Statistical Tests for Detecting Granger Causality

Ribhu Chopra, Chandra R. Murthy, and Govindan Rangarajan

***Abstract*—Detection of a causal relationship between two or more sets of data is an important problem across various scientific disciplines. The Granger causality index and its derivatives are important metrics developed and used for this purpose. However, the test statistics based on these metrics ignore the effect of practical measurement impairments such as subsampling, additive noise, and finite sample effects. In this paper, we model the problem of detecting a causal relationship between two time series as a binary hypothesis test with the null and alternate hypotheses corresponding to the absence and presence of a causal relationship, respectively. We derive the distribution of the test statistic under the two hypotheses and show that measurement impairments can lead to suppression of a causal relationship between the signals, as well as false detection of a causal relationship, where there is none. We also use the derived results to propose two alternative test statistics for causality detection. These detectors are analytically tractable, which allows us to design the detection threshold and determine the number of samples required to achieve a given missed detection and false alarm rate. Finally, we validate the derived results using extensive Monte Carlo simulations as well as experiments based on real-world data, and illustrate the dependence of detection performance of the conventional and proposed causality detectors on parameters such as the additive noise variance and the strength of the causal relationship.**

I. Introduction

A. Motivation

Confirming the presence or absence of a causative relationship between different observed phenomena is an important problem across various disciplines of physical, biological, and social sciences. It has been argued that if the phenomena of interest can be represented as a time series, then the corresponding relationship between them can be quantified by directed mutual information [1]. It is also known that for Gaussian distributed time series, the directed mutual information reduces to the logarithm of ratio of the prediction error variances of the time series of interest with and without including the purported causing time series into the prediction model [2]. This quantity, known as the *Granger causality index* (GCI) [2], [3], has been applied extensively for the detection of causal relationships across numerous research areas such as neuroscience [4]–[11], physics [12], [13], climate science [14], econometrics [15], etc. Most of the present studies on Granger causality assume that the sampling process is noiseless and the sampling frequency is above the Nyquist rate. As a consequence, the tests are not designed to achieve a given false alarm/detection rate. It is also typically assumed that the second-order statistics of the signals of interest are perfectly known. However, in practice, the causality detector has to estimate the second-order statistics from a finite number of possibly undersampled and noisy measurements of the signals of interest. In this paper, our main focus is on the quantification of the effects of these measurement inaccuracies on the performance of Granger causality detectors. This, in turn, allows us to design the detector to achieve a desired false alarm/detection probability.

The phenomenon of undersampling arises mainly due to the inability of the signal acquisition unit to sample the signal of interest above the Nyquist rate. As shown later in the paper, under-sampling can lead to a weakening or even a complete suppression of the causal relationship between two signals. In addition to this, additive noise can lead to suppression of an existing causal relationship, as well as may cause a false detection [16], [17]. The effect of these inaccuracies is further compounded by the fact that the second-order statistics of the signal of interest, required to compute the GCI, are not known, and have to be estimated using the acquired samples. Therefore, finite sample effects, in addition to undersampling and additive noise, also contribute to errors in the detection of a causal relationship between the signals of interest. In this paper, we model the problem of detection of Granger causality as a binary hypothesis test, and evaluate the effects of the measurement impairments listed above on the probabilities of detection and false alarm.

B. Related Work

In the original formulation of Granger causality in [2], a time series $y[n]$ is said to *Granger cause* another time series $x[n]$, when the past samples of the $y[n]$ can be used to improve the estimate of the $x[n]$ over 'what can be predicted using all other information in the universe', conventionally considered to be the past samples of $x[n]$. Therefore, a causal relationship between $x[n]$ and $y[n]$ can be inferred by comparing the mean squared prediction error for $x[n]$ with and without including $y[n]$ in the prediction model [18]. The prediction model considering only the past samples of $x[n]$ is known as the restrictive model (R-model), whereas the one including the effects of $y[n]$ is known as the unrestricted model (U-model). In addition to the GCI, several metrics for quantifying Granger causality have been studied [19], [20]. The system model with two time series [2] has been also generalized to study causal relationships between more than two time series [21]. However, all these studies assume the samples are noiseless, and ignore the finite sample effects on the second-order statistics being estimated.

R. Chopra is with the Indian Institute of Technology Guwahati, Assam, India. (email: ribhu@outlook.com).

C. R. Murthy and G. Rangarajan are with the Indian Institute of Science, Bangalore, India. (emails: cmurthy@iisc.ac.in, rangaraj@iisc.ac.in).

The work of C. R. Murthy was supported in part by a research grant from the Aerospace Network Research Consortium and by the Young Faculty Research Fellowship from the Ministry of Electronics and Information Technology, Govt. of India. The work of G. Rangarajan was supported by research grants from JC Bose National Fellowship, DST Centre for Mathematical Biology Phase II and UGC Centre for Advanced Study.

The effect of noise on Granger causality was first discussed in [22], and was mathematically analyzed in [16] for a bivariate VAR-1 process. These results were extended to a generalized VAR(p) (pth order autoregressive process) in [17], where a Kalman filtering based technique to alleviate the detrimental effect of noise was also developed. An alternative test statistic for causality detection, known as the phase slope index, was introduced in [23] and was shown to be more resilient to noise, as compared to the GCI. Another metric for causality, known as the time reversed Granger causality, has been discussed in [18], [24], [25]. The authors in [26] used simulation techniques to compare the robustness of GCI, time reversed Granger causality, and phase slope index to correlated noise, and found that time reversed Granger causality is the most noise resilient.

The problem of sub-sampling in Granger causality detection has been studied in the literature [27]–[34]. In [27], [28], the authors study the effects of the choice of the sampling frequency on the detectability of a causal relationship between neurological signals. In [28], these derived results are used to identify favorable sampling rates for detecting Granger causality. In [29], two methods, based on the expectation maximization (EM) algorithm and the variational inference framework, are proposed to detect causal relationships from low resolution data, and their applicability to simulated and practical data is tested. The problem of causal inference from an under-sampled time series by using a general purpose Boolean constraint solver has been discussed in [32]. It is shown that the method considered in [32] is independent of the modeling parameters, and can be conveniently scaled to any number of variables. In [33], the authors propose three sampling rate agnostic algorithms for the recovery of a causal relationship between observed time series. The authors in [34] incorporate the effect of both down-sampling and additive noise in their data model. However, none of these studies consider the effect of finite samples on the detection performance, or model the problem of detection of Granger causality as a binary hypothesis testing problem. They also do not analyze the effects of these physical impairments on the probabilities of detection and false alarm.

The authors in [1] consider the related problem of estimation of directed mutual information between two time series, using a finite number of samples. Here, the test for a causal relationship between two time series is modeled as a binary hypothesis test, and it is shown that the probabilities of missed detection and false alarm asymptotically converge to zero as the number of samples is increased. In this paper, we propose to extend the binary hypothesis testing framework for causality detection to GCI.

C. Contributions

In the present work, we consider the problem of detecting a causal relationship between two possibly sub-sampled time series, in the presence of additive white Gaussian noise. We model the time series as two independent pth order autoregressive (AR(p)) processes in the absence of a causal relation, and as a bivariate vector autoregressive (VAR(p)) process in the presence of a causal relationship. We model the test for causality as a binary hypothesis testing problem with the null hypothesis corresponding to the absence of a causal relationship, and the alternate hypothesis to its presence. The system model is described in detail in Section II. Following this, the main objective of this paper is to evaluate the detection performance of GCI for finite samples of the two time series of interest. Our contributions in this direction are as follows:

1) We derive the effects of down-sampling on the causal relationship between two signals. (See Section III.)
2) We derive the GCI for the down-sampled time series in the presence of additive noise assuming knowledge of the second-order statistics of the signals of interest. (See Section IV.)
3) We derive the statistics of the GCI under the two hypotheses for finite number of samples, and use these to derive the probabilities of detection and false alarm. Our analysis accounts for the estimation of the second-order statistics of the two signals using a finite number of samples when the generative model is unknown. (See Section V.)
4) Based on the statistics of GCI, we propose two alternative test statistics to detect Granger causality and derive the corresponding probabilities of detection and false alarm. (See Section VI.)
5) Using detailed simulations and numerical experiments based on real-world data, we validate the derived theory and compare the performance of the two proposed test statistics against the GCI. (See Section VII.)

The results derived in this paper can be used to characterize the behavior of GCI and related detectors under different physical conditions. In addition to this, the two alternative causality detectors discussed in this work are easy-to-use replacements for the GCI. We next describe the system model considered in this work.

Notation: Boldface lowercase and uppercase letters represent vectors and matrices, respectively. The kth column of $\mathbf{A}$ is denoted by $\mathbf{a}_k$. $(.)^H$ represents the Hermitian operation on a vector or a matrix, and $(.)^\dagger$ represents the Moore-Penrose pseudoinverse of a matrix. $\mathbf{I}_K, \mathbf{0}_K$ and $\mathbf{O}_K$ represent the identity matrix, the all-zero vector and the all-zero matrix of dimension K, respectively. The ℓ_2 norm of a vector and the Frobenius norm of a matrix are denoted by $\|\cdot\|_2$ and $\|\cdot\|_F$, respectively. $\mathcal{CN}(\mu, \sigma^2)$ represents a circularly symmetric complex Gaussian random variable with mean μ and variance σ^2. $E[\cdot]$ and var$(\cdot)$ represent the mean and variance of a random variable. In general, $\hat{x}$ and $\tilde{x}$ denote the MMSE estimate and the corresponding estimation error of a random variable x. Table I introduces the different symbols used in this paper.

II. System Model

We consider two complex valued discrete time zero mean Gaussian random processes $c[n]$ and $d[n]$ having variances σ_c^2 and σ_d^2, respectively, and expressed in the form of a bivariate

TABLE I: A list of notation followed in the paper.

$\mathbf{R}_{xy,p}[\tau]$	The $p \times p$ correlation matrix between two random vectors $\mathbf{x}[n]$ and $\mathbf{y}[n-\tau]$ of length p each.
$\mathbf{r}_{xx,p}[\tau]$	The $p \times 1$ correlation vector between the scalar $\mathbf{x}[n]$ and its p past values starting from $\mathbf{x}[n-\tau]$.
$\mathbf{r}_{xy}[\tau]$	The correlation between the scalars $\mathbf{x}[n]$ and $\mathbf{y}[n-\tau]$.
σ_x^2	Variance of the WSS random process $x[n]$.
$d[n]$	A noiseless random process sampled at the Nyquist rate.
$c[n]$	A noiseless random process, potentially caused by $d[n]$, sampled at the Nyquist rate.
q	The model order for the VAR model relating $c[n]$ and $d[n]$.
$\mathbf{A}$	The weight matrix relating $c[n]$ and $d[n]$.
$u[n]$	A downsampled version of $c[n]$.
$v[n]$	A downsampled version of $d[n]$.
p	The model order for the VAR model relating $u[n]$ and $v[n]$.
$\mathbf{B}$	The weight matrix relating $u[n]$ and $v[n]$
$x[n]$	A noisy observation of $u[n]$.
$y[n]$	A noisy observation of $v[n]$.
$\mathbf{z}_p[n]$	The concatenated observation vector of length $2p$.
$\mathbf{w}$	The weight vector relating $x[n]$ and $\mathbf{z}_p[n]$.
$\mathbf{h}$	The weight vector relating $x[n]$ and $\mathbf{x}_p[n]$.

VAR(q) model as

$$c[n] = \sum_{k=1}^{q} a_{1,k}^* c[n-k] + \sum_{k=1}^{q} a_{2,k}^* d[n-k] + \eta_c[n],$$
$$d[n] = \sum_{k=1}^{q} a_{3,k}^* c[n-k] + \sum_{k=1}^{q} a_{4,k}^* d[n-k] + \eta_d[n], \quad (1)$$

with $a_{i,k}$, $i \in \{1,2,3,4\}$ being the regression coefficients, and $\eta_c[n]$ and $\eta_d[n]$ the innovation components in $c[n]$ and $d[n]$, respectively. We assume that $c[n]$ and $d[n]$ are sampled at the Nyquist rate, and the temporally white innovation process vector $\boldsymbol{\eta}[n] = [\eta_c[n], \eta_d[n]]^T$ is distributed as

$$\boldsymbol{\eta}[n] \sim \mathcal{CN}\left(0, \begin{bmatrix} \sigma_{\eta_c}^2 & 0 \\ 0 & \sigma_{\eta_d}^2 \end{bmatrix}\right).$$

For simplicity, we assume that only a unidirectional coupling from $d[n]$ to $c[n]$ can exist, i.e., $a_{3,k} = 0$, for $1 \le k \le q$.

