Dynamic Window-level Granger Causality of Multi-channel Time Series

Anonymous authors
Paper under double-blind review

Abstract

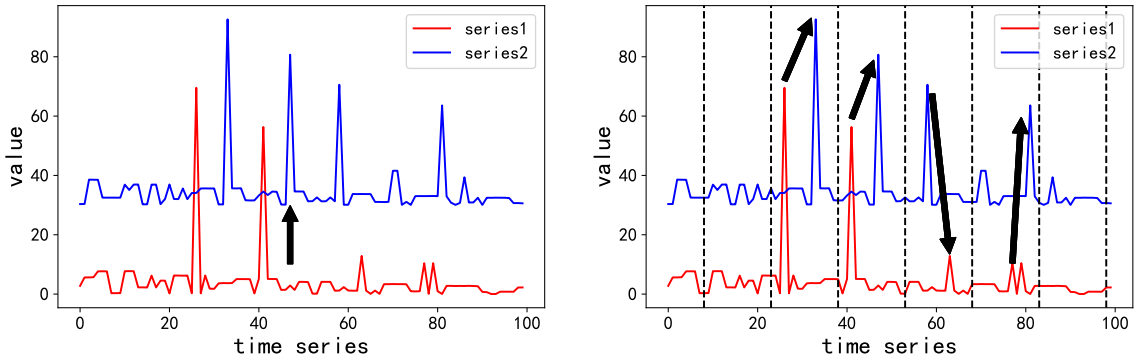
Granger causality (GC) techniques facilitate the examination of temporal causal relationships by assessing the predictive utility of one time series for another. However, conventional GC approaches are limited to linear scenarios and assume that causalities exclusively exist between entire time series channels, remaining constant over time. This assumption falls short when dealing with real-world time series data characterized by dynamic causalities among channels. To address this limitation, we present the dynamic window-level Granger causality with causality index (DWGC-CI) method, which incorporates nonlinear window-level variability into the traditional GC framework. The DWGC-CI method specifically employs F-tests on sliding windows to assess forecasting errors, subsequently extracting dynamic causalities using a causality index. This index involves the introduction of a negative feedback adjustment mechanism, which mitigates window-level noisy fluctuations and helps extract dynamic causalities. We theoretically demonstrate that, compared to traditional channel-level GC methods, DWGC-CI accurately captures window-level GCs. In experimental evaluations involving two synthetic datasets, DWGC-CI significantly surpasses the baseline in terms of accuracy and recall rates. Furthermore, when applied to two real-world datasets, DWGC-CI effectively identifies seasonal and annual varying GCs, demonstrating a markedly better consistency with domain-specific knowledge of seasonal meteorology and stock market dynamics.

1 Introduction

Time series data, characterized by a pre-defined temporal or sequential order Hamilton (1994), is extensively employed in a diverse range of real-world applications, encompassing signal processing Scharf (1991), economics Granger & Newbold (2014), and climatology Mudelsee (2013), among others. Common tasks involving time series data consist of indexing, clustering, classification, regression Keogh & Kasetty (2003), and causality detection Eichler (2012). Among these tasks, causality detection, which emphasizes cognitive reasoning Pearl (2018), offers a more comprehensive understanding of system behaviour compared to traditional statistical machine learning and deep learning methods that rely on correlation fitting. Owing to the inherent nature of temporal precedence in time series data, it not only encapsulates empirical patterns of variable changes but also captures prior knowledge of causal relationships among distinct channels Eichler (2013).

To deal with the causal reasoning task in time series, Granger Granger (1969) first proposed the statistical test to decide whether one time series channel is influential when predicting another, which is known as Granger causality (GC). With the GC method, the causality can be obtained from the temporal data in various domains, such as physics Kaufmann & Stern (1997), biology Angelini et al. (2009) and economy Comincioli (1996).

Nonetheless, conventional methods Xu et al. (2016); Wu et al. (2022); Granger (1969) presuppose that GC remains invariant across time series channels over time. It is a concept referred to as "channel-level GC", more precisely. While this assumption facilitates the resolution of numerous Granger tests, this channel-level static assumption is not obligatory. It mainly serves merely to streamline the Granger testing process Granger



(a) Channel-level GC.

(b) Dynamic window-level GC.

Figure 1: The channel-level approach (a) assumes that causal influences between time series channels remain consistent throughout the duration of the series. In contrast, the window-level causality method (b) involves scrutinizing the dynamic evolution of causal interactions within sliding windows applied to the time series data. This latter approach allows for a more nuanced understanding of shifting causality patterns (shown as four-wave crests) over time.

(1980). More crucially, this assumption contradicts numerous practical situations involving dynamic causalities. In reality, time series data has grown increasingly voluminous, intricate, and uncertain, causing causality relationships to evolve in conjunction with the sliding windows of time series data (refer to the four-wave crests as in Fig. equation 1). Notable applications encompass seasonal meteorology and stock market dynamics, which will be detailed in Sec. 5. Consequently, dynamic window-level causalities, when offering reliable and precise predictions, are better suited for real-world applications.

Addressing the challenge of time-varying window-level causalities, preliminary efforts have been made on the specific application domains. One notable example is the field of neuroscience, where the "dynamic causal modelling" (DCM) method Friston et al. (2003) identifies dynamic causal relationships between neuronal clusters by formulating differential equations that capture neural dynamics. However, the DCM approach is grounded in a domain-specific hypothesis-testing framework, rendering it unsuitable for the broader GC context. Stepping backwards, although alternative time-varying GC methods have been proposed Cekic et al. (2018); Pagnotta et al. (2018); Ming-Hsien & Chih-She (2015), these approaches predominantly rely on linear forecasting models and inadequately account for local fluctuations. For example, it is challenging to linearly model the series data depicted in Fig. equation 1, and more importantly, such fluctuations might result from random noise rather than causal relationships, leading to significant inaccuracies in local-window GC. In summary, current research on the theoretical underpinnings and extraction techniques for window-level time-varying GC is not mature enough, necessitating further investigation.

1.1 Our proposal

In this study, we propose a novel approach to GC analysis by introducing dynamic window-level GC with causality index (DWGC-CI), which extends the conventional GC framework by relaxing the assumption of constant causality on sufficiently long time series Granger (1980). Our method is insensitive to the selection of the prediction model, and we adopt the nonlinear autoregressive (NAR) model Chen et al. (2000) for simplification. Subsequently, we first present the dynamic window-level GC (DWGC) method, which employs the window-level F-test to compare prediction results on sliding windows, with GC being a special case when the entire channel is treated as a single window. Second, to mitigate potential fluctuations in the window-level F-test, we further establish a causality index matrix by optimizing its associated index loss, culminating in the complete DWGC-CI methodology.

In this study, we will provide a comprehensive elucidation of the theoretical guarantees and practical advantages of DWGC-CI. Theoretically, we demonstrate that: 1) the recall rate of DWGC increases with window length, ultimately converging to that of GC; It mathematically reveals a *granularity-efficacy trade-off*, which

drives the proposal of the DWCI-CI technique. By incorporating supplementary CI, our DWGC aspires to overcome this trade-off and 2) surpasses DWGC/GC in terms of recall rate. Experimentally, to evaluate the performance of DWGC-CI, we conduct experiments on four datasets, including two synthetic and two real-world datasets. For the synthetic data, we employ linear autoregressive (AR) and NAR simulators for generation, with DWGC-CI achieving superior window-level causality detection compared to the baseline, regarding the recall rate and accuracy rate. In the case of real-world datasets, DWGC-CI effectively uncovers dynamic causalities in the El-Niño and East Asian monsoon phenomena, as well as the relationship between the stock market and the economy. These findings are consistent with expert knowledge previously reported in the literature. In summary, the contributions of this paper are threefold:

- We are the first to propose a general dynamic window-level GC detection method taking advantage of the causality index, called DWGC-CI in total.
- We provide the theoretical proof demonstrating that the DWGC method contains a granularity-efficacy trade-off. Moreover, the recall rate of DWGC can be enhanced via utilizing the causality index trick, which is integrated within the final comprehensive DWGC-CI methodology.
- We evaluate the performance of our DWGC-CI method through a series of numerical experiments involving two synthetic and two real-world datasets. Our findings reveal that the DWGC-CI approach successfully identifies window-level causalities that align with previously established academic knowledge.

2 Preliminaries of Granger Causality

We use $(Y_{i,1}, Y_{i,2}, \cdots, Y_{i,t}, \cdots)$ to denote the time series channel $i \in \{1, 2, ...d\}$, where $Y_{i,t}$ and $Y_{i,<t}$ are the series channel i at and before time t. [1]. In this paper, we use the double-headed arrow to represent the "channel-level GC" Granger (1969) from channel j to i: $Y_j \Longrightarrow Y_i$. The GC from the series channel j to i holds if channel j provides useful information when predicting the future values of series i. In detail, the traditional GC uses an AR model g to predict future values:

$$\hat{Y}_{i,t} = \quad g_i(Y_{i,<t}), \hat{Y}_{i|j,t} = \quad g_i(Y_{i,<t}, Y_{j,<t}). \tag{1}$$

Then GC can be built based on an F-test to check the difference between $\hat{Y}_{i|j,t}$ and $\hat{Y}_{i,t}$, i.e., the forecasting result of the channel j with and without channel i:

$$F_{j \Longrightarrow i} = \sum_t (\hat{Y}_{i,t} - Y_{i,t})^2 / \sum_t (\hat{Y}_{i|j,t} - Y_{i,t})^2, \tag{2}$$

where g_i is the prediction on time series i (e.g., AR model in traditional GC method). The criteria of this hypothesis test is that we accept that GC $Y_j \Longrightarrow Y_i$ exists when $F_{j \Longrightarrow i}$ in Eqn. equation 2 is larger than a pre-defined causality threshold ϵ $(e.g., \epsilon \geq 1)$.[2]

The generalization of GC methods can be categorized into two primary domains: window-level and nonlinear. However, achieving a concise and effective integration of both within the general GC framework remains a challenging task. Regarding window-level GCs, the hypothesis-testing framework of dynamical causal modelling (DCM)Friston et al. (2003) is incompatible with the general GC framework. Additionally, other time-varying methodsCekic et al. (2018); Pagnotta et al. (2018); Ming-Hsien & Chih-She (2015) simply incorporate time-varying coefficients in regression models, but they exhibit suboptimal performance in handling irregular nonlinear cases. As for nonlinear GCs, Chen et al.Chen et al. (2004) employed a delay embedding technique to extend nonlinear Granger causality. Later, Sun et al.Sun (2008) suggested using

[1] In practice, we follow the GC's basic assumption–the time series Y is stationary; for the generating process, the joint probability distribution function of the piece of time series does not change over time. In order to achieve this objective optimally, it is imperative that we conduct stationarity tests previously and subsequently perform appropriate transformations to ensure stationarity.

