma_{\epsilon\eta}$ (the cross covariance between E_1 and ξ) is zero for simplicity, we get

$$\frac{b^2\sigma^2_{\eta}d}{(1 - d^2)} = \frac{\sigma^2_{\xi}}{(1 - r^2)}, \tag{A18}$$

which can be written as

$$\frac{\sigma^2_{\xi}}{(1 - r^2)} = \frac{b^2\sigma^2_{\eta}d}{(1 - d^2)r}. \tag{A19}$$

Substituting in the variance equation we obtain

$$\frac{b^2\sigma^2_{\eta}}{(1 - d^2)} + \sigma^2_{\epsilon} = \frac{b^2\sigma^2_{\eta}}{(1 - d^2)} \frac{d}{r}. \tag{A20}$$

Thus

$$r = \frac{b^2 d\sigma^2_{\eta}}{b^2\sigma^2_{\eta} + (1 - d^2)\sigma^2_1}. \tag{A21}$$

If $b = 0$, we get $r = 0$ and $\sigma^2_{\xi} = \sigma^2_1$ as expected. Similarly if $d = 0$, we get $r = 0$ and $\sigma^2_{\xi} = \sigma^2_1 + b^2\sigma^2_{\eta}$ as expected. Once r is known, σ^2_{ξ} is given by

$$\sigma^2_{\xi} = (1 - r^2)\left(\frac{b^2\sigma^2_{\eta}}{(1 - d^2)} + \sigma^2_1\right). \tag{A22}$$

We finally have

$$(1 - aB)X(t) = (1 - rB)^{-1}\xi(t). \tag{A23}$$

That is,

$$P_1(B)X(t) = \xi(t), \quad P_1(B) = (1 - rB)(1 - aB). \tag{A24}$$

Consider a white noise process $\xi'(t)$ [which is uncorrelated with $X(t)$] with variance $\sigma^2_{\xi'}$. This is added to $X(t)$ to obtain the noisy process $X^{(c)}(t)$:

$$X^{(c)}(t) = X(t) + \xi'(t). \tag{A25}$$

Applying $P_1(B)$ on both sides of the above equation,

$$P_1(B)X^{(c)}(t) = \xi(t) + P_1(B)\xi'(t). \tag{A26}$$

We need to find a zero-mean white noise process $\xi^{(c)}(t)$ with variance $\sigma^2_{\xi^{(c)}}$ such that

$$\xi(t) + P_1(B)\xi'(t) = P_3(B)\xi^{(c)}(t). \tag{A27}$$

Let

$$P_3(B) = 1 + a'_1 B + a'_2 B^2. \tag{A28}$$

We have

$$\xi(t) + [1 - (a + r)B + arB^2]\xi'(t) = [1 + a'_1 B + a'_2 B^2]\xi^{(c)}(t). \tag{A29}$$

Taking the variances on both sides we get

$$\sigma_\xi^2 + [1 + (a+r)^2 + a^2 r^2]\sigma_{\xi'}^2 = (1 + a_1'^2 + a_2'^2)\sigma_{\xi(c)}^2. \tag{A30}$$

Taking the autocovariance at lag 1 on both sides we obtain

$$-(a+r)\sigma_{\xi'}^2 - ar(a+r)\sigma_{\xi'}^2 = a_1'\sigma_{\xi(c)}^2 + a_1' a_2'\sigma_{\xi(c)}^2. \tag{A31}$$

This can be rewritten as

$$-(a+r)(1+ar)\sigma_{\xi'}^2 = a_1'(1 + a_2')\sigma_{\xi(c)}^2. \tag{A32}$$

Taking the autocovariance at lag 2 on both sides

$$ar\sigma_{\xi'}^2 = a_2'\sigma_{\xi(c)}^2, \tag{A33}$$

which gives

$$\sigma_{\xi(c)}^2 = \frac{ar}{a_2'}\sigma_{\xi'}^2. \tag{A34}$$

Since $\sigma_{\xi(c)}^2 > \sigma_{\xi'}^2$, we see that $|a_2'| < |ar|$ and a_2' has the same sign as ar.

Substituting the last equation in Eqs. (A32) and (A30) we obtain

$$-(a+r)(1+ar)\sigma_{\xi'}^2 = a_1'(1 + a_2')\frac{ar}{a_2'}\sigma_{\xi'}^2, \tag{A35}$$

and

$$\sigma_\xi^2 + [1 + (a+r)^2 + a^2 r^2]\sigma_{\xi'}^2 = [1 + a_1'^2 + a_2'^2]\frac{ar}{a_2'}\sigma_{\xi'}^2. \tag{A36}$$

Thus we get

$$\frac{a_1'(1 + a_2')}{a_2'} = -\frac{(a+r)(1+ar)}{ar} \tag{A37}$$

and

$$\frac{(1 + a_1'^2 + a_2'^2)}{a_2'} = \frac{[1 + (a+r)^2 + a^2 r^2]}{a_r} + \frac{1}{ar}\frac{\sigma_\xi^2}{\sigma_{\xi'}^2}. \tag{A38}$$

We can solve these two equations for a_1' and a_2'. There will be multiple solutions. We choose that solution for which

$|a_2'| < |ar|$. Further, the solution has to be such that the roots of $1 + a_1' B + a_2' B^2 = 0$ lie outside the unit circle. The last condition is required for the invertibility of the moving average (MA) process $(1 + a_1' B + a_2' B^2)\xi^{(c)}(t)$. The expressions for a_1' and a_2' obtained by solving the above equations are very long and therefore we do not list them here. However, we can easily obtain the asymptotic behavior of these solutions as follows.