Defining $\mathbf{c}_q[n] \triangleq [c[n], c[n-1], \ldots, c[n-q+1]]^T$, $\mathbf{d}_q[n] \triangleq [d[n], d[n-1], \ldots, d[n-q+1]]^T$, $\mathbf{a}_i \triangleq [a_{i,1}, a_{i,2}, \ldots, a_{i,q}]^T$, $i \in \{1,2,3,4\}$, we can write (1) as,

$$\begin{bmatrix} c[n] \\ d[n] \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^H & \mathbf{a}_2^H \\ \mathbf{0}_q^H & \mathbf{a}_4^H \end{bmatrix} \begin{bmatrix} \mathbf{c}_q[n-1] \\ \mathbf{d}_q[n-1] \end{bmatrix} + \begin{bmatrix} \eta_c[n] \\ \eta_d[n] \end{bmatrix}, \quad (2)$$

with $\mathbf{0}_q$ being the q dimensional all zero vector. Defining $r_{cc}[\tau] \triangleq E[c[n]c^*[n-\tau]]$, $r_{cd}[\tau] \triangleq E[c[n]d^*[n-\tau]]$, and similarly $r_{dc}[\tau]$ and $r_{dd}[\tau]$, we can write

$$r_{cc}[\tau] = \sum_{k=1}^{q} a_{1,k}^* r_{cc}[\tau-k] + \sum_{k=1}^{q} a_{2,k}^* r_{dc}[\tau-k] + \sigma_{\eta_c}^2 \delta[\tau]. \quad (3)$$

Letting $\mathbf{r}_{cd,q}[\tau] = [r_{cd}[\tau], \ldots, r_{cd}[\tau-q+1]]^T$, and similarly $\mathbf{r}_{cc,q}[\tau]$, etc., the above becomes

$$r_{cc}[\tau] = \mathbf{a}_1^H \mathbf{r}_{cc,q}[\tau-1] + \mathbf{a}_2^H \mathbf{r}_{dc,q}[\tau-1] + \sigma_{\eta_c}^2 \delta[\tau]. \quad (4)$$

We can similarly write

$$r_{cd}[\tau] = \mathbf{a}_1^H \mathbf{r}_{cd,q}[\tau-1] + \mathbf{a}_2^H \mathbf{r}_{dd,q}[\tau-1], \quad (5)$$

and

$$r_{dd}[\tau] = \mathbf{a}_4^H \mathbf{r}_{dd,q}[\tau-1] + \sigma_{\eta_d}^2 \delta[\tau]. \quad (6)$$

The above equations, along with the information that $r_{cc}[0] = \sigma_c^2$, and $r_{dd}[0] = \sigma_d^2$ can be used to compute $r_{cc}[\tau]$, $r_{cd}[\tau]$, and $r_{dd}[\tau]$ for different values of τ.

Defining

$$\mathbf{R}_{cc,q}[\tau] \triangleq E\left[\mathbf{c}_q[n]\mathbf{c}_q^H[n-\tau]\right], \quad (7)$$

$$\mathbf{R}_{cd,q}[\tau] \triangleq E\left[\mathbf{c}_q[n]\mathbf{d}_q^H[n-\tau]\right], \quad (8)$$

and similarly defining $\mathbf{R}_{dc,p}[\tau]$ and $\mathbf{R}_{dd,q}[\tau]$, it can be shown using the Weiner-Hopf equations [35] that

$$\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{cc,q}[0] & \mathbf{R}_{cd,q}[0] \\ \mathbf{R}_{dc,q}[0] & \mathbf{R}_{dd,q}[0] \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{r}_{cc,q}[1] \\ \mathbf{r}_{cd,q}[1] \end{bmatrix}. \quad (9)$$

We can therefore calculate the regression coefficients from the correlation coefficients when the latter are not known. The interested reader is referred to [35, Chapter 2] for a detailed derivation of the Wiener-Hopf equations. We can then use these to compute the estimation error variance as

$$\sigma_{\eta_c}^2 = \sigma_c^2 - \begin{bmatrix} \mathbf{r}_{cc,q}[1] \\ \mathbf{r}_{cd,q}[1] \end{bmatrix}^H \begin{bmatrix} \mathbf{R}_{cc,q}[0] & \mathbf{R}_{cd,q}[0] \\ \mathbf{R}_{dc,q}[0] & \mathbf{R}_{dd,q}[0] \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{r}_{cc,q}[1] \\ \mathbf{r}_{cd,q}[1] \end{bmatrix}. \quad (10)$$

Also, it can be observed from (2) that the causal dependence of $c[n]$ on $d[n]$ is determined by the coefficient vector $\mathbf{a}_2$, that is, there exists a causal relationship only if $\mathbf{a}_2 \neq \mathbf{0}$.

Now, in the absence of a causal relationship, $c[n]$ can alternatively be expressed using a univariate AR(q) model,

$$c[n] = \mathbf{g}^H \mathbf{c}_q[n-1] + \zeta[n], \quad (11)$$

with $\zeta[n]$ being the zero mean temporally white complex Gaussian innovation component having a variance σ_ζ^2. Consequently,

$$r_{cc}[\tau] = \mathbf{g}^H \mathbf{r}_{cc,q}[\tau-1] + \sigma_\zeta^2 \delta[\tau], \quad (12)$$

$$\mathbf{g} = \mathbf{R}_{cc,q}^{-1}[0]\mathbf{r}_{cc,q}[1], \quad (13)$$

and

$$\sigma_\zeta^2 = \sigma_c^2 - \mathbf{r}_{cc,q}^H[1]\mathbf{R}_{cc,q}^{-1}[0]\mathbf{r}_{cc,q}[1]. \quad (14)$$

It is to be noted that in the presence of a causal relationship between $c[n]$ and $d[n]$ the mean squared prediction error for $c[n]$ using the bivariate model, $\sigma_{\eta_c}^2$, will be smaller than the mean

squared prediction error using the univariate signal model, σ_ζ^2, whereas, in the absence of such a causal relationship between them, $\sigma_\zeta^2 = \sigma_{\eta_c}^2$.

The GCI uses this property to quantify the causal relationship, and is defined as [2]

$$T_G \triangleq \log\left(\frac{\sigma_\zeta^2}{\sigma_{\eta_c}^2}\right). \tag{15}$$

In Granger causality, we say that a causal relationship exists between $c[n]$ and $d[n]$ if $T_G > 0$, and no causal relationship exists if $T_G = 0$.

However, both $c[n]$ and $d[n]$ are assumed to be noiseless and sampled at the Nyquist rate. Moreover, it also assumes that the exact second-order statistics for both $c[n]$ and $d[n]$ are known. However, these assumptions do not hold in a practical data acquisition system. Therefore, in the subsequent sections, we introduce these imperfections, viz. down-sampling, measurement noise, and finite sample effects, into a GCI based detector, and analyze their impact on the detection performance. In the next section, we discuss the effect of down-sampling on the GCI.

III. The Effect of Down-sampling on GCI

Let $u[n]$ and $v[n]$ respectively be versions of $c[n]$ and $d[n]$ down-sampled by a factor M, such that,

$$u[n] = c[nM - l], v[n] = d[nM - l]. \tag{16}$$

Since we use the second-order statistics of the observed wide sense stationary random processes for detection, we can set the offset parameter l to zero without loss of generality. We can now express $u[n]$ and $v[n]$ in the form of a bivariate linear regression process of order p as

$$\begin{bmatrix} u[n] \\ v[n] \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1^H & \mathbf{b}_2^H \\ \mathbf{0}_p^H & \mathbf{b}_4^H \end{bmatrix} \begin{bmatrix} \mathbf{u}_p[n-1] \\ \mathbf{v}_p[n-1] \end{bmatrix} + \begin{bmatrix} \epsilon_u[n] \\ \epsilon_v[n] \end{bmatrix}, \tag{17}$$

with $\mathbf{b}_i, i \in \{1, 2, 4\}$ representing new regression coefficients, and $E\left[u[n-k]\epsilon_u^*[n]\right] = 0$, $E\left[v[n-k]\epsilon_v^*[n]\right] = 0$, $E\left[u[n-k]\epsilon_v^*[n]\right] = 0$, and $E\left[v[n-k]\epsilon_u^*[n]\right] = 0$ for $k > 0$. It is important to note that the model order p in this case may be different from the model order q considered in the previous section due to down-sampling.

Since the GCI is a function of the second-order statistics of $c[n]$ and $d[n]$, we need to derive the second-order statistics of $u[n]$ and $v[n]$ in terms of $c[n]$ and $d[n]$ in order to quantify the effect of down-sampling on the GCI.

Let $r_{uv}[\tau] = E[u[n]v^*[n-\tau]]$, $\mathbf{r}_{uv,p}[\tau] = E[\mathbf{u}_p[n]v^*[n-\tau]]$, and $\mathbf{R}_{uv,p}[\tau] = E[\mathbf{u}_p[n]\mathbf{v}_p^H[n-\tau]]$, with the absence of the index $[\tau]$ indicating $\tau = 0$. We can write, $r_{uu}[\tau] = E[c[nM]c^*[nM - \tau M]] = r_{cc}[\tau M]$. Similarly, $\mathbf{r}_{uu,p}[\tau] = [r_{cc}[\tau M], r_{cc}[(\tau+1)M], \ldots, r_{cc}[(\tau+p)M]]^T$, and

$$r_{uu}[\tau] = \mathbf{b}_1^H \mathbf{r}_{uu,p}[\tau - 1] + \mathbf{b}_2^H \mathbf{r}_{vu,p}[\tau - 1] + \sigma_{\eta_c}^2 \delta[\tau], \tag{18}$$

$$r_{uv}[\tau] = \mathbf{b}_1^H \mathbf{r}_{uv,p}[\tau - 1] + \mathbf{b}_2^H \mathbf{r}_{vv,p}[\tau - 1], \tag{19}$$

$$r_{vv}[\tau] = \mathbf{b}_4^H \mathbf{r}_{vv,p}[\tau - 1] + \sigma_{\eta_d}^2 \delta[\tau], \tag{20}$$

and

$$\mathbf{R}_{uu,p} = \begin{bmatrix} r_{cc}[0] & r_{cc}[M] & \ldots & r_{cc}[(p-1)M] \\ r_{cc}[M] & r_{cc}[0] & \ldots & r_{cc}[(p-2)M] \\ \vdots & \vdots & \ddots & \vdots \\ r_{cc}[(p-1)M] & r_{cc}[(p-2)M] & \ldots & r_{cc}[0] \end{bmatrix}. \tag{21}$$

The weight vectors $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_4$ can then be determined via the Wiener-Hopf equations [35] as

$$\begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{uu,p} & \mathbf{R}_{uv,p} \\ \mathbf{R}_{vu,p} & \mathbf{R}_{vv,p} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{r}_{uu,p}[1] \\ \mathbf{r}_{uv,p}[1] \end{bmatrix}, \tag{22}$$

$$\text{and} \quad \mathbf{b}_4 = \mathbf{R}_{vv,p}^{-1} \mathbf{r}_{vv}[1]. \tag{23}$$

We can then write,

$$\epsilon_u[n] = u[n] - \mathbf{b}_1^H \mathbf{u}_p[n-1] - \mathbf{b}_2^H \mathbf{v}_p[n-1], \tag{24}$$

$$\epsilon_v[n] = y[n] - \mathbf{b}_4^H \mathbf{v}_p[n-1]. \tag{25}$$

Consequently,

$$r_{\epsilon_u\epsilon_u}[0] = \sigma_c^2 - \begin{bmatrix} \mathbf{r}_{uu,p}[1] \\ \mathbf{r}_{uv,p}[1] \end{bmatrix}^H \begin{bmatrix} \mathbf{R}_{uu,p} & \mathbf{R}_{uv,p} \\ \mathbf{R}_{vu,p} & \mathbf{R}_{vv,p} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{r}_{uu,p}[1] \\ \mathbf{r}_{uv,p}[1] \end{bmatrix}, \tag{26}$$

$$r_{\epsilon_v\epsilon_v}[0] = \sigma_c^2 - \mathbf{r}_{vv,p}^H[1] \mathbf{R}_{vv,p}^{-1} \mathbf{r}_{vv,p}[1]. \tag{27}$$

$$\begin{aligned} r_{\epsilon_u\epsilon_v}[0] = r_{cd}[0] &- \left[\mathbf{r}_{uu,p}^H[1]\, \mathbf{r}_{uv,p}^H[1]\right] \begin{bmatrix} \mathbf{R}_{uu,p} & \mathbf{R}_{uv,p} \\ \mathbf{R}_{vu,p} & \mathbf{R}_{vv,p} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{r}_{uv,p}[1] \\ \mathbf{r}_{vv,p}[1] \end{bmatrix} \\ &- \mathbf{r}_{uv,p}^H[1] \mathbf{R}_{vv,p}^{-1} \mathbf{r}_{vv,p}[1] + \begin{bmatrix} \mathbf{r}_{uu,p}[1] \\ \mathbf{r}_{vv,p}[1] \end{bmatrix}^H \\ &\times \begin{bmatrix} \mathbf{R}_{uu,p} & \mathbf{R}_{uv,p} \\ \mathbf{R}_{vu,p} & \mathbf{R}_{vv,p} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{R}_{uv,p} \\ \mathbf{R}_{vv,p} \end{bmatrix} \mathbf{R}_{vv,p}^{-1} \mathbf{r}_{vv,p}[1]. \end{aligned} \tag{28}$$

Hence, the covariance matrix of the innovation process need not be diagonal after down-sampling.

