Identifiability of Non-Gaussian Structural VAR Models for Subsampled and Mixed Frequency Time Series

Alex Tank
Department of Statistics
University of Washington
alextank@uw.edu

Emily Fox
Department of Statistics
University of Washington
ebfox@uw.edu

Ali Shojaie
Department of Biostatistics
University of Washington
ashojaie@uw.edu

ABSTRACT

Causal inference in multivariate time series is confounded by subsampling in time between the true causal scale and the observed data sampling rate. In practice, this presents challenges for inferring causal interaction between time series due to differences in sampling rates across time series and generally low sampling rates due to technological limitations. To determine instantaneous and lagged effects between time series at the true causal scale, we take a model based approach based on structural vector autoregressive (SVAR) models. We show that when the underlying noise, or shocks, to the system are non-Gaussian, both the parameters of the true model and its causal structure are identifiable from subsampled and mixed frequency data. Our work builds on the recent work of Gong et al. [1], who established the identifiability of VAR models from subsampled time series with non-Gaussian noise, but with no instantaneous interactions. Here, we generalize their work to the SVAR case to handle instantaneous interactions and, additionally, the multifrequency setting. The resulting approach provides a complete picture of identifiability in non-Gaussian SVAR models under arbitrary mixed frequency subsampling.

Keywords

mixed frequency; subsampled time series; identifiability; structural VAR; non-Gaussian

1. INTRODUCTION

Classical approaches to multivariate time series and Granger causality assume that all time series are sampled at the same sampling rate. However, due to data integration across heterogenous sources, many data sets in econometrics, health care, and neuroscience are composed of multiple time series sampled at different rates, referred to as *mixed frequency* time series. Furthermore, due to the cost of data collection, many time series may be sampled at a rate lower than the true causal scale of the underlying physical process. For example, many econometric indicators, like GDP and housing price data, are recorded at quarterly and monthly scales [2]. However, there may be important interactions between these indicators at the weekly or bi-weekly scales [2, 3, 4]. In neuroscience, imaging technologies with high spatial resolution like functional magnetic resonance imaging or fluorescent calcium imaging have relatively low temporal resolutions. On the other hand, it is well-known that many important neuronal processes and interactions happen at finer time scales [5]. Furthermore, a causal analysis rooted at a slower time scale than the true causal time scale may both miss true interactions and add spurious ones [5, 6, 4, 7]. A comprehensive approach to Granger causality in multivariate time series should be able to simultaneously accommodate both mixed frequency and subsampled data.

A recent set of papers has examined the problem of causal discovery in subsampled time series drawing from methods in causal structure learning using graphical models [8, 9, 10, 11]. These methods are model free, and automatically infer a sampling rate for causal relations most consistent with the data.

In contrast, we take a model based approach and examine the identifiability of structural vector autoregressive models (SVAR) under both subsampling and mixed frequencies. SVARs are an important tool in time series analysis [12, 13] and are a mainstay in econometrics and macro-economic policy analysis. SVAR models combine classical linear autoregressive models with structural equation modeling [14] to allow analysis of both instantaneous and lagged causal effects between time series. However, SVAR models are commonly applied to regularly sampled data, where each series is observed at the same, discrete regular intervals. Typically, the time scale of a causal SVAR analysis is restricted to this shared sampling scale.

Our approach builds on the recent work of Gong et al. [1] who explore identifiability and estimation for VAR models under subsampling with independent innovations, i.e., *no* instantaneous causal effects or error correlations. They show that with non-Gaussian errors, the transition matrix is identifiable under subsampling, implying Granger causality estimation under subsampling is possible. For estimation, they model the non-Gaussian errors using a mixture of normal distributions and develop an EM algorithm for inference.