[2] The well-known premise of Eqn. equation 2 is that series channels i and j meet the "backdoor condition" Pearl & Robins (1995), shown as Appendix. G.

kernel embeddings to address the nonlinearity. Subsequently, Sugihara et al.Sugihara et al. (2012) expanded the traditional GC to the Convergent Cross Mapping (CCM) theorem, which is based on Taken's embedding theoremTakens (1981). Moreover, there are additional attempts via tools such as transfer entropy Syczewska & Struzik (2015), causal structure learning Weichwald et al. (2020), spectrum-based framework Zhang et al. (2016). Eichler Eichler (2013) subsequently re-summarized the challenges associated with transferring GC to the nonlinear version, which includes: 1) aggregating the time-varying coefficients required by Granger causality tests and 2) the instability of the causal structure itself. More recently, neural networks have been introduced Tank et al. (2018); Xu et al. (2019); Duggento et al. (2019); He et al. (2019) to supplant the original autoregressive (AR) model for sequence fitting and prediction. Regrettably, the combination of window-level GC and nonlinear GC is quite challenging since it is hard to get through the local noisy information of complex data to extract informative GC. To our best knowledge, no absolutely universal and simple framework has yet been developed yet, and currently, there are only several attempts in specific engineering applications (such as EEG signals Li et al. (2012), switching systems Ma et al. (2014)).

In addition to GC, causal graph modelling serves as an alternative approach for discerning causal relationships between distinct time series channels Xu et al. (2019). Causal graph models can be constructed using various methodologies, including vector autoregressive Granger analysis (VAR), generalized additive models (GAMs) Lütkepohl (2005), and specific predefined regression models Sommerlade et al. (2012). However, generating a causality graph for each time point presents a significant challenge due to the increased computational complexity, particularly when identifying dynamic window-level causality. Furthermore, after graph construction, A. Mastakouri Mastakouri et al. (2021) theoretically analyze the condition of dynamic causality identification in a time series graph, whereas their theorem is based on complex assumptions which are computational to verify in the real world (such as considerable conditional independence tests), and it is also challenging to concurrently obtain all window-level causal relationships.

3 Method

3.1 DWGC model

To detect dynamic window-level causality between two data series Y_i and Y_j, we first define sliding windows of length k on the same time position t of both series:

$$\{Y_{i,t}, Y_{i,t+1}, \cdots, Y_{i,t+k-1}\}, \quad \{Y_{j,t}, Y_{j,t+1}, \cdots, Y_{j,t+k-1}\}.$$

Then we consider two forms of time-series fitting on the sliding windows on channel i with or without the information from channel j. Specifically, to fit the nonlinearity of the data, we use the NAR model as g in Eqn. equation 1, and modify Eqn. equation 2 as follows:

$$F_{j \Longrightarrow i}^{t,k} = \sum_q (\hat{Y}_{i,q} - Y_{i,q})^2 / \sum_q (\hat{Y}_{i|j,q} - Y_{i,q})^2, q \in [t, t+k-1], \tag{3}$$

where $\hat{Y}_{i,q}$ and $\hat{Y}_{i|j,q}$ denote the predictions on the local sliding window with and without the information of channel j, respectively. $F_{j \Longrightarrow i}^{t,k}$ denotes the F-statistic on the sliding window $[t, t+k-1]$ from channel j to i. Analogously to Eqn. equation 2, the criteria is that we accept the window-level causality $Y_{j,t \sim t+k-1} \Longrightarrow Y_{i,t \sim t+k-1}$ exists if $F_{j \Longrightarrow i}^{t,k}$ is larger than the pre-defined threshold ϵ (≥ 1). In this sense, GC is the special case of the DWGC method when k is equal to the channel length.

3.2 DWGC-CI model

In the preceding section, it was observed that the reduction in length from the channel (Eqn.2) to window (Eqn.3) renders the DWGC method susceptible to instability arising from NAR forecasting results. To address this issue, we introduce the "causality index (CI) matrix", designed to mitigate the temporal fluctuations of NAR forecasting outcomes. This approach can be likened to a negative feedback regulation mechanism, wherein iterative training is performed, and varying weights are assigned to each time point to stabilize the overall F-statistic results.

Algorithm 1 DWGC-CI method

Data: Multi-channel time series $\{Y_{i,t}\}$ and predefined F-test threshold ϵ.
Initialization, Φ set to all-one;
while *NAR forecast loss and* L_{index} *converge* **do**
  Reweight the original time series using causality index matrix Φ via Eqn. equation 4;
  Train an NAR model using the reweighted series $\{Y^{\Phi}_{i,t}\}$;
  Finding dynamic causalities via Eqn. equation 6;
  In all the window pairs with detected causalities, optimize L_{index} in Eqn. equation 5 to update Φ;
end
Result: Window-level GC $\{Y_{j,t\sim t+k-1} \Longrightarrow Y_{i,t\sim t+k-1}\}, i, j \in \{1, 2, ...d\}$.

Specifically, each original point $Y_{i,t}$ could be weighted by an index matrix denoted as $\Phi = \{\Phi_{i,t}\}$, which is taken as an additional input layer of the NAR model, the weighted item $\{\Phi_{i,t} * Y_{i,t}\}$ is denoted as:

$$Y^{\Phi}_{i,t\sim t+k-1} = \Phi_{i,t\sim t+k-1} * Y_{i,t\sim t+k-1},\ i = 1, 2, ...d, \tag{4}$$

where $Y_{i,t\sim t+k-1}$,$\Phi_{i,t\sim t+k-1}$, $Y^{\Phi}_{i,t\sim t+k-1}$ are the column vectors consisting of k elements in the window $[t, t+k-1]$, denoting original data, weighting factor and weighted data respectively, and $*$ is the Hadamard product. We establish an minimization goal to learn the indexing matrix Φ to scale down the temporal fluctuations:

$$L_{index} = \sum_{i,q\in M} \left\| \Phi_{i,q} - h\left(\left| (\hat{Y}^{\Phi}_{i,q} - Y^{\Phi}_{i,q})^2 - \mathbb{E}_q(\hat{Y}^{\Phi}_{i,q} - Y^{\Phi}_{i,q})^2 \right| \right) \right\|_2^2, \tag{5}$$

where M is the set of sliding windows which detected causality and h is a monotonically decreasing function[3]. On this basis, the effectiveness of the "causality index (CI) matrix" can be explained as a *dynamic negative feedback regulation*: In each iteration, it aims to regulate overall forecasting errors on each window which detected causality by reducing the effect of adverse factors such as local noise and local overfitting. For instance, when the term $(\hat{Y}_{i,q} - Y_{i,q})^2$ is much larger (noise) or smaller (overfitting) than others, which is a statistically significant anomaly, then $\left|(\hat{Y}_{i,q} - Y_{i,q})^2 - \mathbb{E}_q(\hat{Y}_{i,q} - Y_{i,q})^2\right|$ will be significantly large. We use the negatively correlated index $\Phi_{i,q}$ to superimpose it to weaken its fluctuation on the local-window F-statistics.

On this basis, the window-level F-statistics in each iteration is addressed by the F-test method analogously to Eqn. equation 3, and we denote the training data/model outcome in each iteration as $Y^{\Phi}_{i,q}, \hat{Y}^{\Phi}_{i,q}, \hat{Y}^{\Phi}_{i|j,q}$. We adopt it to iteratively determine the window set M in Eqn. equation 5:

$$F^{t,k}_{j\Longrightarrow i} = \sum_q (\hat{Y}^{\Phi}_{i,q} - Y^{\Phi}_{i,q})^2 / \sum_q (\hat{Y}^{\Phi}_{i|j,q} - Y^{\Phi}_{i,q})^2, q \in [t, t+k-1]. \tag{6}$$

In summary, the DWGC-CI method uses NAR's prediction errors to reversely control the causality index, which can be seen as the input layer weights of the NAR model, and detect causalities in this process. The final procedure for the conclusion is outlined in Algorithm equation 1. [4]

3.3 Practical Issues

Selection of sliding window size The optimal selection of the sliding window size can be best decided based on the background knowledge of the GC period. In the following experimental part, for a monsoon investigation, we employ a monthly sliding window to investigate the causality of seasonal fluctuations, whereas, in a stock market analysis, an annual sliding window is utilized to examine the causality of yearly variations. We will illustrate in Appendix. E.1 that, when our choice of sliding window size is clearly not in line with a priori intuition, it is likely that we will not reach an informative conclusion about GC.

[3]The choice of h is not unique and in this paper we set $h(\cdot) = (\alpha - \tanh(\cdot)), \alpha > 1$.

[4]In addition, compared with the winsorization Dixon & Yuen (1974) method, our causality index can effectively capture the causal information of each window, while winsorization only concentrates on several points with abnormal prediction errors in each iteration.

In cases where background knowledge is not accessible, our approach relies on previously established sliding window size selection methods found in the time series forecasting literature, which could potentially yield suboptimal outcomes. These methods encompass: 1) preemptively categorizing point-in-time sample features to ensure that samples sharing similar attributes are grouped within the same window; 2) introducing minor perturbations to the dataset and evaluating the robustness of DWGC-CI, with the optimal window length corresponding to the most robust result; and 3) implementing a ROC-based technique to determine the appropriate sliding window size, as detailed in reference Fkih et al. (2012).

Precision-recall curve The precision and recall rates are both intimately related to the selection of the specific threshold value, denoted as ϵ. As ϵ increases, and the likelihood of a window being accepted as window-level GC diminishes, leading to a decrease in the recall rate. Simultaneously, due to the conservative nature of the decision criteria, the precision rate is expected to increase correspondingly. A thorough precision-recall trade-off with ϵ will be presented in Appendix. E.2, in which we also show that DWGC-CI contains a relatively low false-positive rate without GC ground truth between two series.

4 Theoretical Analysis

In this section, we give the theoretical analysis of our methods. Intuitively, the robustness and stability of DWGC are in direct proportion to the window length and can be enhanced by the causal index (CI) based on the view of negative feedback regulation. Mathematically, we re-organized the statement into two theorems: 1) DWGC's recall rate increases with window length and degenerates to GC's, and 2) DWGC-CI's recall rate surpasses DWGC's in window length and GC's in channel length.