Similarly, the single variable linear regression model for the down-sampled signal $u[n]$ can be written as

$$u[n] = \mathbf{f}^H \mathbf{u}_p[n-1] + \xi[n], \tag{29}$$

with $\xi[n]$ being the innovation component. Therefore, we get

$$\mathbf{f} = \mathbf{R}_{uu,p}^{-1} \mathbf{r}_{uu,p}[1], \tag{30}$$

$$\text{and} \quad r_{\xi\xi}[0] = \sigma_u^2 - \mathbf{r}_{uu,p}^H[1] \mathbf{R}_{uu,p}^{-1} \mathbf{r}_{uu,p}[1]. \tag{31}$$

A causal relationship between $u[n]$ and $v[n]$ can now be determined by considering the ratio $T_G = \log_2\left(\frac{r_{\xi\xi}[0]}{r_{\epsilon_u\epsilon_u}[0]}\right)$.

To illustrate the effect of down-sampling on the GCI, consider the simple special case: $c[n] = ad[n-1] + \eta_c[n]$, and $d[n] = \eta_d[n]$. Letting $M = 2$, we can write

$$u[n] = c[2n] = a\eta_d[2n-1] + \eta_c[2n], \tag{32}$$

$$v[n] = d[2n] = \eta_d[2n]. \tag{33}$$

Consequently, $r_{uu}[1] = r_{cc}[2] = 0$, $r_{uv}[1] = r_{cd}[2] = 0$, $r_{uv}[0] = r_{cd}[0] = 0$, and $r_{\xi\xi}[0] = r_{\epsilon_x\epsilon_x}[0] = \sigma_c^2$. We see that there exists a causal relationship between $c[n]$ and $d[n]$ for any $a \neq 0$, but no causal relationship exists between $u[n]$ and $v[n]$. Thus, down-sampling can suppress an existing causal relationship between two signals. However, the degree of suppression of the causal relationship depends on the structure of the VAR model followed by the signals. We next analyze the effect of additive noise on the performance of the GCI.

IV. GCI with Downsampling and Additive Noise

Let us define $x[n]$ and $y[n]$ as the down-sampled signals $u[n]$ and $v[n]$ corrupted by AWGN, such that,

$$\begin{aligned} x[n] &= u[n] + \nu_x[n], \\ y[n] &= v[n] + \nu_y[n], \end{aligned} \tag{34}$$

where $\nu_x[n] \sim \mathcal{CN}(0, \sigma_{\nu_x}^2)$ and $\nu_y[n] \sim \mathcal{CN}(0, \sigma_{\nu_y}^2)$.

Using (17), we can write,

$$\begin{bmatrix} x[n] \\ y[n] \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1^H & \mathbf{b}_2^H \\ \mathbf{0}_p^H & \mathbf{b}_4^H \end{bmatrix} \begin{bmatrix} \mathbf{u}_p[n-1] \\ \mathbf{v}_p[n-1] \end{bmatrix} + \begin{bmatrix} \epsilon_u[n] + \nu_x[n] \\ \epsilon_v[n] + \nu_y[n] \end{bmatrix}. \tag{35}$$

If a causal relationship exists between $c[n]$ and $d[n]$, and is preserved in $u[n]$ and $v[n]$ after down-sampling, it should also exist between $x[n]$ and $y[n]$, and therefore, the past samples of $x[n]$ and $y[n]$ can be used to predict $x[n]$ better than its prediction by using the past samples of $x[n]$ alone. Letting $\mathbf{z}[n] = [x[n]\ y[n]]^T$ and $\mathbf{z}_p[n] = [\mathbf{x}_p^T[n]\ \mathbf{y}_p^T[n]]^T$, we can express $x[n]$ as

$$x[n] = \mathbf{w}^H \mathbf{z}_p[n-1] + \varphi_2[n], \tag{36}$$

with $\varphi_2[n]$ being the prediction error for the bivariate VAR model, and $\mathbf{w}$ being the weight vector minimizing the mean squared value of $\varphi_2[n]$.

Defining $\sigma_{\varphi_2}^2 \triangleq E[|\varphi_2[n]|^2] = E[|x[n] - \mathbf{w}^H \mathbf{z}_p[n-1]|^2]$, it can be shown that [35],

$$\sigma_{\varphi_2}^2 = \sigma_x^2 - \mathbf{w}^H \mathbf{r}_{zx,p}[1] - \mathbf{r}_{zx,p}^H[1]\mathbf{w} + \mathbf{w}^H \mathbf{R}_{zz,p}\mathbf{w}, \tag{37}$$

where

$$\sigma_x^2 = E[x[n]x^*[n]] = \sigma_c^2 + \sigma_{\nu_x}^2, \tag{38}$$

$$\mathbf{r}_{zx,p}[\tau] = E[\mathbf{z}_p[n]x^*[n-\tau]] = [\mathbf{r}_{xx}^H[\tau]\ \mathbf{r}_{yx}^H[\tau]]^H, \tag{39}$$

and

$$\mathbf{R}_{zz,p} = E\left[\mathbf{z}_p[n]\mathbf{z}_p^H[n]\right]. \tag{40}$$

Since the additive noise is independent of both $u[n]$ and $v[n]$, it can be shown that

$$r_{xx}[\tau] = r_{uu}[\tau] + \sigma_{\nu_x}^2\delta[\tau], \tag{41}$$

$$r_{xy}[\tau] = r_{uv}[\tau], \tag{42}$$

$$r_{yy}[\tau] = r_{vv}[\tau] + \sigma_{\nu_y}^2\delta[\tau]. \tag{43}$$

Substituting equations (18)-(20) into (41)-(43) we obtain

$$r_{xx}[\tau] = \mathbf{b}_1^H \mathbf{r}_{uu,p}[\tau-1] + \mathbf{b}_2^H \mathbf{r}_{vu,p}[\tau-1] + (\sigma_{\epsilon_u}^2 + \sigma_{\nu_x}^2)\delta[\tau], \tag{44}$$

$$r_{xy}[\tau] = \mathbf{b}_1^H \mathbf{r}_{uv,p}[\tau-1] + \mathbf{b}_2^H \mathbf{r}_{vv,p}[\tau-1], \tag{45}$$

$$r_{yy}[\tau] = \mathbf{b}_4^H \mathbf{r}_{vv,p}[\tau-1] + (\sigma_{\epsilon_v}^2 + \sigma_{\nu_y}^2)\delta[\tau], \tag{46}$$

and similarly,

$$\begin{aligned} \mathbf{R}_{zz,p} &= \begin{bmatrix} \mathbf{R}_{xx,p} & \mathbf{R}_{xy,p} \\ \mathbf{R}_{yx,p} & \mathbf{R}_{yy,p} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{R}_{uu,p} & \mathbf{R}_{uv,p} \\ \mathbf{R}_{vu,p} & \mathbf{R}_{vv,p} \end{bmatrix} + \begin{bmatrix} \sigma_{\nu_x}^2\mathbf{I}_p & \mathbf{0}_p \\ \mathbf{0}_p & \sigma_{\nu_y}^2\mathbf{I}_p \end{bmatrix}. \end{aligned} \tag{47}$$

The optimal weight vector minimizing $\sigma_{\varphi_2}^2$ can be obtained via the Wiener-Hopf equations as

$$\mathbf{w} = \mathbf{R}_{zz,p}^{-1}\mathbf{r}_{zx,p}[1]. \tag{48}$$

The minimized value of $\sigma_{\varphi_2}^2$ is

$$\sigma_{\varphi_2}^2 = \sigma_x^2 - \mathbf{r}_{zx,p}^H[1]\mathbf{R}_{zz,p}^{-1}\mathbf{r}_{zx,p}[1]. \tag{49}$$

However, in case no causal relationship exists between $c[n]$ and $d[n]$, and consequently between $x[n]$ and $y[n]$, $x[n]$ can be expressed using the single variable AR model as

$$x[n] = \mathbf{h}^H \mathbf{x}_p[n-1] + \varphi_1[n], \tag{50}$$

with $\varphi_1[n]$ being the single variable prediction error, and $\mathbf{h}$ being the optimal weight vector, computed as

$$\mathbf{h} = \mathbf{R}_{xx,p}^{-1}\mathbf{r}_{xx,p}[1]. \tag{51}$$

Similar to the two variable case, the minimized value of the prediction error takes the form

$$\sigma_{\varphi_1}^2 = \sigma_x^2 - \mathbf{r}_{xx,p}[1]^H\mathbf{R}_{xx,p}^{-1}\mathbf{r}_{xx,p}[1]. \tag{52}$$

The GCI therefore becomes

$$T_G = \log_2\left(\frac{\sigma_{\varphi_1}^2}{\sigma_{\varphi_2}^2}\right) = \log_2\left(\frac{\sigma_x^2 - \mathbf{r}_{xx,p}^H[1]\mathbf{R}_{xx,p}^{-1}\mathbf{r}_{xx,p}[1]}{\sigma_x^2 - \mathbf{r}_{zx,p}^H[1]\mathbf{R}_{zz,p}^{-1}\mathbf{r}_{zx,p}[1]}\right). \tag{53}$$

To demonstrate the effects of noise on the GCI, we again consider our running example, $d[n] = \epsilon_v[n]$, $c[n] = u[n] = ad[n-1] + bd[n-2] + \epsilon_c[n]$. Downsampling these by a factor $M = 2$, we obtain, $v[n] = \epsilon_v[n]$, $u[n] = bv[n-1] + \epsilon_u[n]$, with, $\sigma_v^2 = \sigma_{\epsilon_v}^2$, and $\sigma_u^2 = |b|^2\sigma_{\epsilon_v}^2 + \sigma_{\epsilon_u}^2$, and consequently,

$$\sigma_x^2 = |b|^2\sigma_{\epsilon_v}^2 + \sigma_{\epsilon_u}^2 + \sigma_{\nu_x}^2, \quad \sigma_y^2 = \sigma_{\epsilon_v}^2 + \sigma_{\nu_y}^2, \tag{54}$$

$$r_{xy}[0] = 0, \quad r_{xy}[1] = b\sigma_{\epsilon_v}^2. \tag{55}$$

Therefore,

$$\mathbf{R}_{zz,1} = \begin{bmatrix} |b|^2\sigma_{\epsilon_v}^2 + \sigma_{\epsilon_u}^2 + \sigma_{\nu_x}^2 & 0 \\ 0 & \sigma_{\epsilon_v}^2 + \sigma_{\nu_y}^2 \end{bmatrix}, \tag{56}$$

$$\sigma_{\varphi_2}^2 = \sigma_{\epsilon_u}^2 + \sigma_{\nu_x}^2 + \frac{|b|^2\sigma_{\epsilon_v}^2\sigma_{\nu_y}^2}{\sigma_{\epsilon_v}^2 + \sigma_{\nu_y}^2}, \tag{57}$$

and

$$\sigma_{\varphi_1}^2 = \sigma_x^2 = |b|^2\sigma_{\epsilon_v}^2 + \sigma_{\epsilon_u}^2 + \sigma_{\nu_x}^2. \tag{58}$$

We can simplify the GCI as

$$T_G = \log_2\left(\frac{|b|^2\sigma_{\epsilon_v}^2 + \sigma_{\epsilon_u}^2 + \sigma_{\nu_x}^2}{|b|^2\sigma_{\epsilon_v}^2\left(\frac{\sigma_{\nu_y}^2}{\sigma_{\epsilon_v}^2 + \sigma_{\nu_y}^2}\right) + \sigma_{\epsilon_u}^2 + \sigma_{\nu_x}^2}\right). \tag{59}$$

In the absence of additive noise in both $x[n]$ and $y[n]$,

T_G reduces to $T_G = \log_2\left(\frac{|b|^2\sigma^2_{\epsilon_y}+\sigma^2_{\epsilon_u}}{\sigma^2_{\epsilon_u}}\right)$, as in the previous section. However, as either $\sigma^2_{\nu_x}$, or $\sigma^2_{\nu_y}$ get large, the second term in the expression for T_G converges to zero. Thus, additive noise can also lead to the suppression of a causal relationship between two signals.