Interestingly, non-Gaussian errors have also been shown to aide model identifiability in structural VAR models with standard sampling assumptions [15, 16, 17, 18, 19]. This line of work applies techniques originally developed for both structural equation modeling with non-Gaussian errors and independent component analysis (ICA) [20] to the structural

WOODSTOCK '97 El Paso, Texas USA

DOI: 10.475/123_4

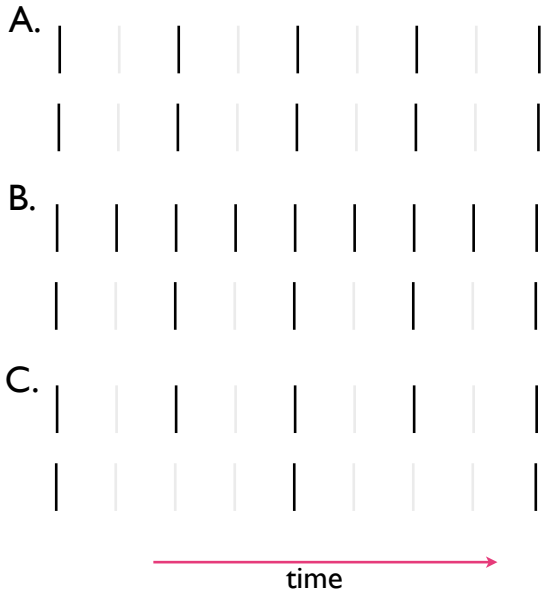

Figure 1: Schematic representing three different types of structured sampling. Black lines indicate observed data and light grey indicate missing data. A) Both series are subsampled at a rate of two. B) Mixed frequency example series one has no subsampling while series two is subsampled. C) Subsampled mixed frequency example where each series is subsampled, but at different rates.

VAR context. Importantly, non-Gaussian errors allow identification of the structural VAR model without any other identifying restrictions [16] and further allow identification of the causal ordering of the instantaneous effects if these are known to follow a directed acyclic graph (DAG) [17].

Our work concretely unifies the non-Gaussian subsampled VAR with independent innovations of Gong et al. [1] with the now extensive non-Gaussian structural VAR framework [15, 16, 17, 18, 19] by proving identifiability of a full structural VAR model of order one under arbitrary subsampling. As a result, we find that not only can one identify the causal structure of lagged effects from subsampled data, but also the DAG of the instantaneous effects without prior knowledge of the causal ordering. We generalize our results to mixed frequency time series with potential subsampling, where the subsampling level may be different for each time series.

While this work focuses on issues of model identifiability, we describe model estimation and its difficulties in Section 5, and leave it as future work.

2. BACKGROUND

2.1 SVAR models

Let $x_t \in \mathbb{R}^p$ be a p-dimensional multivariate time series for $t = 1, \ldots, T$ generated at a fixed sampling rate. We collect the entire set of x_ts into the matrix $\mathbf{X} = (x_1, \ldots, x_T)$. We assume the dynamics of x_t follow a combination of instantaneous effects, lagged autoregressive effects, and independent noise

$$x_t = \mathbf{B}x_t + \mathbf{D}x_{t-1} + e_t, \tag{1}$$

where $\mathbf{B} \in \mathbb{R}^{p\times p}$ is the structural matrix that determines the instantaneous time linear effects, $\mathbf{D} \in \mathbb{R}^{p\times p}$ is an autoregressive matrix that specifies the lag one effects conditional on the instantaneous effects, and $e_t \in \mathbb{R}^p$ is a white noise process such that $E(e_t) = 0\ \forall t$, e_{tj} is distributed as $e_{tj} \sim p_{e_j}$, and e_{ti} is independent of $e_{t'j}\ \forall i, j, t, t'$ such that $i \neq j$ and $t \neq t'$. $E(e_t e_t^T) = \Lambda$ where Λ is diagonal. Solving Eq. (1) in terms of x_t gives the following equation for the evolution of x_t:

$$x_t = (I - \mathbf{B})^{-1}\mathbf{D}x_{t-1} + (I - \mathbf{B})^{-1}e_t \tag{2}$$
$$= \mathbf{A}x_{t-1} + \mathbf{C}e_t \tag{3}$$

Under the representation in Eq. (3), each $\mathbf{A}_{ji}$ element denotes the lag one linear effect of series i on series j and $\mathbf{C} \in \mathbb{R}^{p\times p}$ is the structural matrix. Element e_{tj} is refered to as the *shock* to time series j and element $\mathbf{C}_{ji}$ is the linear effect of the shock on series j to time series i.