For definitions, we analytically consider the series channel length as L and the sample pairs T_j, T_i. Naturally, the recall rate as below refers to the overall probability of detecting $F_{j\Longrightarrow i} > \epsilon$ on retrieved windows with causalities. The definition of accuracy is analogous.

4.1 DWGC: granularity-efficacy trade-off

We first analyze the granularity-efficacy trade-off of DWGC. We choose the sliding window length as granularity and the recall rate as efficacy. We refer readers to some additional criteria of efficacy (such as accuracy rate) in the experiment part.

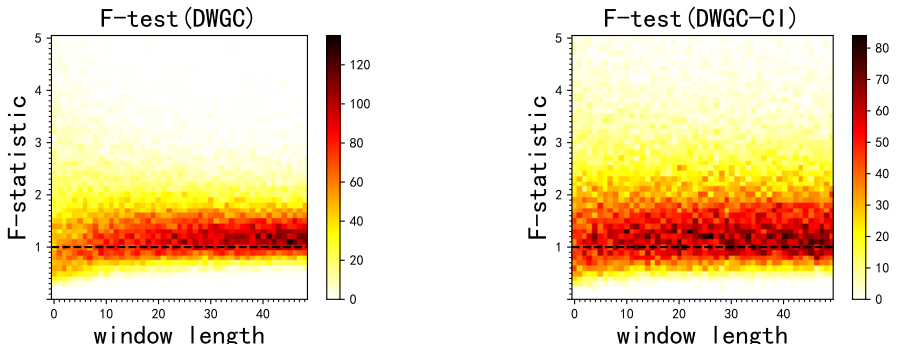
Assumption 1. *The output of the machine learning model $g(\cdot)$ approximates a Gaussian distribution. Namely, $g_i(Y_{i,<t}) \sim \mathcal{N}(Y_{i,t}, \cdot)$, $g_i(Y_{i,<t}, Y_{j,<t}) \sim \mathcal{N}(Y_{i,t}, \cdot)$.*

Theorem 1 (Positive correlation between recall rate and window length of DWGC)**.** *Suppose Assumption. equation 1 holds. Assume the channel level GC exists (we have enough confidence to reject it only when $F_{j\Longrightarrow i} < \epsilon$). Then the DWGC's recall rate satisfies $\mathbb{P}(F^{t,k_2}_{j\Longrightarrow i} > \epsilon) - \mathbb{P}(F^{t,k_1}_{j\Longrightarrow i} > \epsilon) = \mathcal{O}(\frac{k_2-k_1}{2^{k_1}})$. Specifically, $\forall k_1, k_2 \in [0, L], k_1 < k_2$,*

$$\mathbb{P}(F^{t,k_2}_{j\Longrightarrow i} > \epsilon) - \mathbb{P}(F^{t,k_1}_{j\Longrightarrow i} > \epsilon) \in \left[0, C(k_2 - k_1)2^{-k_1}\right], \textit{where } C \textit{ is a constant.} \tag{7}$$

Corollary 1 (Granularity-efficacy trade-off)**.** *DWGC's recall rate monotonically increases with (efficacy increases) the sliding window length k (granularity decreases). Naturally, DWGC degenerates to the traditional GC method in the process of window length k approaching L.*

The strict proof is referred to in Appendix B-D. This theorem encapsulates a critical insight: directly transitioning the GC detection from channel-level to window-level may imply a significant compromise in performance, particularly with regard to the recall rate. Owing to the "*trade-off*" between the granularity of the results and the efficacy, we are compelled to propose the DWCI-CI method. We endeavour to strike this "*trade-off*" by introducing additional CI, thereby controlling local noise. We assert that, through the appropriate selection of CI, the performance of DWGC-CI will surpass that of DWGC, which will be illustrated in the following part. We also refer readers to Fig. equation 2 for deeper comprehension.



(a) (b)

Figure 2: (a) shows the granularity-efficacy trade-off of DWGC on a simulation dataset (NAR dataset generation in Sec 5.1). The recall rate $P(F^{t,k}_{j\Longrightarrow i} > \epsilon = 1)$ is monotonically increasing with the window length k, and when k is large (e.g., $k > 20$), its performance stabilises quickly. This is consistent with the monotonicity and convergence property as reflected in Corollary. 1. Based on this insight, DWGC-CI (b) is exactly what helps to improve the performance of DWGC, especially in smaller windows (e.g., $k < 20$).

4.2 DWGC-CI: enhancing the granularity while keeping efficacy

In the above section, we introduce the granularity-efficacy trade-off implicated by DWGC. In this part, we claim the causality index (CI) can help enhance the efficacy (recall rate) while keeping the granularity (window level).

Theorem 2 (DWGC-CI outperforms DWGC)**.** *Suppose Assumption. equation 1 holds. Assume the window length is chosen as k, and the dynamic window-level GC on $[t, t+k-1]$ exists (we only have enough confidence to reject it only when $F^{t,k}_{j\Longrightarrow i} < \epsilon$). Then causality index (CI) matrix Φ exists to make DWGC-CI's recall rate outperform DWGC's.*

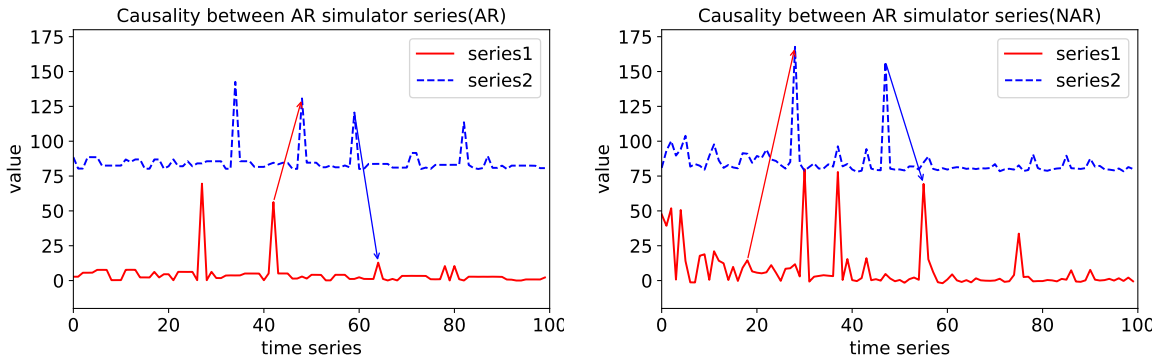
We prove it via constructing sufficient conditions and leave the proof in Appendix C. Intuitively, we impose smaller CI to the time point with local fluctuations (noise/overfit) in order to make stabilization, and hence the F-test is more informative and the recall rate is enhanced. Seen from Fig. equation 2, most F-statistics of DWGC-CI are steadily higher than DWGC's on the same window length k, especially when k is relatively small (e.g., $k < 20$). Moreover, in the process of training approximate Φ, we adopt the sufficient condition of Φ as the regularization items for Eqn. equation 5 (see Appendix. C for detail) and conduct automatic differentiation.

5 Experiments

In this section, we employ our proposed techniques to analyze two synthetic and two real-world datasets. As previously elucidated in the preliminary discussion, the integration of window-level and non-linear approaches represent a novel avenue of research, which consequently leads to a scarcity of well-established baselines. To ensure rigorous and impartial evaluation, our experimental analysis primarily concentrates on validating the theoretical propositions (Sec. 5.1) and assessing the congruence with a priori knowledge (Sec. 5.2).

5.1 Experiment on synthetic AR/NAR simulation dataset

We apply DWGC/DWGC-CI on synthetic (N)AR dataset and choose Rolling-window GC Ming-Hsien & Chih-She (2015) methods as the relatively best baseline (shown in related work). These results further verify the theoretical results in the last section.



(a) Linear AR simulation.

(b) NAR simulation.

Figure 3: The partial schematic in AR/NAR experiment. The blue/red curve represents two channels of time series (series 1: T_1, series 2: T_2), and arrows represent causalities. The starting position of each causality arrow is at the point when m_1 (m_2) is abnormally large. Noting that some mutations are caused by noise instead of causalities in our construction, which are served as confounding factors.

5.1.1 Experimental setting

The AR and NAR simulations generate two synthetic datasets, and the generation process is detailed as follows.

1) *The linear AR simulation construction*: We simulate two linear AR time series with a lag value randomly picked from the integers from one to nine ($t = 0, 1, ...$): $T_{1,t} = m_1 T_{2,t-lag} + 0.02n_1(t)$, $T_{2,t} = m_2 T_{1,t-lag} + 0.02n_2(t)$. For $i = 1, 2$, $n_i(t)$ is the Gaussian noise term ($n_i(t) \sim N(0, 1)$), and the initial values are $T_{i,t} = t\mathbb{1}_{t\in[1,5]} + (11-t)\mathbb{1}_{t\in[6,10]}, m_i = 0.9\mathbb{1}_{rand(0,1)\leq 0.95} + 10\mathbb{1}_{rand(0,1)>0.95}$.[5] **2)** *The NAR simulation construction*: we simulate two NAR time series: $T_{1,t} = \mathrm{Re}(m_1\sqrt{{T_{2,t-lag}}^2 - 1 + n_1(t)})$, $T_{2,t} = \mathrm{Re}(m_2\sqrt{{T_{1,t-lag}}^2 - 1 + n_2(t)})$, where $n_1(t)$, $n_2(t)$, the lag and the initial value of T_1, T_2 are the same as AR's. Let $f(s) = \sin(0.1s)$, s is random. Then the values of m_1, m_2 are $m_1 = 10\mathbb{1}_{f(s)>0.9} + 0.9\mathbb{1}_{f(s)<0.9}$, $m_2 = 10\mathbb{1}_{f(s)<-0.9} + 0.9\mathbb{1}_{f(s)>-0.9}$. The time point of $m_i = 10, i = 1, 2$ denotes the beginning of the causality[6]. At last, $\mathrm{Re}(\cdot)$ takes the real part of the square root.