The discussion till now assumed that the second-order statistics of both the signals of interest are available at the detector. However, in practice, we have to estimate the second-order statistics using a finite number of samples of $x[n]$ and $y[n]$. The effects of the use of estimated second-order statistics on the detection of Granger causality in an under-sampled noisy environment are discussed in the next section.

V. Finite Sample Effects on the GCI

In case only N samples of each of $x[n]$ and $y[n]$ are observed, and the generative model is unknown, we need to estimate the optimal predictors for the single and two variable cases, the corresponding prediction error variances, and the GCI or an equivalent test statistic using these samples. In this section, we first discuss the properties of the optimal predictor weights for the one variable and two variable prediction models. Then, we use these weights to calculate the statistics of the finite sample estimates of the corresponding mean square prediction errors, and finally derive expressions for the performance of GCI with finite number of observations.

A. Properties of Optimal Predictor Weights

It is known that given a finite number of observations, the least squares estimate is the best linear unbiased estimator for $x[n]$ [35]. Letting $x[1],\ldots,x[N]$ be the N available samples of $x[n]$, $y[1],\ldots,y[N]$ be the N available samples of $y[n]$, and $\hat{x}[n] = \hat{\mathbf{w}}^H\mathbf{z}_p[n-1]$ be the least squares (LS) estimate of $x[n]$ generated using the two variable model, we can write

$$\mathbf{x}_{N-p}[N] = \hat{\mathbf{x}}_{N-p}[N] + \tilde{\mathbf{x}}_{N-p}[N], \tag{60}$$

where $\tilde{\mathbf{x}}_{N-p}[N]$ is the error term orthogonal to the measurement space. Also, $\hat{\mathbf{w}}$ is the least squares estimate of the weight vector, given as [35]

$$\hat{\mathbf{w}} = \mathbf{Z}_p^{\dagger}[N]\mathbf{x}_{N-p}[N]. \tag{61}$$

Here, $\hat{\mathbf{w}}$ is a finite sample estimate of the true weight vector $\mathbf{w}$ with $\mathbf{Z}_p[N] = [\mathbf{z}_p[N-1],\mathbf{z}_p[N-2],\ldots,\mathbf{z}_p[p]]^H$, and $()^{\dagger}$ represents the Moore-Penrose inverse of a matrix. Since the innovation process and the measurement noise are white and zero mean, it can be shown that $\hat{\mathbf{w}}$ has the following properties [35]:

1) $\hat{\mathbf{w}}$ is a an unbiased estimator of $\mathbf{w}$.
2) The covariance matrix of $\hat{\mathbf{w}}$ is $\text{cov}(\hat{\mathbf{w}}) = \frac{\sigma^2_{\varphi_2}}{N-p}\mathbf{R}^{-1}_{zz,p}$.

Therefore, $\hat{\mathbf{w}} = \mathbf{w} + \tilde{\mathbf{w}}$, with $\tilde{\mathbf{w}} \sim \mathcal{CN}\left(\mathbf{0}, \frac{\sigma^2_{\varphi_2}}{N-p}\mathbf{R}^{-1}_{zz,p}\right)$.

Similarly, considering the single variable prediction model, we can write $\mathbf{x}_{N-p}[N]$ in terms of its projection, $\bar{\mathbf{x}}_{N-p}[N]$ on the data matrix, , $\mathbf{X}_p[N] = [\mathbf{x}_p[N-1],\mathbf{x}_p[N-2],\ldots,\mathbf{x}_p[p]]^H$, and the error component $\dot{\mathbf{x}}_{N-p}[N] = \mathbf{x}_{N-p}[N] - \bar{\mathbf{x}}_{N-p}[N]$. Again, $\bar{\mathbf{x}}_{N-p}[N] = \mathbf{X}_p[N]\hat{\mathbf{h}}$ with $\hat{\mathbf{h}} = \mathbf{X}_p^{\dagger}[N]\mathbf{x}_{N-p}[N]$, and

$$\hat{\mathbf{h}} \sim \mathcal{CN}\left(\mathbf{h}, \frac{\sigma^2_{\varphi_1}}{N-p}\mathbf{R}^{-1}_{xx,p}\right). \tag{62}$$

We next use the above properties of the estimated regression weights to derive the statistics of the finite sample estimate of the mean squared error for the two prediction models.

B. Mean Squared Prediction Errors

First, considering the prediction error for the two variable model, we can express $\tilde{x}[n]$ as

$$\begin{aligned}\tilde{x}[n] &= x[n] - \hat{x}[n] \\ &= x[n] - \mathbf{w}^H\mathbf{z}_K[n-1] - \tilde{\mathbf{w}}^H\mathbf{z}_K[n-1].\end{aligned} \tag{63}$$

Since the innovation process of $u[n]$ and the additive noise are white, $\tilde{x}[n]$ is zero mean i.i.d. Gaussian with variance

$$\begin{aligned}E[|\tilde{x}[n]|^2] = \sigma^2_{\varphi_2} &+ E[\tilde{\mathbf{w}}^H\mathbf{z}_p[n-1]\mathbf{z}_p^H[n-1]\tilde{\mathbf{w}}] \\ &- 2\Re\{E[\tilde{\mathbf{w}}^H\mathbf{z}_p[n-1]x^*[n]]\} \\ &- 2\Re\{E[\tilde{\mathbf{w}}^H\mathbf{z}_p[n-1]\mathbf{z}_p^H[n-1]\mathbf{w}]\}.\end{aligned} \tag{64}$$

It is common in the adaptive filtering literature to assume the weight error vector to be independent of the regression vectors as well as the desired output [35]. Hence,

$$E[\tilde{\mathbf{w}}^H\mathbf{z}_p[n-1]x^*[n]] = E[\tilde{\mathbf{w}}^H\mathbf{z}_p[n-1]\mathbf{z}_p^H[n-1]\mathbf{w}] = 0, \tag{65}$$

and

$$\begin{aligned}&E[\tilde{\mathbf{w}}^H\mathbf{z}_p[n-1]\mathbf{z}_p^H[n-1]\tilde{\mathbf{w}}] \\ &\quad= \text{Tr}(E[\mathbf{z}_p[n-1]\mathbf{z}_p^H[n-1]E[\tilde{\mathbf{w}}\tilde{\mathbf{w}}^H]) \\ &\quad= \text{Tr}\left(\mathbf{R}_{zz,p}\frac{\sigma^2_{\varphi_2}}{N-p}\mathbf{R}^{-1}_{zz,p}\right) = \sigma^2_{\varphi_2}\frac{2p}{N-p}.\end{aligned} \tag{66}$$

Substituting these in (64), we can write

$$E[|\tilde{x}[n]|^2] = \sigma^2_{\varphi_2}\left(1 + \frac{2p}{N-p}\right). \tag{67}$$

We can similarly argue, for the one variable model, that $\dot{x}[n]$ is also zero mean i.i.d. Gaussian distributed, with

$$E[|\dot{x}[n]|^2] = \sigma^2_{\varphi_1}\left(1 + \frac{p}{N-p}\right). \tag{68}$$

Thus, the estimates of the prediction error variances from the one variable and two variable regression models of the system can be expressed as

$$t_1 = \frac{1}{N-p}\sum_{n=p+1}^{N}|\dot{x}[n]|^2, \tag{69}$$

$$t_2 = \frac{1}{N-p}\sum_{n=p+1}^{N}|\tilde{x}[n]|^2. \tag{70}$$

These can now be used to calculate the GCI as

$$T_G = \log_2\left(\frac{t_1}{t_2}\right), \tag{71}$$

However, since T_G is a logarithm of ratios of random variables, its statistics become hard to determine. Instead, we define $T_E \triangleq 2^{-T_G} = \frac{t_2}{t_1}$ as a test statistic, for simplicity of analysis.

Since $\tilde{x}[n]$ and $\dot{x}[n]$ are the measurement errors for the same process, t_1 and t_2 cannot be considered independent. However, since both t_1 and t_2 are sums of a large number of i.i.d. random variables, we can use the central limit theorem to approximate these as Gaussian r.v.s [36]. Note that this approximation is valid when the number of samples N is much larger than the model order p. Now

$$E[t_2] = \sigma^2_{\varphi_2}\left(1 + \frac{2p}{N-p}\right), \tag{72}$$

and

$$\text{var}(t_2) = \frac{1}{N-p}\sigma^4_{\varphi_2}\left(1 + \frac{2p}{N-p}\right)^2. \tag{73}$$

Similarly,

$$E[t_1] = \sigma^2_{\varphi_1}\left(1 + \frac{p}{N-p}\right). \tag{74}$$

However, in the presence of a causal relationship between $y[n]$ and $x[n]$, $\mathbf{w}^H\mathbf{z}_p[n-1]$ will be a better estimate of $x[n]$ than $\mathbf{h}^H\mathbf{x}_p[n-1]$, therefore, $\mathbf{w}^H\mathbf{z}_p[n-1]$ can be written as $\mathbf{w}^H\mathbf{z}_p[n-1] = \mathbf{h}^H\mathbf{x}_p[n-1] + \acute{x}[n]$, such that, $\acute{x}[n]$ is orthogonal to $\mathbf{h}^H\mathbf{x}_p[n-1]$. Consequently,

$$\varphi_1[n] = \varphi_2[n] + \acute{x}[n], \tag{75}$$

with $E[\acute{x}[n]\varphi_2^*[n]] = 0$, and

$$\sigma^2_{\varphi_1} = \sigma^2_{\varphi_2} + \sigma^2_{\acute{x}}, \tag{76}$$

where $\sigma^2_{\acute{x}} \triangleq \text{var}(\acute{x}^2)$.

Now, since $\acute{x}[n] = \mathbf{w}^H\mathbf{z}_p[n] - \mathbf{h}^H\mathbf{x}_p[n]$, it can be shown that $E[\acute{x}[n]] = 0$, and,

$$\begin{aligned} E[|\acute{x}[n]|^2] &= \mathbf{w}^H R_{zz,p}\mathbf{w} + \mathbf{h}^H R_{xx,p}\mathbf{h} - 2\Re\{\mathbf{w}^H\mathbf{R}_{zx}\mathbf{h}\} \\ &= \mathbf{r}^H_{zz,p}\mathbf{R}^{-1}_{zz,p}\mathbf{r}_{zz,p} + \mathbf{r}^H_{xx,p}\mathbf{R}^{-1}_{xx,p}\mathbf{r}_{xx,p} \\ &\quad -2\Re\{\mathbf{r}^H_{zz,p}\mathbf{R}^{-1}_{zz,p}\mathbf{R}_{zx,p}\mathbf{R}^{-1}_{xx,p}\mathbf{r}_{xx,p}\}. \end{aligned} \tag{77}$$

Therefore, $\dot{x}[n] = \tilde{x}[n] + \acute{x}[n] - \tilde{\mathbf{h}}^H\mathbf{x}_p[n-1]$ and

$$\begin{aligned} E[t_1] &= \left(\sigma^2_{\varphi_2} + \sigma^2_{\acute{x}}\right)\left(1 + \frac{p}{N-p}\right) \\ &= E[t_2] + \sigma^2_{\acute{x}_2}\left(1 + \frac{p}{N-p}\right) - \frac{\sigma^2_{\varphi_2}}{N-p}. \end{aligned} \tag{78}$$

Similarly,

$$E[t_1^2] = E^2[t_1] + \frac{1}{N-p}(\sigma^2_{\varphi_2} + \sigma^2_{\acute{x}})^2\left(1 + \frac{p}{N-p}\right)^2, \tag{79}$$

and,

$$\begin{aligned} \text{var}(t_1) &= \frac{1}{N-p}\left(\sigma^4_{\varphi_2} + \sigma^4_{\acute{x}} + 2\sigma^2_{\varphi_2}\sigma^2_{\acute{x}}\right) \\ &\quad \times\left(1 + \frac{p^2}{(N-p)^2} + \frac{2p}{N-p}\right) \\ &= \text{var}(t_2) + \frac{1}{N-p}\Bigg(\sigma^4_{\acute{x}} + 2\sigma^2_{\varphi_2}\sigma^2_{\acute{x}} \\ &\quad -\sigma^4_{\varphi_2}\left(\frac{3p^2}{(N-p)^2} + \frac{2p}{N-p}\right)\Bigg). \end{aligned} \tag{80}$$

Consequently, it can be shown that

$$\text{cov}(t_1, t_2) = \frac{\sigma^4_{\varphi_2}}{N-p}. \tag{81}$$

The GCI can therefore be approximated as the ratio of two correlated Gaussian r.v.s t_1 and t_2 whose statistics are derived above. In the next subsection, we use the above statistics to calculate the probabilities of detection and false alarm.