For our bivariate AR(1)process to be stable, we require that the roots of

$$\det[\lambda I - A(1)] = 0 \tag{A39}$$

lie within the unit circle, i.e., the eigenvalues of $A(1)$ should have absolute value less than 1. In our case

$$A(1) = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix},$$

which is an upper triangular matrix. Hence the eigenvalues are a and d. Therefore, for stability we require that $|a| < 1$ and $|d| < 1$.

As already derived, we have

$$r = d\left(\frac{b^2\sigma_\eta^2}{b^2\sigma_\eta^2 + (1 - d^2)\sigma_\xi^2}\right). \tag{A40}$$

Since $|d| < 1$, the term within parentheses is always positive and less than 1. It becomes zero only when $b = 0$. Hence $|r| < |d|$ and r has same sign as d. As $|d| \to 1$, $|r| \to 1$. As $|d| \to 0$ or $|b| \to 0$, we see that $|r| \to 0$.

We have already seen that $|a_2'| < |ar|$. Since $|r| < |d|$, we obtain the further results that $|a_2'| < |a||d|$ and a_2' has same sign as ad. Since $|a|, |d| < 1$, we get

$$0 < |a_2'| < |a||d| < 1.$$

As $|a|, |d| \to 1$, $|a_2'|$ also $\to 1$. As $a \to 1$, $d \to 1$, and the ratio $\sigma_\xi^2 / \sigma_{\xi'}^2 \to 0$, we have

$$a_1' \to -2; \quad a_1' \to 1.$$

As the variance ratio $\to \infty$

$$a_1' \to 0; \quad a_2' \to 0,$$

as expected. The parameter a_1' is hardly affected by the value of the parameter b. On the other hand, $a_2' \to 0$ as $b \to 0$ and saturates rapidly for $b > 0.5$.

[1] C. W. J. Granger, Econometrica **37**, 424 (1969).

[2] K. J. Blinowska, R. Kus, and M. Kaminski, Phys. Rev. E **70**, 050902(R) (2004).

[3] D. Marinazzo, M. Pellicoro, and S. Stramaglia, Phys. Rev. E **73**, 066216 (2006).

[4] P. F. Verdes, Phys. Rev. E **72**, 026222 (2005).

[5] M. G. Rosenblum and A. S. Pikovsky, Phys. Rev. E **64**, 045202(R) (2001).

[6] X. Hu and V. Nenov, Phys. Rev. E **69**, 026206 (2004).

[7] L. Xu, Z. Chen, K. Hu, H. E. Stanley and P. C. Ivanov, Phys. Rev. E **73**, 065201(R) (2006).

[8] M. Ding, S. L. Bressler, W. Yang, and H. Liang, Biol. Cybern. **84**, 463 (2000).

[9] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. L. Bressler, Proc. Natl. Acad. Sci. U.S.A. **101**, 9849 (2004).

[10] D. L. Thornton and D. S. Batten, J. Money, Credit Banking **17**, 164 (1985).

[11] T. E. Hall and N. R. Noble, J. Money, Credit Banking **19**, 112

(1987).

[12] C. Hiemstra and J. D. Jones, J. Financ. **49**, 1639 (1994).

[13] Y. Chen, G. Rangarajan, J. Feng, and M. Ding, Phys. Lett. A **324**, 26 (2004).

[14] R. Ganapathy, G. Rangarajan, and A. K. Sood, Phys. Rev. E **75**, 016211 (2007).

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

[16] P. Newbold, Int. Econom. Rev. **19**, 787 (1978).

[17] D. A. Pierce and L. D. Haugh, J. Econometr. **5**, 265 (1977).

[18] C. W. J. Granger and M. J. Morris, J. R. Stat. Soc. Ser. A (Gen.) **139**, 246 (1976).

[19] A. Maravall and A. Mathis, J. Econometr. **61**, 197 (1994).

[20] C. Chatfield, *The Analysis of Time Series* (Chapman and Hall, Boca Raton, FL, 2004).

[21] M. Morf, A. Vieira, D. Lee, and T. Kailath, IEEE Trans. Geosci. Electron. **16**, 85 (1978).

[22] J. Geweke, J. Am. Stat. Assoc. **77**, 304 (1982).

[23] Y. Hosoya, Probab. Theory Relat. Fields **88**, 429 (1991).

[24] A. P. Dempster, N. M. Laird, and D. B. Rubin, J. R. Stat. Soc. Ser. B (Methodol.) **39**, 1 (1977).

[25] S. Haykin, *Adaptive Filter Theory* (Prentice-Hall, New York, 2001).

[26] Z. Gahramani and G. E. Hinton, University of Toronto, Technical Report No. CRG-TR-96–2, 1996 (unpublished).

[27] E. Weinstein, A. V. Oppenheim, M. Feder, and J. R. Buck, IEEE Trans. Signal Process. **42**, 846 (1994).

[28] V. Digalakis, J. R. Rohlicek, and M. Ostendorf, IEEE Trans. Speech Audio Process. **1**, 431 (1993).

[29] H. Akaike, IEEE Trans. Autom. Control **AC-19**, 716 (1974).

[30] Y. T. Chan and R. Langford, IEEE Trans. Acoust., Speech, Signal Process. **30**, 689 (1980).

[31] J. A. Cadzow, Proc. IEEE **70**, 907 (1982).

[32] M. Kaminski, M. Ding, W. A. Truccolo, and S. L. Bressler, Biol. Cybern. **85**, 145 (2001).

[33] B. Efron, *The Jackknife, the Bootstrap, and Other Resampling Plans* (SIAM, Philadephia, 1982).