Table 1: Recall (left) and accuracy (right) of three window-level GC methods on two AR/NAR simulation datasets.

dataset	k / method	10	20	30	100
AR	Rolling-window GC (baseline)	0.20 0.44	0.29 0.52	0.53 **0.99**	0.75 **0.99**
	DWGC (ours)	0.59 0.42	0.60 0.76	0.73 0.93	0.87 **0.99**
	DWGC-CI (ours)	**0.84** 0.44	**0.90 0.79**	**0.84** 0.94	**0.88 0.99**
NAR	Rolling-window GC (baseline)	0.22 **0.25**	0.47 0.42	0.40 0.61	0.75 0.70
	DWGC (ours)	0.48 0.22	0.49 0.43	0.77 0.82	0.77 0.90
	DWGC-CI (ours)	**0.72** 0.23	**0.72 0.45**	**0.80 0.84**	**0.88 0.92**

5.1.2 Experimental process and result

For the two AR/NAR simulation datasets, we consider the case of $m_1, m_2 = 10$ as the beginning of the causality. To determine ϵ and the parameter of the function $h(\cdot)$, we use the 5-fold cross-validation, during which we randomly choose one hundred time points as the test set and the other four hundred as the training set. We choose the function $h(\cdot) = (\frac{6}{5} - \tanh(\cdot))$ and the causality threshold $\epsilon = 1$ (the sensitivity test of ϵ is

[5]Let $\mathcal{X}$ be the set of possible values of x, then $\mathbb{1}_{x\in\mathcal{X}}$ denotes the characteristic function. Moreover, $rand(0, 1)$ denotes the random values ranging from 0 to 1.

[6]This chronological construction is consistent with the very nature of GC, i.e., the 'precedence' principle as we illustrated in preliminaries.

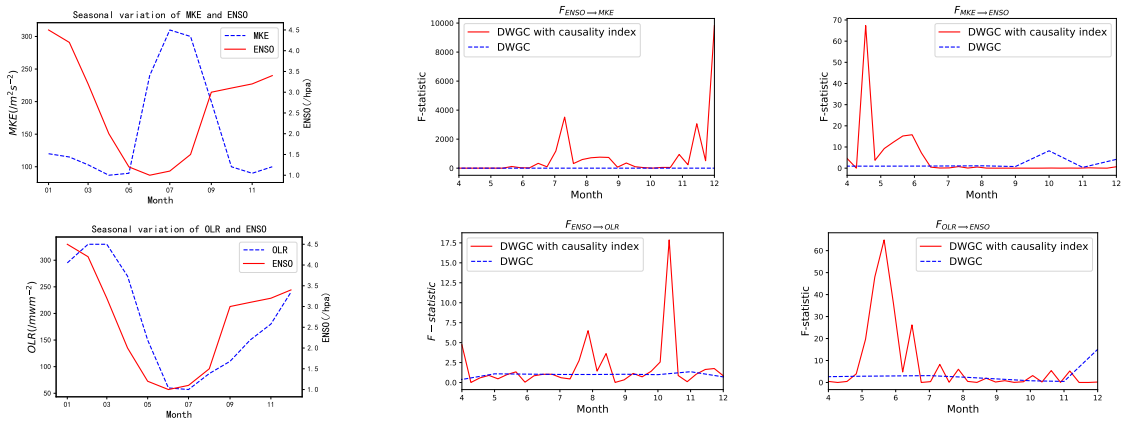


Figure 4: The experimental results on Climate Dataset for ElNiño by DWGC and DWGC-CI methods. The first column is the original data of MKE & OLR and ENSO (1992). The second and third column is the F-test result between ENSO and MKE & OLR using DWGC and DWGC-CI respectively. In all mentioned cases, our DWGC-CI method is closer to the prior knowledge of Kumar's Kumar et al. (1999) and Yasunari's Yasunari (1990).

in Appendix. D). Moreover, the rolling window intervals are taken as $\{[i*k,(i+1)k] : i*k \in [0,L], i \in N\}$, where $k = 10, 20, 30, 100$.

Fig. equation 3 shows the partial schematic of AR/NAR simulations, where the arrows mean the lag relation between the two series and we consider them as the ground truth of the dynamic causalities. Following the definition in Section 4, if $F^{t,k}_{2\Longrightarrow 1} > 1$ or $F^{t,k}_{1\Longrightarrow 2} > 1$ is detected on a window with at least a set of causality pairs, our causality extraction on this window is successful.

In this case, we calculate the recall/accuracy of our methods in Table equation 1 after making stationary processing. As seen from the results, in terms of the recall and accuracy rate, the DWGC-CI method performs better than DWGC, at least at these sampling window lengths k, and better than the baseline Ming-Hsien & Chih-She (2015), especially at $k = 20, 100$. Besides, The recall rate of the DWGC method increases with the sliding window length. These results are consistent with the Theorem. equation 1 and Theorem. equation 2.

5.2 Experiment on climate dataset for ElNiño

In this part, we verify our methods on a real-world climate dataset, which has widely recognized seasonal causalities. Due to space limitation, we refer readers to Appendix. A for domain-specific knowledge.

5.2.1 Experimental processing and result

ENSO and Asian monsoon data in 1992 Yang et al. (2018) can be seen from the first column in Fig. equation 4. The first four months are taken as the training data and the rest as the testing data. In order to facilitate comparison with prior knowledge, we selected the window length k as one month.

We use our model to judge the window-level causalities between ENSO and MKE & OLR every month. In the second column of Fig. equation 4, we show the F-statistic of every month of DWGC-CI and DWGC for two series (ENSO to MKE) and (ENSO to OLR). The x-axis is the month-time and the y-axis is the F-statistic. By DWGC, the causality between ENSO and MKE & OLR can hardly be successfully detected, not to mention the difference of causality between May-Aug (spring & summer) and Aug-Dec (autumn & winter). However, by DWGC-CI, we not only obtain the basic causalities of ENSO $\Longrightarrow$ MKE and ENSO $\Longrightarrow$ OLR, but also find that the causality in Sep-Dec (autumn & winter) is more significant than that in May-Aug (spring & summer). This significant causality variation detected by DWGC-CI is more consistent with the academic knowledge Kumar et al. (1999) than that by DWGC. In the third column of Fig. equation 4, we show the F-statistic of each month of DWGC and DWGC-CI for two series (MKE to ENSO) and (OLR

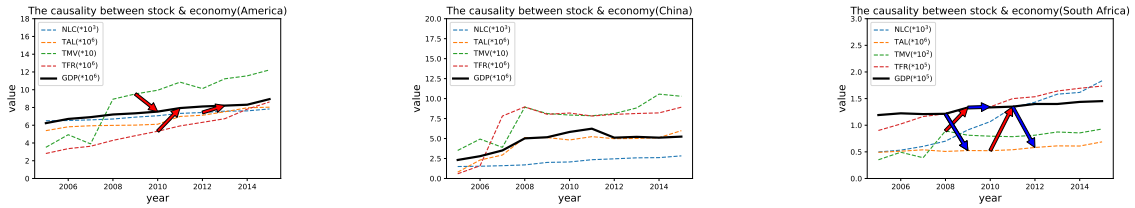


Figure 5: The causality between the stock market and economy by DWGC-CI method. Three subfigures show the trend of five variables (NLC, TAL (million), TMV (‰), TFR (million), GDP (million)) during 2007-2015 in three countries. The arrows represent causalities between different curves (the red is to GDP while the blue is from GDP). Results of our DWGC method are consistent with academic knowledge in three areas. Detailed F-statistic results are shown in Table 2,3,4.

to ENSO). By DWGC, we detect the causality MKE & OLR $\Longrightarrow$ ENSO in Sep-Dec (autumn & winter). However, by DWGC-CI, we detect the causality MKE & OLR $\Longrightarrow$ ENSO in May-Aug (spring & summer). Compared to the DWGC method, our causalities detected by DWGC-CI are more aligned with the academic knowledge Yasunari (1990).

5.3 Experiment on Stock Market and Economy

In this part, we use DWGC and DWGC-CI methods to detect causality between the stock market and the economy of three representative economies: the United States, China and South Africa. The a priori knowledge is shown in Appendix. A.

5.3.1 Experimental processing and result

In this experiment, we choose four widely-acknowledged indexes Edou (2017) to jointly evaluate the prosperity of the stock market: the number of listed companies (NLC), the total assets listed (TAL), the total market value (TMV), and the total funds raised (TFR). On the other hand, Economic growth is measured in terms of gross domestic product (GDP). Considering the authorization limit of data sources, we selected data from three countries from 2005-2015 (5 dimensions in total) and choose 2005-2007 among them as the training data. For the sake of comparing with the prior knowledge, we set the window length k as one year to observe the annual change trend of stock-economy causalities in three countries, and causal threshold ϵ = 500.

In this experiment, we test causalities between GDP and NLC, TAL, TMV, and TFR by DWGC-CI and DWGC methods, and the DWGC-CI results are illustrated by red/blue arrows in figure. equation 5. The results show that 1) in America, causalities NLC (2012), TMV (2009), TFR (2010) $\Longrightarrow$ GDP exist, 2) there are no specific causalities between the stock market and economy in China, and 3) bidirectional causalities TMV (2008), TAL (2010) $\Longrightarrow$ GDP, GDP $\Longrightarrow$ TAL (2008), TFR (2009,2012) exist in South Africa. However, the DWGC results don't show such a good correlation with the prior conclusions. In all, these causalities between stock & market of three areas in fig. equation 5 are consistent with our academic knowledge Pearce (1983); Aziz & Duenwald (2006); Ghirmay (2004) respectively by DWGC-CI, better than by DWGC. The experimental results on the stock market and economy are in table 1,2,3. In these tables, the data beyond the causal threshold $\epsilon = 500$ are bolded, which matches the causal arrow in Fig (5).

Table 2: $F_{X \Longrightarrow GDP}$ (left) and $F_{GDP \Longrightarrow X}$ (right) of America.

X	F-test / method	2008	2009	2010	2011	2012	2013	2014
NLC	DWGC-CI	32.67/1.34	3.65/2.56	289.09/2.68	2.30/21.34	**1269.82**/390.30	51.11/330.92	29.59/99.81
	DWGC	4.35/1.17	3.64/1.54	9.74/0.28	0.22/5.62	22.50/26.10	5.92/13.71	5.32/9.66
TAL	DWGC-CI	0.64/23.02	1.03/169.64	1.00/0.23	1.00/1.68	1.00/1.02	1.00/0.99	1.00/1.00
	DWGC	1.29/5.03	10.23/10.14	1.60/0.48	0.020/1.29	12.55/1.01	2.21/0.99	2.14/1.00
TMV	DWGC-CI	15.42/0.82	**9480.66**/0.99	0.78/0.02	0.0036/0.024	91.19/0.0056	0.93/0.038	0.96/0.062
	DWGC	4.20/6.8	41.38/169.6	0.86/0.56	0.059/3.13	3.76/1.08	1.08/3.09	1.03/2.07
TFR	DWGC-CI	28.37/2.05	3.49/1.17	**616.96**/2.47	0.078/0.13	20.81/0.17	411.65/1.67	11.23/5.56
	DWGC	3.94/0.07	3.71/0.31	17.10/0.43	0.036/1.93	3.68/1.47	9.11/10.86	3.23/7.12

Table 3: $F_{X \Longrightarrow GDP}$ (left) and $F_{GDP \Longrightarrow X}$ (right) of China.