C. Performance of GCI

Now, we declare a causal relationship to exist between the two signals if the ratio $T_E = \frac{t_2}{t_1}$ is below a threshold λ. Since t_1 is the sum of non-negative terms, it is almost surely positive, and therefore

$$\Pr\{T_E < \lambda\} = \Pr\{\lambda t_1 - t_2 > 0\}. \tag{82}$$

Since t_1 and t_2 are approximated as correlated Gaussian r.v.s, therefore, from (78), $T_D \triangleq \lambda t_1 - t_2$ can also be approximated as a Gaussian r.v. with mean and variance

$$E[T_D] = (\lambda-1)E[t_2] + \lambda\sigma^2_{\acute{x}}\left(1 + \frac{p}{N-p}\right) - \lambda\sigma^2_{\varphi_2}\frac{p}{N-p}, \tag{83}$$

and

$$\text{var}(T_D) = \lambda^2\,\text{var}(t_1) + \text{var}(t_2) - 2\lambda\,\text{cov}(t_1, t_2), \tag{84}$$

respectively.

Now, under the null hypothesis, $\mathcal{H}_0$, $\sigma^2_{\varphi_1} = \sigma^2_{\varphi_2}$ and $\sigma^2_{\acute{x}|\mathcal{H}_0} = 0$. Hence, under $\mathcal{H}_0$, the above simplify to

$$\begin{aligned} \mu_{T_D,0} &\triangleq E[T_D|\mathcal{H}_0] = \\ &(\lambda-1)\sigma^2_{\varphi_2}\left(1 + \frac{p}{N-p}\right) - \sigma^2_{\varphi_2}\left(\frac{p}{N-p}\right), \end{aligned} \tag{85}$$

$$\begin{aligned} \sigma^2_{T_D,0} &\triangleq \text{var}(T_D|\mathcal{H}_0) \\ &= (\lambda^2+1)\frac{\sigma^4_{\varphi_2}}{N-p}\left(1 + \frac{2p}{N-p}\right)^2 \\ &\quad - \lambda^2\frac{\sigma^4_{\varphi_2}}{N-p}\left(\frac{3p^2}{(N-p)^2} + \frac{2p}{N-p}\right) - 2\lambda\frac{\sigma^4_{\varphi_2}}{N-p}. \end{aligned} \tag{86}$$

Similarly, under the alternate hypothesis, $\mathcal{H}_1$, $\sigma^2_{\varphi_1} > \sigma^2_{\varphi_2}$,

and therefore,

$$\mu_{T_D,1} \triangleq E[T_D|\mathcal{H}_1] = \lambda\sigma_{\acute{x}}^2\left(1+\frac{p}{N-p}\right) + (\lambda-1)\sigma_{\varphi_2}^2\left(1+\frac{p}{N-p}\right) - \sigma_{\varphi_2}^2\frac{p}{N-p}, \quad (87)$$

$$\sigma_{T_D,1}^2 \triangleq \text{var}(T_D|\mathcal{H}_1) = (\lambda^2+1)\frac{\sigma_{\varphi_2}^4}{N-p}\left(1+\frac{2p}{N-p}\right)^2 + \lambda^2\frac{\sigma_{\acute{x}}^4}{N-p}\left(1+\frac{p}{N-p}\right)^2 + 2\lambda^2\frac{\sigma_{\varphi_2}^2\sigma_{\acute{x}_2}^2}{N-p}\left(1+\frac{p}{N-p}\right)^2 - \lambda^2\frac{\sigma_{\varphi_2}^4}{N-p}\left(\frac{3p^2}{(N-p)^2}+\frac{2p}{N-p}\right) - 2\lambda\frac{\sigma_{\varphi_2}^4}{N-p}. \quad (88)$$

Based on these, the probability of false alarm can be calculated as

$$P_{\text{FA}} = \Pr\{T_D > 0|\mathcal{H}_0\} = Q\left(\frac{\mu_{T_D,0}}{\sigma_{T_D,0}}\right). \quad (89)$$

Similarly, the probability of correct detection of a causal relationship, P_D, can then be calculated as

$$P_{\text{D}} = \Pr\{T_D > 0|\mathcal{H}_1\} = Q\left(\frac{\mu_{T_D,1}}{\sigma_{T_D,1}}\right). \quad (90)$$

Using (89), a detector providing a given false alarm rate of P_{FA}^* can be designed by selecting a threshold λ that satisfies

$$\left[\sigma_{\varphi_2}^2(\lambda-1)\left(1+\frac{p}{N-p}\right) - \lambda\sigma_{\varphi_2}^2\frac{p}{N-p}\right]^2 = Q^{-1}(P_{FA}^*)\left[(\lambda^2+1)\frac{\sigma_{\varphi^2}^4}{N-p}\left(1+\frac{p}{N-p}\right)^2 - \lambda^2\frac{\sigma_{\varphi_2}^4}{N-p}\left(\frac{3p^2}{(N-p)^2}+\frac{2p}{N-p}\right) - 2\lambda\frac{\sigma_{\varphi_2}^4}{N-p}\right]. \quad (91)$$

The above reduces to a quadratic equation in λ that can be solved to obtain the detection threshold for the detector.

As an example, consider N samples of signals $u[n]$ and $v[n]$ corrupted by AWGN and observed as $x[n]$ and $y[n]$, respectively, such that $v[n] = \epsilon_v[n]$ and

$$u[n] = \begin{cases} \epsilon_{u,0}[n] & \mathcal{H}_0 \\ a\epsilon_v[n] + \epsilon_{u,1}[n] & \mathcal{H}_1 \end{cases}, \quad (92)$$

with $\text{var}(\epsilon_v) = \sigma_{\epsilon_v}^2$, $\text{var}(\epsilon_{u,1}) = \sigma_{\epsilon_u}^2$, and $\text{var}(\epsilon_{u,0}) = |a|^2\sigma_{\epsilon_v}^2 + \sigma_{\epsilon_u}^2$. Then, $K = 1$, $\sigma_{\varphi_1}^2$ is given by (58), and

$$\sigma_{\varphi_2}^2 = \begin{cases} \sigma_{\varphi_1}^2 & \mathcal{H}_0 \\ \sigma_{\epsilon_u}^2 + \sigma_{\nu_x}^2 + \frac{|a|^2\sigma_{\epsilon_v}^2\sigma_{\nu_y}^2}{\sigma_{\epsilon_u}^2+\sigma_{\nu_y}^2} = \sigma_{\varphi_1}^2 - \Delta & \mathcal{H}_1 \end{cases}, \quad (93)$$

with $\Delta = |a|^2\sigma_{\epsilon_v}^2\left(1-\frac{\sigma_{\nu_y}^2}{\sigma_{\epsilon_v}^2+\sigma_{\nu_y}^2}\right)$.

Therefore,

$$\mu_{T_D,0} = \sigma_{\varphi_1}^2\left(\lambda - 1 + \frac{\lambda-2}{N-1}\right), \quad (94)$$

and

$$\sigma_{T_D,0}^2 = \frac{1}{N-1}\left[(\lambda^2+1)\sigma_{\varphi_1}^4\left(1+\frac{2}{N-1}\right)^2 - \lambda^2\sigma_{\varphi_1}^4\left(\frac{3}{(N-1)^2}+\frac{2}{N-1}\right) - 2\lambda\frac{\sigma_{\varphi_1}^2}{N-1}\right]. \quad (95)$$

Consequently, the probability of false alarm is given as (96) at the top of the next page.

Under the alternate hypothesis,

$$E[T_D|\mathcal{H}_1] = (\lambda-1)\sigma_{\varphi_1}^2\left(1+\frac{1}{N-1}\right) - \Delta\left(1+\frac{1}{N-1}\right) - (\sigma_{\varphi_1}^2-\Delta)\frac{1}{N-1}, \quad (97)$$

and

$$\text{var}(T_D|\mathcal{H}_1) = \frac{1}{N-1}\left[(\lambda^2+1)\sigma_{\varphi_1}^4\left(1+\frac{1}{N-1}\right)^2 + \Delta^2\left(1+\frac{2}{N-1}\right)^2 - 2\sigma_{\varphi_2}^2\Delta\left(1+\frac{2}{N-1}\right)^2 + \sigma_{\varphi_1}^4\left(\frac{3}{(N-1)^2}+\frac{2}{N-1}\right)\right]. \quad (98)$$

The probability of detection can thus be expressed as (99) in the next page.

We observe from (99) that the probability of detection depends on the ratio $\frac{\Delta}{\sigma_{\varphi_1}^2}$, which can in turn be viewed as the relative difference between the variances of the prediction error under the one variable and the two variable models. Consequently, the difference between the mean square prediction error for the single and two variable prediction models, i.e., $t_1 - t_2$ can be used as a test statistic to detect the presence of a causal relationship. We inspect the use of $t_1 - t_2$ as a test statistic in the next section.

VI. Granger Causality Detection Based on the Difference of Error Terms

Let $T_C \triangleq t_1 - t_2$ denote the difference in the finite sample estimates of the mean squared errors of the one variable and the two variable prediction models. It can be argued that T_C is the difference of two correlated Chi-squared r.v.s., each with $2(N-p)$ degrees of freedom. Therefore, the pdf of T_C can be written as [37]

$$f_{T_C}(y) = \frac{|y|^{N-p-1}}{(m-1)![2\text{var}(t_1)\text{var}(t_2) - \text{cov}^2(t_1,t_2)\gamma]^{(N-p)}} \times \exp\left(\frac{1}{4}\alpha^- y\right)\sum_{i=0}^{N-p-1}\frac{(N-p+i-1)!}{(N-p-i-1)!}\left(\frac{2}{\gamma|y|}\right)^i, \; y < 0;$$

$$P_{\text{FA}} = Q\left(\frac{\sqrt{(N-1)}\left(\lambda - 1 + \frac{\lambda-2}{N-1}\right)}{\sqrt{(\lambda^2+1)\sigma_{\varphi_2}^4\left(1+\frac{2}{N-1}\right)^2 - \lambda^2\sigma_{\varphi_2}^4\left(\frac{3}{(N-1)^2}+\frac{2}{N-1}\right) - 2\lambda\frac{\sigma_{\varphi_1}^4}{N-1}}}\right). \tag{96}$$

$$P_{\text{D}} = Q\left(\frac{\sqrt{N-1}\left((\lambda-1)\sigma_{\varphi_1}^2\left(1+\frac{1}{N-1}\right) - \Delta\left(1+\frac{1}{N-1}\right) - (\sigma_{\varphi_1}^2-\Delta)\frac{1}{N-1}\right)}{\sqrt{(\lambda^2+1)\sigma_{\varphi_1}^4\left(1+\frac{1}{N-1}\right)^2 + \Delta^2\left(1+\frac{2}{N-1}\right)^2 - 2\sigma_{\varphi_2}^2\Delta\left(1+\frac{2}{N-1}\right)^2 + \sigma_{\varphi_1}^4\left(\frac{3}{(N-1)^2}+\frac{2}{N-1}\right)}}\right). \tag{99}$$

$$f_{T_C}(y) = \frac{|y|^{N-p-1}}{(m-1)![2\text{var}(t_1)\text{var}(t_2) - \text{cov}^2(t_1,t_2)\gamma]^{(N-p)}} \times \exp\left(-\frac{1}{4}\alpha^+ y\right) \sum_{i=0}^{N-p-1} \frac{(N-p+i-1)!}{(N-p-i-1)!}\left(\frac{2}{\gamma|y|}\right)^i,\ y \geq 0, \tag{100}$$

with

$$\gamma = \frac{[(\text{var}(t_1)+\text{var}(t_2))^2 - 4\text{cov}^2(t_1,t_2)]^{\frac{1}{2}}}{\text{var}(t_1)\text{var}(t_2) - \text{cov}^2(t_1,t_2))}, \tag{101}$$

and

$$\alpha^{\pm} = \gamma \pm \frac{[(\text{var}(t_1) - \text{var}(t_2))]}{\text{var}(t_1)\text{var}(t_2) - \text{cov}^2(t_1,t_2))}. \tag{102}$$