Conditions on $\mathbf{C}$, or equivalently $\mathbf{B}$, for model identifiability and estimation have been heavily explored [13]. Perhaps the most typical condition is that $\mathbf{C}$ is a lower triangular matrix with ones on the diagonal, implying a known causal ordering to the instantaneous effects. In this case, one may interpret the instantaneous effects as a directed acyclic graph (DAG) [21]. A DAG is a directed graph, $G = (V, E)$, with vertices $V = \{1, \ldots, p\}$ and directed edge set E, with no directed cycles. A causal ordering for a DAG is an ordering of the vertices into a sequence, π, such that if j comes before i in π then E does not contain any edge $i \to j$; see, e.g., [22] for more details. In the context of SVARs, for $i \neq j$ there exists a directed edge $i \to j$ from x_i to x_j in E, if and only if $\mathbf{C}_{ji}$ is nonzero.

Classical estimation may typically proceed by fitting $\mathbf{A}$ and $\mathbf{C}$ simultaneously with the identifiability conditions, or equivalently fitting $\mathbf{A}$ and $\Sigma = \mathbf{C}\Lambda\mathbf{C}^T$ and letting $\mathbf{C}$ be the unique Cholesky factorization of the Σ estimate. The downside of the classical SVAR methods is that they assume a known causal ordering of the instantaneous effects that may in practice be unknown. A recent line of work [15, 16, 17] instead assumes that the errors, e_t, are *non-Gaussian*. By leveraging the non-Gaussianity of errors, the causal ordering in $\mathbf{C}$, or $\mathbf{B}$, may be inferred directly from the data using techniques common in independent component analysis (ICA) [17]. Alternatively, one may dispense with causal orderings and lower triangular restrictions all together and directly estimate $\mathbf{C}$ [16]. Our analysis continues this line of work, leveraging non-Gaussianity of SVAR with subsampling and/or mixed frequency sampling.

3. SUBSAMPLED SVAR

3.1 The subsampled process

Subsampling occurs when, due to low temporal resolution, we only observe x_t every k time steps. In this case, we only have access to the observations $\tilde{\mathbf{X}} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_T) \equiv \big(x_1, x_{1+k}, \ldots, x_{1+(T-1)k}\big)$. We may marginalize out the un-

observed x_t to obtain the evolution equations for $\tilde{x}_t$:

$$\begin{aligned}\tilde{x}_{t+1} &= x_{1+tk} = \mathbf{A}x_{1+tk-1} + \mathbf{C}e_{1+tk} \\ &= \mathbf{A}\left(\mathbf{A}x_{1+tk-2} + \mathbf{C}e_{t+tk-1}\right) + \mathbf{C}e_{1+tk} \\ &= \dots \\ &= \mathbf{A}^k\tilde{x}_{t-1} + \sum_{l=0}^{k-1}\mathbf{A}^l\mathbf{C}e_{1+tk-l} \quad (4) \\ &= \mathbf{A}^k\tilde{x}_{t-1} + \mathbf{L}\tilde{e}_t, \quad (5)\end{aligned}$$

where $\tilde{e}_t = \left(e_{1+tk}^T, \dots e_{tk+k}^T\right)^T$ and $\mathbf{L} = \left(\mathbf{C}, \dots, \mathbf{A}^{k-1}\mathbf{C}\right)$. Eq. (5) appears to take a similar form to a SVAR process in Eq. (1); however, now the vector of shocks, $\tilde{e}_t$, is of dimension kp with special structure on both the structural matrix $\mathbf{L}$ and the distributions of the elements in $\tilde{e}_t$. Unfortunately, this representation no longer has the interpretation of instantaneous causal effects as described in Section 2.1 since there are now multiple shocks per individual time series. We will refer to the full parametrization of the subsampled SVAR model in Eq. (5) as $(\mathbf{A}, \mathbf{C}, k, e)$. Identifiability of the SVAR model means that there is a unique pair of $\mathbf{A}$ and $\mathbf{C}$ for the SVAR model consistent with the joint disitrbution $\tilde{\mathbf{X}}$ at subsampling rate k.