X	F-test / method	2008	2009	2010	2011	2012	2013	2014
NLC	DWGC-CI	0.77/0.95	3.66/2.06	1.01/0.72	1.01/1.00	0.99/0.99	0.98/1.00	0.99/1.00
	DWGC	0.87/1.20	2.20/3.34	1.01/2.65	1.00/5.93	0.99/16.86	0.99/17.71	0.99/5.47
TAL	DWGC-CI	2.28/0.88	0.22/1.23	1.12/0.99	1.05/1.01	0.90/0.97	0.87/0.94	0.90/1.05
	DWGC	1.70/0.94	0.50/1.11	1.06/0.99	1.03/1.01	0.95/0.98	0.930.97	0.95/1.02
TMV	DWGC-CI	0.27/0.92	0.06/1.00	0.86/0.98	0.93/0.98	1.18/0.98	1.26/1.15	1.18/1.01
	DWGC	0.33/1.00	0.40/9.35	0.86/1.01	0.93/1.01	1.20/1.01	1.30/0.96	1.19/1.00
TFR	DWGC-CI	9.94/1.40	0.02/3.13	0.83/1.55	0.90/2.26	1.33/0.02	1.33/0.02	1.23/4.36
	DWGC	2.83/1.16	0.10/1.73	0.93/1.26	0.95/1.60	1.15/0.13	1.12/0.07	1.09/2.47

Table 4: $F_{X \Longrightarrow GDP}$(left) and $F_{GDP \Longrightarrow X}$(right) of South Africa.

X	F-test / method	2008	2009	2010	2011	2012	2013	2014
NLC	DWGC-CI	1.47/414.05	0.042/4.43	0.87/2.58	0.14/1.00	8.97/1.00	54.10/1.00	3.10/1.00
	DWGC	1.22/0.023	0.20/0.027	0.94/2.40	0.36/44.09	3.01/4.99	7.59/4.58	1.77/3.7
TAL	DWGC-CI	1.98/**8245.55**	1.10/150.84	**772.42**/0.45	0.0047/0.21	137.48/1.22	162.61/1.01	4.03/1.00
	DWGC	1.36/0.85	3.93/0.03	15.82/0.57	0.19/0.05	12.38/11.09	12.25/2.27	1.85/2.41
TMV	DWGC-CI	**2730.82**/1.65	0.0018/297.02	0.49/0.27	0.0012/0.38	0.22/0.0039	0.27/2.53	0.36/4.72
	DWGC	53.10/1.31	0.41/14.96	0.87/0.55	0.25/0.63	0.49/0.19	0.51/1.47	0.64/1.88
TFR	DWGC-CI	1.81/1.33	0.45/**1205.46**	267.35/93.50	0.036/0.19	85.53/**579.83**	204.13/6.06	3.52/3.16
	DWGC	102.98/0.57	0.38/2.12	0.92/1.29	0.25/1.85	0.49/78.81	0.51/10.05	0.64/4.63

6 Conclusions and Discussions

In this paper, we focus on a new task to detect the dynamic window-level Granger causality (DWGC) between the time series. We proposed the DWGC method by conducting the F-test on comparing the window-level forecasting predictions, whereas there is demonstrated to be a granularity-efficacy tradeoff and DWGC's power of window-level GC extraction is limited. We then introduce a "causality index (CI)" method to weigh the original time series, whose purpose is to decrease the local fluctuations of the window-level F-test. The full method is DWGC-CI and is theoretically proven to have better recall rates. In the experiments on two synthetic and two real datasets, the DWGC-CI method successfully detects the window causalities, which cannot be accomplished by GC. It also outperforms DWGC, both on recall rate and accuracy rate.

This is fundamental work with many areas for future improvement and expansion. Subsequently, we will be interested in further detecting the principle of DWGC-CI on 1) window adjustment, such as dynamic rolling window intervals with different starting points, 2) weight strategy enhancement, such as assigning dynamic causality index with domain-specific knowledge, 3) graph extension, such as constructing time series causal graph combined with additional graph theory, and 4) further thoughts on the mechanism design, such as how to strike the granularity-efficacy trade-off in a more general intervention/reweighting framework.

References

Leonardo Angelini, Mario Pellicoro, and Sebastiano Stramaglia. Granger causality for circular variables. *Physics Letters A*, 373(29):2467–2470, 2009.

Jahangir Aziz and Christoph Duenwald. Growth-financial intermediation nexus in china. *Imf Working Papers*, 02(194), 2006.

Sezen Cekic, Didier Grandjean, and Olivier Renaud. Time, frequency, and time-varying granger-causality measures in neuroscience. *Statistics in Medicine*, 37(11):1910–1931, 2018.

S. Chen, S. A. Billings, and P. M. Grant. Non-linear system identification using neural networks. *International Journal of Control*, 51(6):1191–1214, 2000.

Yonghong Chen, Govindan Rangarajan, Jianfeng Feng, and Mingzhou Ding. Analyzing multiple nonlinear time series with extended granger causality. *Physics Letters A*, 324(1):26–35, 2004.

Brad Comincioli. The stock market as a leading indicator: An application of granger causality. *University avenue undergraduate journal of economics*, 1(1):1, 1996.

Wilfrid J Dixon and Kareb K Yuen. Trimming and winsorization: A review. *Statistische Hefte*, 15(2-3): 157–170, 1974.

Andrea Duggento, Maria Guerrisi, and Nicola Toschi. Echo state network models for nonlinear granger causality. *bioRxiv*, pp. 651679, 2019.

TCHIDI Guillaume Edou. Stock market and economic growth: A comparative study – evidences from china, usa, and south africa. *INTERNATIONAL JOURNAL OF ECONOMICS AND FINANCIAL ISSUES*, 2017.

Michael Eichler. *Causal inference in time series analysis*. Wiley Online Library, 2012.

Michael Eichler. Causal inference with multiple time series: principles and problems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1997):20110613, 2013.

Fethi Fkih, Mohamed Nazih Omri, et al. Learning the size of the sliding window for the collocations extraction: A roc-based approach. In *Proceedings of the 2012 International Conference on Artificial Intelligence (ICAI'12)*, pp. 1071–1077, 2012.

K.J. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *NeuroImage*, 19(4):1273 – 1302, 2003. ISSN 1053-8119. doi: https://doi.org/10.1016/S1053-8119(03)00202-7. URL `http://www.sciencedirect.com/science/article/pii/S1053811903002027`.

Teame Ghirmay. Financial development and economic growth in sub-saharan african countries: Evidence from time series analysis. *African Development Review*, 16(3):415–432, 2004. doi: 10.1111/j.1017-6772.2004.00098.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1017-6772.2004.00098.x`.

Clive William John Granger and Paul Newbold. *Forecasting economic time series*. Academic Press, 2014.

Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pp. 424–438, 1969.

C.W.J. Granger. Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, 2:329 – 352, 1980. ISSN 0165-1889. doi: https://doi.org/10.1016/0165-1889(80)90069-X. URL `http://www.sciencedirect.com/science/article/pii/016518898090069X`.

James D Hamilton. *Time series analysis*, volume 2. Princeton New Jersey, 1994.

David Hammarwall, Mats Bengtsson, and Björn Ottersten. Acquiring partial csi for spatially selective transmission by instantaneous channel norm feedback. *IEEE Transactions on Signal Processing*, 56(3): 1188–1204, 2008.

Godfrey Harold Hardy. *Orders of infinity: The'infinitärcalcül'of Paul du Bois-Reymond*. University Press, 1924.

Miao He, Weixi Gu, Ying Kong, Lin Zhang, Costas J Spanos, and Khalid M Mosalam. Causalbg: Causal recurrent neural network for the blood glucose inference with iot platform. *IEEE Internet of Things Journal*, 2019.

Robert K Kaufmann and David I Stern. Evidence for human influence on climate from hemispheric temperature relations. *Nature*, 388(6637):39–44, 1997.

Eamonn Keogh and Shruti Kasetty. On the need for time series data mining benchmarks: a survey and empirical demonstration. *Data Mining and knowledge discovery*, 7(4):349–371, 2003.

K Krishna Kumar, Balaji Rajagopalan, and Mark A Cane. On the weakening relationship between the indian monsoon and enso. *Science*, 284(5423):2156–2159, 1999.

Yang Li, Hua-Liang Wei, Steve A Billings, and Xiao-Feng Liao. Time-varying linear and nonlinear parametric model for granger causality analysis. *Physical Review E*, 85(4):041906, 2012.

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

Huanfei Ma, Kazuyuki Aihara, and Luonan Chen. Detecting causality from nonlinear dynamics with short-term time series. *Scientific reports*, 4(1):1–10, 2014.

Atalanti A Mastakouri, Bernhard Schölkopf, and Dominik Janzing. Necessary and sufficient conditions for causal feature selection in time series with latent common causes. In *International Conference on Machine Learning*, pp. 7502–7511. PMLR, 2021.

YANG Ming-Hsien and WU Chih-She. Revisit export and gdp nexus in china and taiwan: A rolling window granger causality test. *Theoretical & Applied Economics*, 22(3), 2015.

Manfred Mudelsee. *Climate time series analysis*, volume 30. Springer, 2013.

Mattia F. Pagnotta, Mukesh Dhamala, and Gijs Plomp. Benchmarking nonparametric granger causality: Robustness against downsampling and influence of spectral decomposition parameters. *NeuroImage*, 183: 478 – 494, 2018. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2018.07.046. URL `http://www.sciencedirect.com/science/article/pii/S1053811918306621`.

John Pearce. The relationship of internal versus external orientations to financial measures of strategic performance. *Strategic Management Journal*, 4:297 – 306, 10 1983. doi: 10.1002/smj.4250040402.

Judea Pearl. Theoretical impediments to machine learning with seven sparks from the causal revolution. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 3–3, 2018.

Judea Pearl and James Robins. Probabilistic evaluation of sequential plans from causal models with hidden variables. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI'95, pp. 444–453, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc. ISBN 1558603859.

CS Ramage. guillemotright monsoon meteorology. *International Geophysics Series*, 15, 1971.

Louis L Scharf. *Statistical signal processing*, volume 98. Addison-Wesley Reading, MA, 1991.