The p.d.f. of T_C is in the form of a series, which makes it difficult to obtain the probabilities of detection and false alarm in closed form. However, for large N, T_C can be approximated as a Gaussian r.v., such that,

$$E[T_C|\mathcal{H}_0] = -\sigma_{\varphi_2}^2\frac{p}{N-p}, \tag{103}$$

$$E[T_C|\mathcal{H}_1] = \sigma_{\acute{x}}^2\left(1+\frac{p}{N-p}\right) - \sigma_{\varphi_2}^2\frac{p}{N-p}, \tag{104}$$

and

$$\text{var}(T_C|\mathcal{H}_0) = \frac{1}{N-p}\Bigg[2\sigma_{\varphi_2}^4\left(1+\frac{2p}{N-p}\right)^2 - \sigma_{\varphi_2}^4\left(\frac{3p^2}{(N-p)^2}+\frac{2p}{N-p}\right) - 2\sigma_{\varphi_2}^4\Bigg]. \tag{105}$$

$$\text{var}(T_D|\mathcal{H}_1) = \frac{1}{N-p}\Bigg[2\sigma_{\varphi_2}^4\left(1+\frac{2p}{N-p}\right)^2 + \sigma_{\acute{x}}^4\left(1+\frac{p}{N-p}\right)^2 + 2\sigma_{\varphi_2}^2\sigma_{\acute{x}_2}^2\left(1+\frac{p}{N-p}\right)^2 - \sigma_{\varphi_2}^4\left(\frac{3p^2}{(N-p)^2}+\frac{2p}{N-p}\right) - 2\sigma_{\varphi_2}^4\Bigg]. \tag{106}$$

Using the above statistics, it is now straightforward to design the detector to achieve a given probability of false alarm, and obtain closed-form expressions for the probability of false alarm and detection, similar to (89) and (90), respectively. In particular, when $N \gg p$, i.e., the number of samples is much larger than the model order, it can be easily shown that the detection threshold λ is well approximated by

$$\lambda = \frac{\sqrt{2}Q^{-1}(P_{FA})\sigma_{\varphi_2}^2}{\sqrt{N-1}}, \tag{107}$$

and the probability of detection as a function of the target false alarm probability P_{FA}^* can be approximated as

$$P_D \approx Q\left(\frac{Q^{-1}\left(P_{FA}^*\right)\sqrt{2}\sigma_{\varphi_2}^2 - \sqrt{N-p}\sigma_{\acute{x}}^2}{\sqrt{\sigma_{\acute{x}}^4 + 2\sigma_{\varphi_2}^2\sigma_{\acute{x}_2}^2}}\right). \tag{108}$$

Based on the above, the number of samples required to attain a given pair of probabilities of detection and false alarm can be approximated as

$$N \approx p + \left(Q^{-1}(P_{FA})\sqrt{2}\frac{\sigma_{\varphi_2}^2}{\sigma_{\acute{x}}^2} + Q^{-1}(P_D)\ \sqrt{1+2\frac{\sigma_{\varphi_2}^2}{\sigma_{\acute{x}}^2}}\right)^2. \tag{109}$$

Revisiting the example considered in the previous sections, we observe that, $\sigma_{\varphi_1}^2 = \sigma_x^2$, $\sigma_{\varphi_2}^2 = \sigma_x^2 - \Delta$, using which, it is easy to obtain the expressions for the probabilities of false alarm and detection. We omit writing out the expressions for the sake of brevity.

We can also use the mean squared value of the difference of error terms as a test statistic to detect the presence of a causal relationship between $x[n]$ and $y[n]$. Letting $\check{x}[n] \triangleq \hat{x}[n] - \bar{x}[n]$, and

$$T_A \triangleq \frac{1}{N-p}\sum_{n=p+1}^{N}|\check{x}[n]|^2, \tag{110}$$

it can be shown that

$$E[T_A|\mathcal{H}_0] = \sigma_{\varphi_1}^2\left(\frac{3p}{N-p}\right), \tag{111}$$

and

$$E[T_A|\mathcal{H}_1] = \mathbf{r}_{zz,p}^H\mathbf{R}_{zz,p}^{-1}\mathbf{r}_{zz,p} + \mathbf{r}_{xx,p}^H\mathbf{R}_{xx,p}^{-1}\mathbf{r}_{xx,p} - 2\Re\{\mathbf{r}^H\mathbf{R}_{zz,p}^{-1}\mathbf{R}_{zx}\mathbf{R}_{xx,p}^{-1}\mathbf{r}_{xx,p}\} + \frac{2\sigma_{\varphi_2}^2 p}{N-p} + \frac{\sigma_{\varphi_1}^2 p}{N-p}. \tag{112}$$

However, the higher order statistics and the exact distribution of T_A cannot be found in closed form. Therefore, we evaluate its performance via simulations. To the best of our knowledge, the test statistics T_C and T_A have not been used in the literature to detect Granger causality.

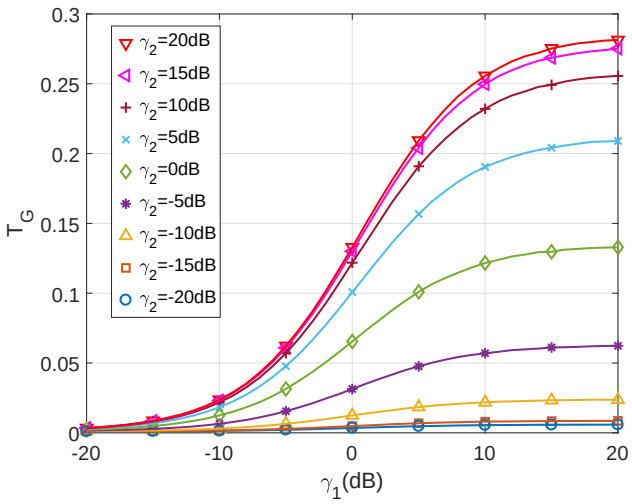


Fig. 1: Value of T_E at different SNRs, for $a = 0.5$, with known second-order statistics.

VII. Simulation Results

In this section, we use Monte Carlo simulations to corroborate our analytical expressions and illustrate the relative performance of different test statistics for detecting Granger causality. We also illustrate the causality detection performance of T_C using real-world data with known ground truth, available in [38]

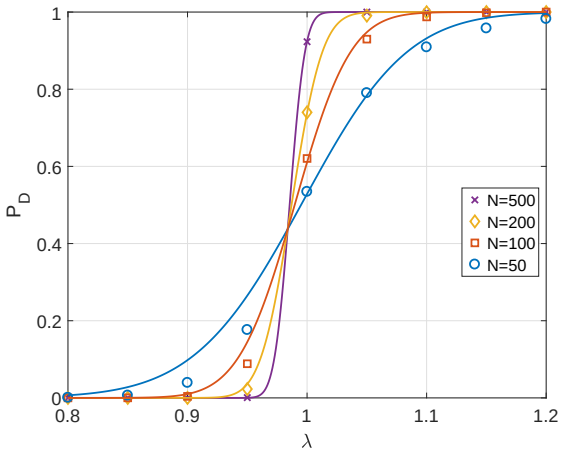
For the simulation-based study, we consider a system model similar to the example discussed in the previous sections, for different values of the variance of the additive noise and varying number of samples. The probabilities of detection and false alarm are obtained by averaging over $10,000$ independent realizations of the signals of interest.

A. Effect of Additive Noise

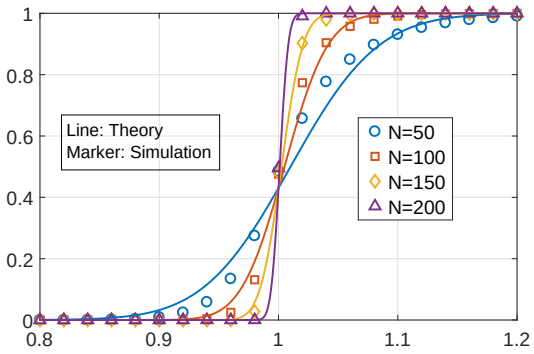
In Fig. 1, we plot the mean value of the test statistic T_G for the example considered in Section IV, with $b = 0.5$. Here we inspect the effects of noise on the GCI. We define $\gamma_1 \triangleq \frac{\sigma_u^2}{\sigma_{\nu_1}}$, and $\gamma_2 \triangleq \frac{\sigma_v^2}{\sigma_{\nu_2}}$ as the observation SNRs for $x[n]$ and $y[n]$. We see that the mean value of T_G converges to a value close to zero as the noise variance increases. The reduction in the SNR will make it harder to detect the presence of a causal relationship between the signals of interest.

B. Finite Sample Effects

We now consider the effects of using a finite number of samples on the GCI, as discussed in Section V. In Figs. 2a and 2b, we compare the simulated detection and false alarm probability obtained by using the test statistic T_E as a function of the detection threshold λ and for different number of samples, with $a = 0.25$ and $\gamma_1 = \gamma_2 = 0$ dB. The simulated performance of T_E is found to be in close agreement with the values computed using our analytical expressions in (96) and (99). It is also observed that the behavior of both the probability of detection and false alarm as a function of the



(a)



(b)

Fig. 2: Probabilities of (a) detection and (b) false alarm as a function of the detection threshold λ for the test statistic T_E.

detection threshold becomes sharper with an increase in the number of samples, and asymptotically approximates a step function. The transition point of the step function occurs at the mean of the test statistic under the respective hypotheses, which corresponds to the 50% probability of detection/ false alarm point on Figs. 2a and 2b. Therefore, the threshold $\lambda = 1$ gives a zero error rate when infinite samples of the two signals of interest are available. However, in case of a finite number of samples, λ has to be chosen carefully to achieve reliable performance.

In Fig. 3, we compare the theoretical and simulated receiver operating characteristics (ROCs) for the example considered in Section V for $a = 0.25$ with different values of N, at an SNR of 0 dB. It is observed that for a small number of samples, the ROC is close to the $P_{FA} = P_D$ line, whereas, for a large number of samples, the probability of detection becomes almost independent of P_{FA}. This is in line with the behavior predicted in (96) and (99).

Having established that the use of finite samples indeed causes a deterioration in the performance of the CGI, we now discuss the performance of the other two test statistics, viz. T_C and T_A, as discussed in Section VI. In Fig. 4a, we compare the theoretical and simulated ROCs for the example considered in Section VI, with $a = 0.25$, for different number of samples,

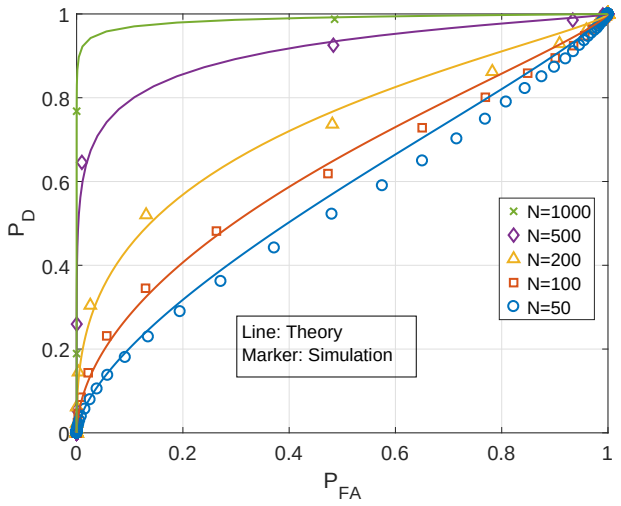


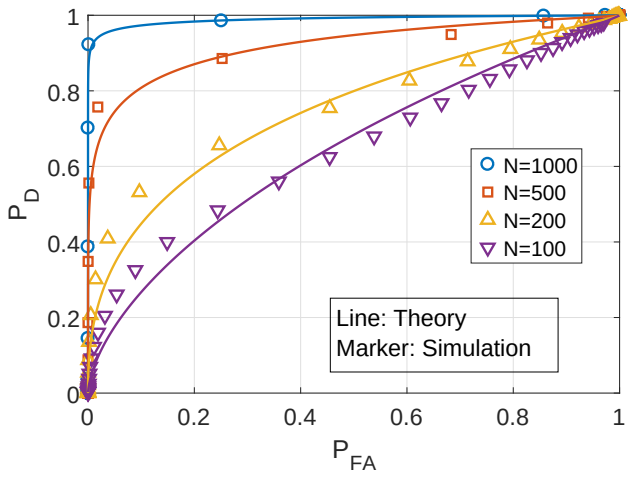
Fig. 3: ROC for detection based on T_E, at an SNR of 0 dB, for different number of samples.

and at an SNR of 0 dB. We observe that the ROC for T_C shows a behavior similar to that for T_E with an increase in the number of samples. Therefore, an appropriate choice of the detection threshold λ becomes important for T_C as well.