Linda Sommerlade, Marco Thiel, Bettina Platt, Andrea Plano, Gernot Riedel, Celso Grebogi, Jens Timmer, and Björn Schelter. Inference of granger causal time-dependent influences in noisy multivariate time series. *Journal of neuroscience methods*, 203(1):173–185, 2012.

George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *science*, 338(6106):496–500, 2012.

Xiaohai Sun. Assessing nonlinear granger causality from multivariate time series. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 440–455. Springer, 2008.

E Syczewska and Z Struzik. Granger causality and transfer entropy for financial returns. *Acta Physica Polonica A*, 127(3A), 2015.

Floris Takens. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pp. 366–381. Springer, 1981.

Alex Tank, Ian Covert, Nicholas Foti, Ali Shojaie, and Emily Fox. Neural granger causality for nonlinear time series. *arXiv preprint arXiv:1802.05842*, 2018.

Alexey Tsymbal. The problem of concept drift: definitions and related work. *Computer Science Department, Trinity College Dublin*, 106(2):58, 2004.

Sebastian Weichwald, Martin E Jakobsen, Phillip B Mogensen, Lasse Petersen, Nikolaj Thams, and Gherardo Varando. Causal structure learning from time series: Large regression coefficients may predict causal links better in practice than small p-values. In *NeurIPS 2019 Competition and Demonstration Track*, pp. 27–36. PMLR, 2020.

John Woods and Douglas Walton. Post hoc, ergo propter hoc. *The Review of Metaphysics*, pp. 569–593, 1977.

Alexander P Wu, Rohit Singh, and Bonnie Berger. Granger causal inference on dags identifies genomic loci regulating transcription. In *International Conference on Learning Representations*, 2022.

Chenxiao Xu, Hao Huang, and Shinjae Yoo. Scalable causal graph learning through a deep neural network. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 1853–1862. ACM, 2019.

Hongteng Xu, Mehrdad Farajtabar, and Hongyuan Zha. Learning granger causality for hawkes processes. In *International conference on machine learning*, pp. 1717–1726. PMLR, 2016.

Song Yang, Kaiqiang Deng, and Wansuo Duan. Selective interaction between monsoon and enso: Effects of annual cycle and spring predictability barrier. *Chinese Journal of Atmospheric Sciences*, 2018.

T Yasunari. Impact of indian monsoon on the coupled atmosphere/ocean system in the tropical pacific. *Meteorology and Atmospheric Physics*, 44(1-4):29–41, 1990.

Yaoyu Zhang, Yanyang Xiao, Douglas Zhou, and David Cai. Granger causality analysis with nonuniform sampling and its application to pulse-coupled nonlinear dynamics. *Physical Review E*, 93(4):042217, 2016.

Ethics Statement

This study is conducted in accordance with the general academical principles and adhered to the corresponding ethical guidelines and standards. The research design, procedures, and data collection methods were reviewed and approved by the corresponding institutes of authors.

All participants involved in the study were informed of the study's objectives, procedures, potential risks, and benefits. Written informed consent was obtained from each participant prior to their involvement in the research. Participants were assured of the confidentiality of their personal information and were informed that they had the right to withdraw from the study at any time without any consequences.

All data collected in this study were anonymized to protect the privacy of the participants. Any identifiable information was removed, and the data were securely stored in compliance with the data protection policies. Access to the data was granted only to the research team members involved in the study.

The authors declare that they have no competing interests or conflicts of interest related to this research. The study was conducted with the utmost respect for the participants and their rights, ensuring their welfare and privacy throughout the research process.

Supplemental Materials

A Domain-specific knowledge in experiments

A.1 Academic knowledge of ElNiño and monsoon

The climate academia already have the following definition on ElNiño: a climate phenomenon in the Pacific equatorial belt where the ocean and atmosphere interact with each other and lose their balance Ramage (1971), while monsoon generally eqrefers to the seasonal conversion of atmospheric circulation and precipitation in tropical and subtropical areas, which is measured by two parameters: OLR (Outgoing Longwave Radiation) and MKE (Monsoon Kinetic Energy).

The causalities between ENSO and the East Asian monsoon (MKE, OLR) have been extensively explored. Up till now, it is widely believed that the causal direction and intensity between the two has seasonal variation summarized as below:

1. The causality ENSO $\Longrightarrow$ MKE & OLR exists, more intense in autumn & winter than in spring & summer.
2. The reverse causality MKE & OLR $\Longrightarrow$ ENSO exists in spring & summer.

However, with the increasing instability and complexity of ENSO, the above are only observational conclusions based on meteorology knowledge and lacks highly interpretable causality analysis. In all, a prior-knowledge-fitted result in this experiment will strongly demonstrate the effectiveness of our method.

A.2 Academic knowledge of stock market and economy in America, China and South Africa

The relationship between the development of the stock market and economic growth has long been a hot topic in financial circles. It's one of the most classical cases where a mutual causal-and-effect relationship exists. These causalities are hard to be captured since they show different characteristics in various social environments and economic development stages.

In the world, the United States, China and South Africa, as three highly representative countries in the Americas, Asia and Africa, contain the empirical knowledge of stock-economy causality summarized as follows:

1 the United States: Pearce Pearce (1983) showed that the stock market was a long-term stability indicator of economic growth, i.e., Stock Market $\Rightarrow$ Economy.

2 China: Aziz Aziz & Duenwald (2006) used the panel data method, and found that China's stock market development had little impact on GDP growth during the post-1978 eqreform period.

3 South Africa: Ghirmay Ghirmay (2004) studied the causal relationship between financial development and economic growth in 13 sub-Saharan African countries by bivariate VAR model, results showed that financial development promoted economic growth in eight countries, and among which there were bidirectional causalities in six countries, i.e., Economy $\Rightarrow$ Stock Market.

B Proof of Theorem. 1

Proof. Without loss of generation, we can choose $\epsilon = 1$, and the approximate Gaussian distribution on each time point shares a similar variance as in Assumption. equation 1. $Y_{i,t}$ is partitioned into **o**riginal real data $Y^o_{i,t}$ and **n**oise $Y^n_{i,t}$ as

$$Y_{i,t} = Y^o_{i,t} + Y^n_{i,t}. \tag{8}$$

We assume that $\widehat{Y_{i,t}} \sim \mathcal{N}(Y^o_{i,t}, \sigma^2)$, $\widehat{Y_{i|j,t}} \sim \mathcal{N}(Y^o_{i,t}, \epsilon_0^2\sigma^2)$, where $\epsilon_0 \in (0,1)$, σ are constants. Moreover, the Gaussian noise satisfies $Y^n_{i,t} \sim \mathcal{N}(0, \sigma_0)$. In Equation. 3 , if we denote $\frac{\sum_q (\hat{Y}_{i,q} - Y_{i,q})^2}{\sum_q (\hat{Y}_{i|j,q} - Y_{i,q})^2} = \frac{\sigma^2+\sigma_0^2}{\epsilon_0^2\sigma^2+\sigma_0^2} X$, then we conclude that X satisfies the $\mathcal{F}$ distribution. Its density probability function is

$$f(x) = \begin{cases} 0, x \leq 0 \\ \frac{1}{\mathcal{B}(\frac{k}{2}, \frac{k}{2})} x^{\frac{k}{2}-1}(1+x)^{-k}, x > 0. \end{cases} \tag{9}$$

In our proof, we choose the threshold 1. On this basis, we have

$$\begin{aligned} P(F^{t,k}_{j \Longrightarrow i} > 1) &= \int_{\frac{\epsilon_0^2\sigma^2+\sigma_0^2}{\sigma^2+\sigma_0^2}}^{+\infty} \frac{1}{\mathcal{B}(\frac{k}{2}, \frac{k}{2})} x^{\frac{k}{2}-1}(1+x)^{-k} dx \\ &= \frac{1}{\mathcal{B}(\frac{k}{2}\frac{k}{2})} \int_{x_0}^{1} x^{\frac{k}{2}-1}(1-x)^{\frac{k}{2}-1} dx, \quad (x_0 := \frac{\epsilon_0^2\sigma^2+\sigma_0^2}{(\epsilon_0^2+1)\sigma^2+2\sigma_0^2} \in (0, \frac{1}{2})) \\ &=: 1 - I_{x_0}(\frac{k}{2}, \frac{k}{2}) \end{aligned} \tag{10}$$

We introduce the following lemma:

Lemma 1. *If $x_0 \in (0, \frac{1}{2})$, then*

$$\begin{aligned} I_{x_0}(\frac{k}{2}, \frac{k}{2}) &= \frac{\Gamma(k)}{\Gamma(\frac{k}{2}+1)\Gamma(\frac{k}{2})}(1-2x_0)x_0^{\frac{k}{2}}(1-x_0)^{\frac{k}{2}} + I_{x_0}(\frac{k+2}{2}, \frac{k+2}{2}) \\ &\geq I_{x_0}(\frac{k+2}{2}, \frac{k+2}{2}). \end{aligned} \tag{11}$$

We make contradictions. If there is $k = i_0$ such that $I_{x_0}(\frac{i_0}{2}, \frac{i_0}{2}) > I_{x_0}(\frac{i_0-1}{2}, \frac{i_0-1}{2})$. Then we have

$$\begin{aligned} & I_{x_0}(\frac{i_0+2}{2}, \frac{i_0+2}{2}) - I_{x_0}(\frac{i_0+1}{2}, \frac{i_0+1}{2}) \\ \geq & I_{x_0}(\frac{i_0}{2}, \frac{i_0}{2}) - I_{x_0}(\frac{i_0-1}{2}, \frac{i_0-1}{2}) + \frac{\Gamma(k-1)(1+\frac{2(k-1)}{k+1}x_0)x_0^{\frac{k-1}{2}}(1-x_0)^{\frac{k-1}{2}}}{\Gamma(\frac{k-1}{2}+1)\Gamma(\frac{k-1}{2})} \\ & - \frac{\Gamma(k)(1+\frac{2k}{k+2}x_0)x_0^{\frac{k}{2}}(1-x_0)^{\frac{k}{2}}}{\Gamma(\frac{k}{2}+1)\Gamma(\frac{k}{2})} \\ = & I_{x_0}(\frac{i_0}{2}, \frac{i_0}{2}) - I_{x_0}(\frac{i_0-1}{2}, \frac{i_0-1}{2}). \end{aligned} \tag{12}$$

Hence we have $0 = \lim_{k \to +\infty} |I_{x_0}(\frac{k+1}{2}, \frac{k+1}{2}) - I_{x_0}(\frac{k}{2}, \frac{k}{2})| \geq I_{x_0}(\frac{i_0}{2}, \frac{i_0}{2}) - I_{x_0}(\frac{i_0-1}{2}, \frac{i_0-1}{2}) > 0$. It's contradictive! Therefore, we have $\forall k, I_{x_0}(\frac{k}{2}, \frac{k}{2}) > I_{x_0}(\frac{k+1}{2}, \frac{k+1}{2})$. Thus

$$\forall k, \ P(F^{t,k}_{j \Longrightarrow i} > 1) < P(F^{t,k+1}_{j \Longrightarrow i} > 1). \tag{13}$$

Moreover, according to lemma. 1, we further have

$$\begin{aligned} 0 < \mathbb{P}(F^{t,k_2}_{j \Longrightarrow i} > 1) - \mathbb{P}(F^{t,k_1}_{j \Longrightarrow i} > 1) &= \sum_{i=k_1}^{k_2} [I_{x_0}(\frac{i}{2}, \frac{i}{2}) - I_{x_0}(\frac{i+1}{2}, \frac{i+1}{2})] \\ &= \sum_{i=k_1}^{k_2} \mathcal{O}\left(\frac{\Gamma(i)}{\Gamma(\frac{i}{2}+1)\Gamma(\frac{i}{2})}(1-2x_0)x_0^{\frac{i}{2}}(1-x_0)^{\frac{i}{2}}\right) \\ &< C\left((\frac{1}{4})^{\frac{k_1}{2}}(k_2-k_1)\right). \end{aligned} \tag{14}$$

Proved.

□

B.1 The proof of lemma. 1

Proof. Via integration by parts, we have

$$
\begin{aligned}
& I_{x_0}(p,q) - I_{x_0}(p+1,q) \\
= & \frac{\int_0^{x_0} x^{p-1}(1-x)^{q-1}}{\mathcal{B}(p,q)} - \frac{\int_0^{x_0} x^{p}(1-x)^{q-1}}{\mathcal{B}(p+1,q)} \\
= & \frac{1}{\mathcal{B}(p,q)}[\int_0^{x_0} x^{p-1}(1-x)^{q-1}dx - \frac{q+p}{p}\int_0^{x_0} x^{p}(1-x)^{q-1}dx] \\
= & \frac{1}{\mathcal{B}(p,q)}[\int_0^{x_0} x^{p-1}(1-x)^{q}dx - \frac{q}{p}\int_0^{x_0} x^{p}(1-x)^{q-1}dx] \\
= & \frac{1}{\mathcal{B}(p,q)}[\int_0^{x_0} x^{p-1}(1-x)^{q}dx - (\frac{1}{p}x^p(1-x)^q \mid_0^{x_0} + \int_0^{x_0} x^{p-1}(1-x)^{q}dx)] \\
= & \frac{\Gamma(p+q)}{\Gamma(p+1)\Gamma(q)} x_0^p (1-x_0)^q.
\end{aligned}
\tag{15}
$$

If we choose $p = \frac{k}{2}, q = \frac{k}{2}$, then

$$
I_{x_0}(\frac{k}{2},\frac{k}{2}) = \frac{\Gamma(k)}{\Gamma(\frac{k}{2}+1)\Gamma(\frac{k}{2})} x_0^{\frac{k}{2}}(1-x_0)^{\frac{k}{2}} + I_{x_0}(\frac{k+2}{2},\frac{k}{2}) \tag{16}
$$

On the other hand, we also have

$$
\begin{aligned}
I_{x_0}(p,q) &= 1 - \frac{\int_{x_0}^{1} x^{p-1}(1-x)^{q-1}dx}{\mathcal{B}(p,q)} \\
&= 1 - \frac{\int_0^{1-x_0} x^{q-1}(1-x)^{p-1}dx}{\mathcal{B}(p,q)} \\
&= 1 - I_{1-x_0}(q,p) \\
&\overset{*}{=} 1 - I_{1-x_0}(q+1,p) - \frac{\Gamma(p+q)}{\Gamma(q+1)\Gamma(p)}(1-x_0)^q x_0^p \\
&= I_{x_0}(p,q+1) - \frac{\Gamma(p+q)}{\Gamma(q+1)\Gamma(p)}(1-x_0)^q x_0^p,
\end{aligned}
\tag{17}
$$

where $*$ denotes Eqn. equation 15. If we choose $p = \frac{k+2}{2}, q = \frac{k}{2}$, then

$$
I_{x_0}(\frac{k+2}{2},\frac{k}{2}) = -\frac{\Gamma(k+1)}{\Gamma(\frac{k}{2}+1)\Gamma(\frac{k+2}{2})} x_0^{\frac{k+2}{2}}(1-x_0)^{\frac{k}{2}} + I_{x_0}(\frac{k+2}{2},\frac{k+2}{2}). \tag{18}
$$

Combined Eqn. equation 16 and Eqn. equation 18, and due to $x_0 \in (0, \frac{1}{2})$, we conclude that

$$
\begin{aligned}
I_{x_0}(\frac{k}{2},\frac{k}{2}) &= \frac{\Gamma(k)}{\Gamma(\frac{k}{2}+1)\Gamma(\frac{k}{2})}(1-2x_0)x_0^{\frac{k}{2}}(1-x_0)^{\frac{k}{2}} + I_{x_0}(\frac{k+2}{2},\frac{k+2}{2}) \\
&> I_{x_0}(\frac{k+2}{2},\frac{k+2}{2}).
\end{aligned}
\tag{19}
$$

Proved. □

C The proof of Corollary. 1

Proof. We have

$$
\begin{aligned}
\mid P(F^{t,k}_{j\Longrightarrow i} > 1) - P(F^{t,L}_{j\Longrightarrow i} > 1) \mid =& \sum_{i=k}^{L} [I_{x_0}(\frac{i}{2}, \frac{i}{2}) - I_{x_0}(\frac{i+1}{2}, \frac{i+1}{2})] \\
=& \sum_{i=k}^{L} \mathcal{O}\left(\frac{\Gamma(i)}{\Gamma(\frac{i}{2}+1)\Gamma(\frac{i}{2})} (1-2x_0) x_0^{\frac{i}{2}} (1-x_0)^{\frac{i}{2}} \right) \\
=& \mathcal{O}\left((\frac{1}{4})^{\frac{k}{2}} (L-k) \right),
\end{aligned}
\tag{20}
$$

$\mathcal{O}$ is the asymptotic notation denoting the orders of approximation following Hardy (1924). In conclusion, if GC exists, i.e., $\epsilon_0 \in (0, 1)$, the recall rate of DWGC is positively related to the window length k and is convergent to 1 when $k \to +\infty$. Proved. □

D Proof of Theorem 3

Analogous to the above section, the re-weighted. $Y_{i,t}^{\Phi}$ is partitioned as

$$Y_{i,t}^{\Phi} := \Phi_t Y_{i,t} = \Phi_t Y_{i,t}^{o} + \Phi_t Y_{i,t}^{n}. \tag{21}$$

On this basis, we assume $\widehat{Y_{i,t}^{\Phi}} \sim \mathcal{N}(\Phi_t Y_{i,t}^{o}, \Phi_t^2\sigma^2)$, $\widehat{Y_{i|j,t}^{\Phi}} \sim \mathcal{N}(\Phi_t Y_{i,t}^{o}, \Phi_t^2\epsilon_0^2\sigma^2)$, $\Phi_t\widehat{Y_{i,t}^{n}} \sim \mathcal{N}(0, \Phi_t^2\sigma_0^2)$. Moreover, for simplification, we assume $\Phi_{t_i} \neq \Phi_{t_j}, t_i \neq t_j$, or else a small perturbation is applied. According to the generalized chi-square distribution in Hammarwall et al. (2008), the summarization satisfies the following probability density distribution:

$$\begin{aligned}
&\sum_{q=t}^{t+k-1} (\hat{Y}_{i,q}^{\Phi} - Y_{i,q}^{\Phi})^2 \sim f_1(x; \Phi_t, ...\Phi_{t+k-1}) \\
&:= \sum_{q=t}^{t+k-1} \frac{e^{-\frac{x}{\Phi_q^2(\sigma^2+\sigma_0^2)}}}{\Phi_q^2(\sigma^2+\sigma_0^2)\prod_{j=t,j\neq q}^{t+k-1}\left(1-\frac{\Phi_j^2}{\Phi_q^2}\right)}, \text{ for } x \geq 0. \\
&\sum_{q=t}^{t+k-1} (\hat{Y}_{i|j,q}^{\Phi} - Y_{i,q}^{\Phi})^2 \sim f_2(x; \Phi_t, ...\Phi_{t+k-1}) \\
&:= \sum_{q=t}^{t+k-1} \frac{e^{-\frac{x}{\Phi_q^2(\epsilon_0^2\sigma^2+\sigma_0^2)}}}{\Phi_q^2(\epsilon_0^2\sigma^2+\sigma_0^2)\prod_{j=t,j\neq q}^{t+k-1}\left(1-\frac{\Phi_j^2}{\Phi_q^2}\right)}, \text{ for } x \geq 0.
\end{aligned} \tag{22}$$

On this basis, we construct $\{\Phi_t, ...\Phi_{t+k-1}\}$ to make $P(\frac{\sum_q(\hat{Y}_{i,q}^{\Phi}-Y_{i,q}^{\Phi})^2}{\sum_q(\hat{Y}_{i|j,q}^{\Phi}-Y_{i,q}^{\Phi})^2} > 1) > P(\frac{\sum_q(\hat{Y}_{i,q}-Y_{i,q})^2}{\sum_q(\hat{Y}_{i|j,q}-Y_{i,q})^2} > 1)$. In this sense, we equivalently transform the Theorem. 2 to the following lemma:

Lemma 2. *If we have* $\sum_{q_1=t}^{t+k-1}\sum_{q_2=t}^{t+k-1} \frac{1}{(c_0\frac{\Phi_{q_2}^2}{\Phi_{q_1}^2}+1)\prod_{j=t,j\neq q_1}^{t+k-1}(1-\frac{\Phi_j^2}{\Phi_{q_1}^2})\prod_{j=t,j\neq q_2}^{t+k-1}(1-\frac{\Phi_j^2}{\Phi_{q_2}^2})} > \int_{c_0}^{+\infty}\frac{1}{\mathcal{B}(\frac{k}{2},\frac{k}{2})}x^{\frac{k}{2}-1}(1+x)^{-k}dx$, *where* $c_0 = \frac{\epsilon_0^2\sigma^2+\sigma_0^2}{\sigma^2+\sigma_0^2}$. *Then the recall rate of DWGC-CI outperforms the DWGC's in window length* k.