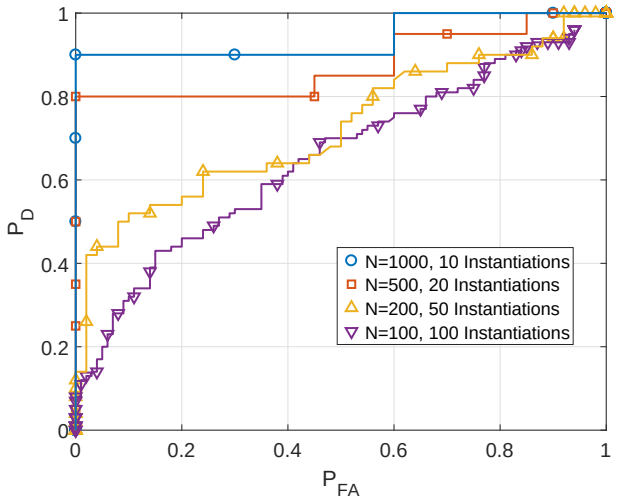
In practice, one is typically provided a given number of samples of a pair of time-series data to perform detection. In this case, there is a trade-off between the window size N and the quality of the ROC obtained. Using a larger N improves the reliability of detection, but the resulting ROC has more "jumps" because it is computed by averaging over fewer instantiations. We illustrate this in Fig. 4b, where we plot the ROC for the setup of Fig. 4a with the total number of available samples fixed at $10,000$. We see that the ROC with $N = 100$ (averaged over 100 instantiations) matches well with the corresponding curve in Fig. 4a, where $10,000$ instantiations were used. For higher N, the performance improves, but the quality of the ROC deteriorates, because the probability of detection is averaged over a correspondingly fewer number of instantiations. Thus, given a data record of a given size, a good choice for N is thus the largest one such that the ROC is averaged over about 100 instantiations.

In Fig. 5, we repeat the experiment considered in Fig. 4a by fixing the number of samples of $x[n]$ and $y[n]$ at 100 and varying the SNRs. Again, the theoretical and simulated plots are in close agreement, with the concavity of the ROC increasing with the SNR.

In Figs. 6 and 7, we evaluate the performance of T_A as a test statistic for detection of a causal relation between $x[n]$ and $y[n]$ for a setup similar to the example considered in Section V, with $a = 0.25$. Since theoretical expressions for the behavior of T_A are not derived, these are not considered here. In Fig. 6, we plot the ROC of T_A for different SNRs for 100 samples of each of $x[n]$ and $y[n]$ being used for detection. The performance of T_A under different SNRs is seen to mimic that of T_C and T_E, with the concavity of the ROC increasing with the SNR. In Fig 7, we plot the probabilities of detection for different numbers of samples as a function of the signal



(a)



(b)

Fig. 4: ROC for detection based on T_C, at an SNR of 0 dB, for different number of samples (N) (a) averaged over 10,000 realizations in all cases, (b) obtained for a fixed data record length (10,000) divided into windows of length N.

SNRs. In this case, we use the Neyman-Pearson criterion to fix the probability of false alarm at 10%. As expected, an increase in the number of samples leads to improved detection performances for the same SNR. Therefore, T_A can also be used to detect a causal relationship between a pair of signals in the finite sample regime.

In Fig. 8, we compare the simulated performance of T_E, T_C and T_A for the example considered previously, with $b = 0.25$, at an SNR of 0 dB and for 100 samples. We observe that T_A, despite being theoretically intractable, offers the best performance among the three test statistics considered in this paper, followed by T_C and T_E respectively. It is interesting to note that if the exact knowledge about the second-order statistics is available, then the performance of the three test statistics should be identical.

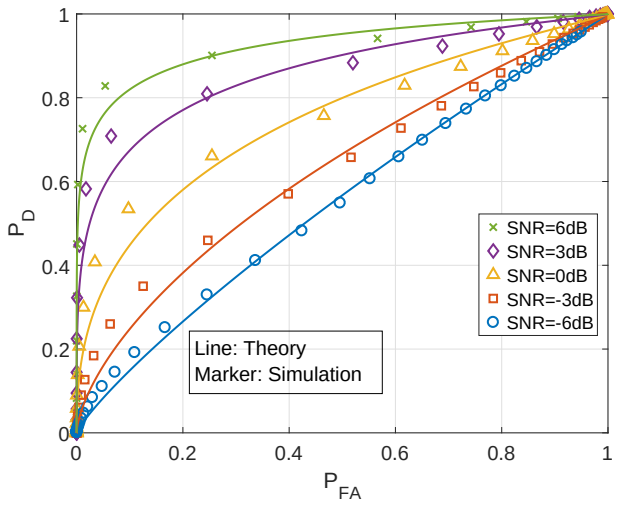


Fig. 5: ROC for detection based on T_C, with $N = 200$ samples, for various values of the SNR.

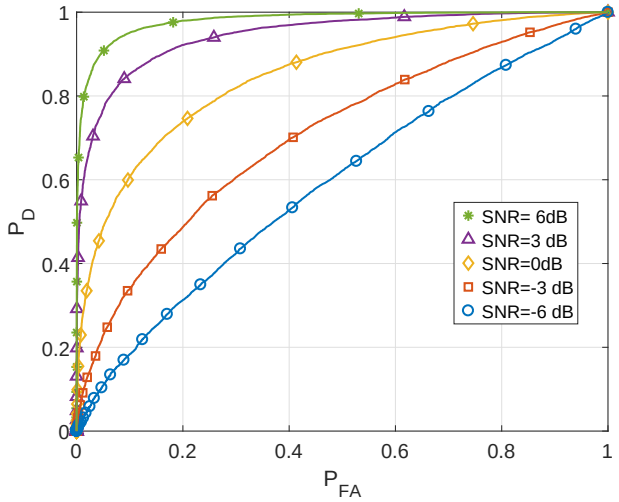


Fig. 6: ROC for detection based on T_A, with $N = 100$ samples, for various values of the SNR.

C. Detection Performance of T_C with Real World Data

In Figs. 9 and 10, we plot the performance obtained from the test statistic T_C on a real-world dataset with known ground truth, available in [38]. Fig 9 corresponds to data pairs $65 - 67$ of this database. These data pairs consist of the returns from related stocks over a period of five years. Similarly, Fig 10 corresponds to data pair 69 of the database, containing temperature measurements from inside and outside a room taken every five minutes. These four data sets were selected out of the 108 available data sets because:

1) They represent time series data, in accordance with our system model.
2) The number of samples in these time series were sufficiently large to estimate of probabilities of detection and false alarm with good accuracy.

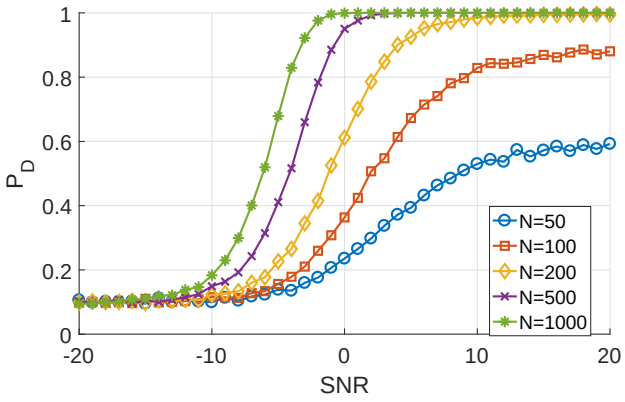


Fig. 7: Probability of detection using T_A at a fixed false alarm rate of 10% as a function of the SNR.

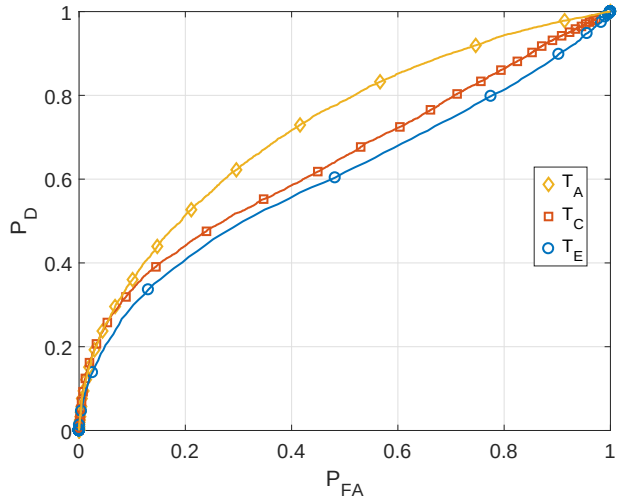


Fig. 8: ROC of test statistics T_E, T_C and T_A for $N = 100$ samples at an SNR of 0 dB.

For generating the experimental results, we assumed that the data was uniformly sampled at Nyquist rate. However, neither the model parameters of the data nor the additive noise variance are known. Hence, an order 1-VAR generative model is assumed, and each data set is preprocessed to make it zero mean. The detection threshold for a given false alarm rate (P_{FA}) is calculated assuming the two sequences to be independent and white, and using the sample variance of the caused sequence as $\sigma^2_{\varphi_2}$ in (107).

The data pairs are divided into windows of length N, and the test statistic T_C is calculated and compared to the threshold in (107). The results of these comparisons are averaged over all the sample windows to obtain the probability of detection for a given probability of false alarm. The jitters in the plots are due to the averaging over the relatively small number of samples that were available. The performance of data set 69 is poorer than data sets 65-67, for the same value of N, indicating either a weaker causal relationship, or larger amount of additive noise, or both. Nonetheless, the ROC follows a pattern similar to that predicted by (108).

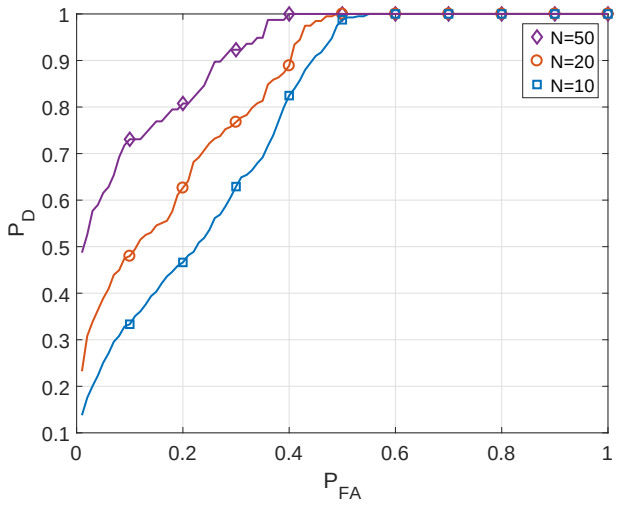


Fig. 9: Performance of the test statistic T_C on the real-world datasets $67-69$ in [38] for different numbers of samples.

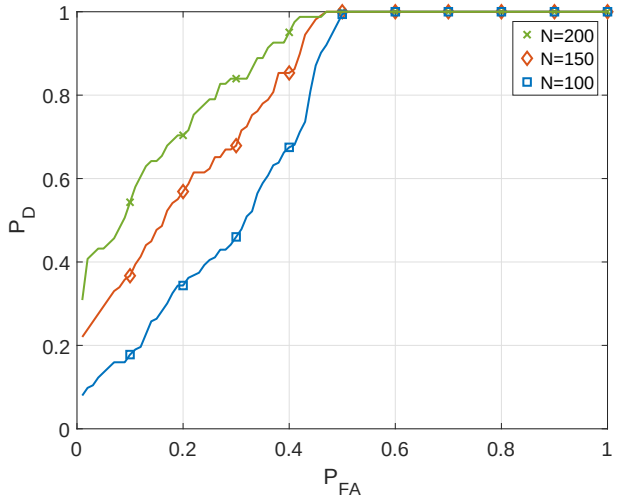


Fig. 10: Performance of the test statistic T_C on the real-world dataset 69 in [38] for different numbers of samples.

In Fig. 11, we plot the performance of T_C for real-world data used in Fig. 10 with different amounts of added noise for a window length of $N = 200$. Noisy samples are obtained by adding real valued white Gaussian noise to the caused sequence, suitably scaled to achieve a given SNR. These samples are used to calculate the test statistic T_C, which is then compared against the detection threshold in (107), with $\sigma^2_{\varphi_1}$ as the sample variance of the corrupted caused signal. It is observed that for a fixed number of samples per window, the addition of noise worsens the detection performance of T_C, and the variation in the detection performance approximately matches with the Q-function of the square root of the SNR, as expected.

Acknowledgments

The authors are grateful to the anonymous reviewers for their feedback, which significantly improved the overall presentation, and in particular the simulation results in the paper.

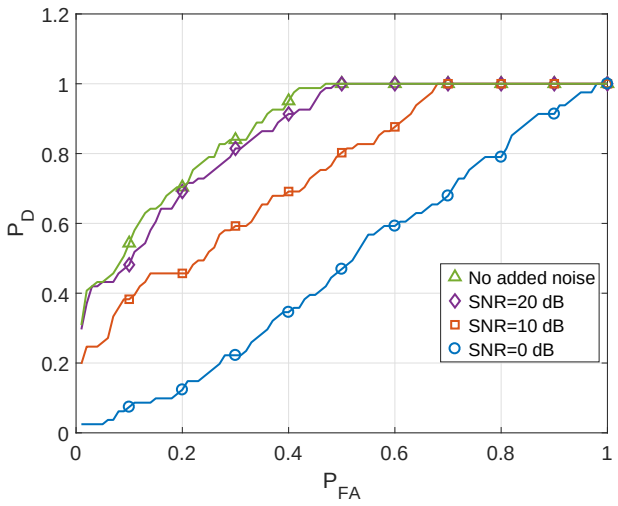


Fig. 11: Performance of T_C for the dataset considered in Fig. 10 for $N = 200$ with different amounts of added noise.

VIII. Conclusions

In this work, we derived the effects of different data acquisition impairments on the detection of a causal relation between two signals. We showed that these effects, viz. downsampling, additive noise, and finite sample effects, could lead to a significant deterioration in the performance of the GCI, both individually as well as cumulatively. We derived the probabilities of detection and false alarm for the GCI under the impairments. Following this, we proposed two alternative test statistics, based on the difference of mean squared error terms, and the mean squared value of the difference of error terms. Via extensive simulations, we showed that the derived results corroborate the simulated as well as real-world data, with the test statistic based on the mean squared value of the difference of error terms outperforming the other two statistics. A more precise analysis of this test statistic is an interesting direction for future research.

References

[1] I. Kontoyiannis and M. Skoularidou, "Estimating the directed information and testing for causality," *IEEE Trans. Inf. Theory*, vol. 62, pp. 6053–6067, Nov. 2016.

[2] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.

[3] D. A. Pierce, "R^2 measures for time series," *Journal of the American Statistical Association*, vol. 74, no. 368, pp. 901–910, 1979.

[4] R. Goebel, A. Roebroeck, D.-S. Kim, and E. Formisano, "Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping," *Magnetic Resonance Imaging*, vol. 21, no. 10, pp. 1251 – 1261, 2003.

[5] W. Hesse, E. Mller, M. Arnold, and B. Schack, "The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies," *Journal of Neuroscience Methods*, vol. 124, no. 1, pp. 27 – 44, 2003.

[6] Y. Chen, G. Rangarajan, J. Feng, and M. Ding, "Analyzing multiple nonlinear time series with extended Granger causality," *Physics Letters A*, vol. 324, no. 1, pp. 26 – 35, 2004.

[7] M. Dhamala, G. Rangarajan, and M. Ding, "Analyzing information flow in brain networks with nonparametric Granger causality," *NeuroImage*, vol. 41, no. 2, pp. 354 – 362, 2008.

[8] S. Hu and H. Liang, "Causality analysis of neural connectivity: New tool and limitations of spectral Granger causality," *Neurocomputing*, vol. 76, no. 1, pp. 44 – 47, 2012. Seventh International Symposium on Neural Networks (ISNN 2010)Advances in Web Intelligence.

[9] D. Sampath, K. Sabitha, P. Hegde, H. Jayakrishnan, B. M. Kutty, S. Chattarji, G. Rangarajan, and T. Laxmi, "A study on fear memory retrieval and {REM} sleep in maternal separation and isolation stressed rats," *Behavioural Brain Research*, vol. 273, pp. 144 – 154, 2014.

[10] M. Hu and H. Liang, "A copula approach to assessing Granger causality," *NeuroImage*, vol. 100, pp. 125 – 134, 2014.

[11] A. K. Seth, A. B. Barrett, and L. Barnett, "Granger causality analysis in neuroscience and neuroimaging," *Journal of Neuroscience*, vol. 35, no. 8, pp. 3293–3297, 2015.

[12] R. Ganapathy, G. Rangarajan, and A. K. Sood, "Granger causality and cross recurrence plots in rheochaos," *Phys. Rev. E*, vol. 75, p. 016211, Jan. 2007.

[13] S. Stramaglia, J. M. Cortes, and D. Marinazzo, "Synergy and redundancy in the Granger causal analysis of dynamical networks," *New Journal of Physics*, vol. 16, no. 10, p. 105003, 2014.

[14] E. Kodra, S. Chatterjee, and A. R. Ganguly, "Exploring Granger causality between global average observed time series of carbon dioxide and temperature," *Theoretical and Applied Climatology*, vol. 104, no. 3, pp. 325–335, 2011.

[15] H. White and D. Pettenuzzo, "Granger causality, exogeneity, cointegration, and economic policy analysis," *Journal of Econometrics*, vol. 178, Part 2, pp. 316 – 330, 2014. Recent Advances in Time Series Econometrics.

[16] H. Nalatore, M. Ding, and G. Rangarajan, "Mitigating the effects of measurement noise on Granger causality," *Phys. Rev. E*, vol. 75, p. 031123, Mar 2007.

[17] H. Nalatore, S. N, and G. Rangarajan, "Effect of measurement noise on Granger causality," *Phys. Rev. E*, vol. 90, p. 062127, Dec 2014.

[18] I. Winkler, D. Panknin, D. Bartz, K. R. Mller, and S. Haufe, "Validity of time reversal for testing Granger causality," *IEEE Transactions on Signal Processing*, vol. 64, pp. 2746–2760, June 2016.

[19] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, pp. 461–464, July 2000.

[20] L. A. Baccalá and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biological Cybernetics*, vol. 84, no. 6, pp. 463–474, 2001.

[21] E. Siggiridou and D. Kugiumtzis, "Granger causality in multivariate time series using a time-ordered restricted vector autoregressive model," *IEEE Transactions on Signal Processing*, vol. 64, pp. 1759–1773, April 2016.

[22] P. Newbold and N. Davies, "Error mis-specification and spurious regressions," *International Economic Review*, vol. 19, no. 2, pp. 513–19, 1978.

[23] G. Nolte, A. Ziehe, V. V. Nikulin, A. Schlögl, N. Krämer, T. Brismar, and K.-R. Müller, "Robustly estimating the flow direction of information in complex physical systems," *Phys. Rev. Lett.*, vol. 100, p. 234101, Jun 2008.

[24] S. Haufe, V. V. Nikulin, and G. Nolte, *Alleviating the Influence of Weak Data Asymmetries on Granger-Causal Analyses*, pp. 25–33. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[25] S. Haufe, V. V. Nikulin, K.-R. Mller, and G. Nolte, "A critical assessment of connectivity measures for {EEG} data: A simulation study," *NeuroImage*, vol. 64, pp. 120 – 133, 2013.

[26] M. Vinck, L. Huurdeman, C. A. Bosman, P. Fries, F. P. Battaglia, C. M. Pennartz, and P. H. Tiesinga, "How to detect the Granger-causal flow direction in the presence of additive noise?," *NeuroImage*, vol. 108, pp. 301 – 318, 2015.

[27] A. K. Seth, P. Chorley, and L. C. Barnett, "Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling," *NeuroImage*, vol. 65, no. Supplement C, pp. 540 – 555, 2013.

[28] L. Barnett and A. K. Seth, "Detectability of Granger causality for subsampled continuous-time neurophysiological processes," *Journal of Neuroscience Methods*, vol. 275, no. Supplement C, pp. 93 – 121, 2017.

[29] M. Gong, K. Zhang, B. Schoelkopf, D. Tao, and P. Geiger, "Discovering temporal causal relations from subsampled data," in *Proceedings of the 32nd International Conference on Machine Learning* (F. Bach and D. Blei, eds.), vol. 37 of *Proceedings of Machine Learning Research*, (Lille, France), pp. 1898–1906, PMLR, 07–09 Jul 2015.

[30] M. Gong, K. Zhang, B. Schölkopf, C. Glymour, and D. Tao, "Causal discovery from temporally aggregated time series," in *Proceedings Conference on Uncertainty in Artificial Intelligence (UAI) 2017*, p. ID 269, Association for Uncertainty in Artificial Intelligence (AUAI), Aug. 2017.

[31] A. Tank, E. B. Fox, and A. Shojaie, "Identifiability of non-gaussian structural var models for subsampled and mixed frequency time series," in *Proceedings SIGKDD Workshop on Causal Discovery 2016*, 2016.

[32] A. Hyttinen, S. Plis, M. Jrvisalo, F. Eberhardt, and D. Danks, "A constraint optimization approach to causal discovery from subsampled time series data," *International Journal of Approximate Reasoning*, vol. 90, pp. 208 – 225, 2017.

[33] S. Plis, D. Danks, C. Freeman, and V. Calhoun, "Rate-agnostic (causal) structure learning," in *Advances in Neural Information Processing Systems 28* (C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, eds.), pp. 3303–3311, Curran Associates, Inc., 2015.

[34] X. Wen, G. Rangarajan, and M. Ding, "Is Granger causality a viable technique for analyzing fMRI data?," *PLOS ONE*, vol. 8, pp. 1–11, 07 2013.

[35] S. Haykin, *Adaptive Filter Theory*. Pearson Education, 2002.

[36] R. Tandra and A. Sahai, "SNR walls for signal detection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, pp. 4–17, Feb 2008.

[37] M. K. Simon, *Probability Distributions Involving Gaussian Random Variables*. Springer US, 2002.

[38] J. M. Mooij, J. Peters, D. Janzing, J. Zscheischler, and B. Schölkopf, "Distinguishing cause from effect using observational data: Methods and benchmarks," *Journal of Machine Learning Research*, vol. 17, no. 32, pp. 1–102, 2016.

Ribhu Chopra (S'11–M'17) received the B.E. degree in Electronics and Communication Engineering from Panjab University, Chandigarh, India in 2009, and the M. Tech. and Ph. D. Degrees in Electronics and Communication Engineering from the Indian Institute of Technology Roorkee, India in 2011 and 2016 respectively. He worked as a project associate at Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore from Aug. 2015, till May 2016. From May 2016 to March 2017 he worked as an institute research associate at the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India. In April 2017, he joined the department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Assam, India. His research interests include statistical and adaptive signal processing, massive MIMO communications, and cognitive communications.

Chandra R. Murthy (S'03–M'06–SM'11) received the B. Tech. degree in Electrical Engineering from the Indian Institute of Technology Madras, India, in 1998, the M. S. and Ph. D. degrees in Electrical and Computer Engineering from Purdue University and the University of California, San Diego, USA, in 2000 and 2006, respectively. From 2000 to 2002, he worked as an engineer for Qualcomm Inc., where he worked on WCDMA baseband transceiver design and 802.11b baseband receivers. From Aug. 2006 to Aug. 2007, he worked as a staff engineer at Beceem Communications Inc. on advanced receiver architectures for the 802.16e Mobile WiMAX standard. In Sept. 2007, he joined the Department of Electrical Communication Engineering at the Indian Institute of Science, Bangalore, India, where he is currently working as a Professor.

His research interests are in the areas of energy harvesting communications, multiuser MIMO systems, and sparse signal recovery techniques applied to wireless communications. His paper won the best paper award in the Communications Track at NCC 2014 and a paper co-authored with his student won the student best paper award at the IEEE ICASSP 2018. He has 50+ journal papers and 80+ conference papers to his credit. He was an associate editor for the IEEE Signal Processing Letters during 2012-16. He is an elected member of the IEEE SPCOM Technical Committee for the years 2014-16, and has been re-elected for the 2017-19 term. He is a past Chair of the IEEE Signal Processing Society, Bangalore Chapter. He is currently serving as an associate editor for the IEEE Transactions on Signal Processing, the Sadhana Journal and as an editor for the IEEE Transactions on Communications.