Notice that the left part in the lemma. 2 holds the property

$$\sup_{\{\Phi_t,...\Phi_{t+k-1}\}\in R^k} \sum_{q_1=t}^{t+k-1}\sum_{q_2=t}^{t+k-1} \frac{1}{(c_0\frac{\Phi_{q_2}^2}{\Phi_{q_1}^2}+1)\prod_{j=t,j\neq q_1}^{t+k-1}(1-\frac{\Phi_j^2}{\Phi_{q_1}^2})\prod_{j=t,j\neq q_2}^{t+k-1}(1-\frac{\Phi_j^2}{\Phi_{q_2}^2})} = +\infty, \tag{23}$$

thus such set $\{\Phi_t, ...\Phi_{t+k-1}\}$ in lemma. 2 naturally exists. On this basis, following the lemma. 2, the regularization term can be designed as

$$\left\| \sum_{q_1=t}^{t+k-1}\sum_{q_2=t}^{t+k-1} \frac{1}{(c_0\frac{\Phi_{q_2}^2}{\Phi_{q_1}^2}+1)\prod_{j=t,j\neq q_1}^{t+k-1}(1-\frac{\Phi_j^2}{\Phi_{q_1}^2})\prod_{j=t,j\neq q_2}^{t+k-1}(1-\frac{\Phi_j^2}{\Phi_{q_2}^2})} - \int_{c_0}^{+\infty}\frac{1}{\mathcal{B}(\frac{k}{2},\frac{k}{2})}x^{\frac{k}{2}-1}(1+x)^{-k}dx \right\|_2. \tag{24}$$

D.1 The proof of lemma. 2

Proof.

$$
\begin{aligned}
&P\left(\frac{\sum_q (\hat{Y}_{i,q} - Y_{i,q})^2}{\sum_q (\hat{Y}_{i|j,q} - Y_{i,q})^2} > 1\right) \\
&= \int_{c_0}^{+\infty} \frac{1}{\mathcal{B}(\frac{k}{2}, \frac{k}{2})} x^{\frac{k}{2}-1} (1+x)^{-k} dx \\
&\overset{*}{\leq} \sum_{q_1=t}^{t+k-1} \sum_{q_2=t}^{t+k-1} \frac{1}{(c_0 \frac{\Phi_{q_2}^2}{\Phi_{q_1}^2} + 1) \prod_{j=t, j\neq q_1}^{t+k-1} (1 - \frac{\Phi_j^2}{\Phi_{q_1}^2}) \prod_{j=t, j\neq q_2}^{t+k-1} (1 - \frac{\Phi_j^2}{\Phi_{q_2}^2})} \\
&= \sum_{q_1=t}^{t+k-1} \sum_{q_2=t}^{t+k-1} \frac{\frac{1}{\frac{1}{\Phi_{q_1}^2(\sigma^2+\sigma_0^2)} + \frac{1}{\Phi_{q_2}^2(\epsilon_0^2\sigma^2+\sigma_0^2)}}}{\prod_{j=t, j\neq q_1}^{t+k-1} (1 - \frac{\Phi_j^2}{\Phi_{q_1}^2}) \Phi_{q_2}^2 (\epsilon_0^2\sigma^2 + \sigma_0^2) \prod_{j=t, j\neq q_2}^{t+k-1}} \\
&= \int_0^{+\infty} \sum_{q_1=t}^{t+k-1} \sum_{q_2=t}^{t+k-1} \frac{e^{-\frac{x}{\Phi_{q_1}^2(\sigma^2+\sigma_0^2)} - -\frac{x}{\Phi_{q_2}^2(\epsilon_0^2\sigma^2+\sigma_0^2)}}}{\prod_{j=t, j\neq q_1}^{t+k-1} (1 - \frac{\Phi_j^2}{\Phi_{q_1}^2}) \Phi_{q_2}^2 (\epsilon_0^2\sigma^2 + \sigma_0^2) \prod_{j=t, j\neq q_2}^{t+k-1}} dx \\
&= \int_0^{+\infty} \int_x^{+\infty} \sum_{q=t}^{t+k-1} \frac{e^{-\frac{m}{\Phi_q^2(\sigma^2+\sigma_0^2)}}}{\Phi_q^2(\sigma^2+\sigma_0^2) \prod_{j=t, j\neq q}^{t+k-1} \left(1 - \frac{\Phi_j^2}{\Phi_q^2}\right)} dm \\
&\quad \sum_{q=t}^{t+k-1} \frac{e^{-\frac{x}{\Phi_q^2(\epsilon_0^2\sigma^2+\sigma_0^2)}}}{\Phi_q^2(\epsilon_0^2\sigma^2+\sigma_0^2) \prod_{j=t, j\neq q}^{t+k-1} \left(1 - \frac{\Phi_j^2}{\Phi_q^2}\right)} dx \\
&= \int_0^{+\infty} \int_x^{+\infty} f_2(m) f_1(x) dm dx \\
&= P\left(\frac{\sum_q (\hat{Y}_{i,q}^{\Phi} - Y_{i,q}^{\Phi})^2}{\sum_q (\hat{Y}_{i|j,q}^{\Phi} - Y_{i,q}^{\Phi})^2} > 1\right)
\end{aligned}
\tag{25}
$$

Here $*$ is from the inequality in lemma. 2. Hence we proved $P(\frac{\sum_q (\hat{Y}_{i,q}^{\Phi} - Y_{i,q}^{\Phi})^2}{\sum_q (\hat{Y}_{i|j,q}^{\Phi} - Y_{i,q}^{\Phi})^2} > 1) > P(\frac{\sum_q (\hat{Y}_{i,q} - Y_{i,q})^2}{\sum_q (\hat{Y}_{i|j,q} - Y_{i,q})^2} > 1)$. □

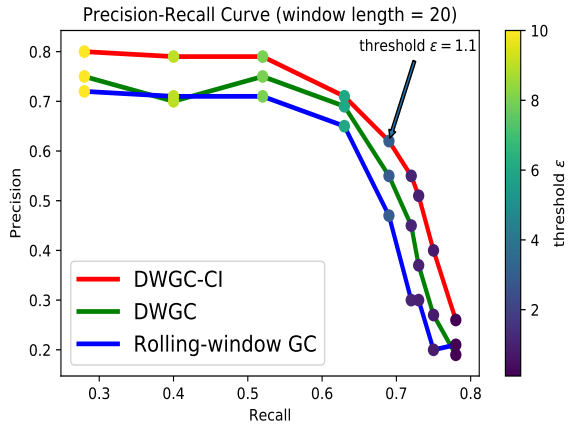
E Sensitivity analysis

E.1 Sliding window size selection

The impact of window size selection is shown in Tab. 5.

Table 5: To illustrate the impact of the window size selection, we add an experimental comparison using the American data of 2008 in the stock-and-market experiment. We change the sliding window size to a month and the results are as follows. Compared with the results in the paper, the detected causality does not completely fit with the prior knowledge and possibly misses some part of the dynamic causality directions.

X	F-test / method	Jan-Feb	Mar-Apr	May-Jun	Jul-Aug	Sept-Oct	Nov-Dec
NLC	DWGC-CI	21.67/12.30	38.6/22.51	28.09/2.60	7.20/34.54	12.62/39.36	34.76/53.65
	DWGC	32.64/43.65	5.32/5.32	6.43/54.32	5.23/46.23	53.23/54.23	54.32/45.32
TAL	DWGC-CI	3.21/32.12	32.12/54.23	43.23/0.54	32.11/14.22	1.23/2.34	32.43/0.53
	DWGC	43.12/43.23	34.21/32.12	54.23/0.43	0.32/54.23	54.23/56.65	54.23/0.43
TMV	DWGC-CI	43.23/54.23	43.23/0.65	0.76/0.89	0.43/0.67	91.54/54.23	0.76/0.32
	DWGC	54.34/23.32	54.32/34.23	43.23/43.54	12.23/45.23	34.12/35.32	4.12/4.23
TFR	DWGC-CI	47.32/56.09	4.23/45.21	54.23/56.64	54.23/65.23	43.23/1.24	43.23/23.12
	DWGC	43.23/49.34	2.24/43.23	34.23/0.34	1.24/1.65	9.34/4.34	4.32/65.23



E.2 Precision-recall rate (threshold selection)

This is an auxiliary experiment for our AR/NAR simulation. In this section, we set window length = 20 in the NAR case, for instance, to test the correlation between the causality threshold (ϵ) and the performance (precision and recall). The experimental results show that our method DWGC-CI is robust to the selection of epsilon values and consistently outperforms DWGC and Rolling-window GC. Moreover, the difference between the performances of various methods is most significant at $\epsilon = 1.1$.

Very few false-positive rates without GC ground truth If the data has no GC relation, our DWGC-CI captures very few false positives. We add the following result to the synthetic data experiment: two series with no correlation (and certainly no GC) are constructed as $T_{i,t} = rand(0,1), i = 0, 2, t = 0\ 1000$, where is a random variable ranging from to . We find that the probability of being judged as GC by DWGC-CI is relatively low at all window sizes, as follows. When there is no correlation between the two series, the false-positive probability is relatively low.

Sliding window size	10	20	30	100
False-positive rate of DWGC	0.142	0.190	0.072	0.032
False-positive rate of DWGC-CI	**0.041**	**0.024**	**0.039**	**0.00**

Table 6: Very few false-positive rates without GC ground truth of DWGC-CI.

F Concept drift

Besides, we can also treat this problem given "concept drift" Tsymbal (2004), which means that "changes in the hidden context can induce more or less radical changes in the target concept" Tsymbal (2004). "Concept drift" falls into two categories: sudden (instantaneous) and gradual. The traditional GC only extracts the causal information from the latter and ignores the former, while our DWGC-CI method can extract both.

G Back door criteria

The premise of Eqn. equation 2 is that series channels i and j meet the "backdoor condition" Pearl & Robins (1995). That is, we should first exclude common confounding from other channels that both effect series i and j. Moreover, in a more strict sense, Granger causality is considered as a non-causal "precedence" Pearl (2018), according to the informal fallacy "Post hoc ergo propter hoc" Woods & Walton (1977) which means that "after this, therefore because of this". It is a flaw inherent in GC's approach itself. Thus the objective of our paper is to generalise the existing GC rather than sticking to such philosophical dilemmas.