3.2 Lagged and Instantaneous Causality Confounds of Subsampling

A classical SVAR analysis on the $\tilde{x}_t$ that does not account for subsampling would incorrectly estimate lagged Granger causal effects in $\mathbf{A}^k$ since $\mathbf{A}_{ij}$ being zero does not imply that $(\mathbf{A}^k)_{ij} = 0$, and vice versa [1]. Zeros in the estimated structural matrix may also be incorrect if subsampling is ignored. Furthermore, classical SVAR estimation methods that assume a known causal ordering to the instantaneous shocks simply estimate the covariance of the error process, $\Sigma = E(\mathbf{C}e_t e_t^T\mathbf{C}^T) = \mathbf{C}\Lambda\mathbf{C}^T$, and let the estimated structural matrix be the Cholesky decomposition of Σ. Under subsampling, the covariance of the error process is given by

$$E(\mathbf{L}\tilde{e}_t\tilde{e}_t^T\mathbf{L}^T) = \mathbf{L}\left(I_k \otimes \Lambda\right)\mathbf{L}^T, \quad (6)$$

where $\otimes$ is the Kronecker product and I_k is the identity matrix of size k. Note that the causal structure given by zeros in the Cholesky decomposition of Eq. (6) need not be the same as those implied by $\mathbf{C}$.

As an example, consider the following process [1]:

$$\mathbf{A} = \begin{pmatrix} .8 & .5 \\ 0 & -.5 \end{pmatrix}$$

and $\mathbf{C} = \Lambda = I_p$. Then,

$$\mathbf{A}^k = \begin{pmatrix} .64 & 0 \\ 0 & .64 \end{pmatrix}$$

$$\mathbf{L}\left(I_k \otimes \Lambda\right)\mathbf{L}^T = \begin{pmatrix} 1.89 & -.4 \\ -.4 & 1.64 \end{pmatrix},$$

implying *no* lagged causal effect between x_1 and x_2 but a relatively large instantaneous interaction; this is the opposite of the true generating model! A graphical depiction of this example is given in Figure 2.

3.3 Identifiability under subsampling

While both lagged Granger causality and instantaneous structural interactions are confounded by subsampling, we show here that if one accounts for subsampling, then under some conditions one may still estimate the $\mathbf{A}$ and $\mathbf{C}$ matrices of the underlying SVAR process directly from the subsampled data.

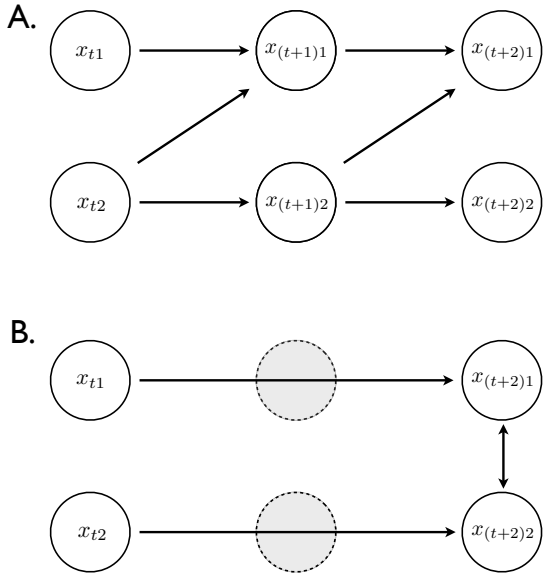

Figure 2: Graphical depiction of how subsampling confounds both causal analysis of lagged and instantaneous effects. A) The true causal diagram for the regularly sampled data. B) The estimated causal structure of the subsampled process when the effects of subsampling are ignored.

3.3.1 *Identifiability of* $\mathbf{L}$

As a first step to proving the identifiability result for $\mathbf{A}$ and $\mathbf{C}$ we note that that the matrix $\mathbf{L} = \left(\mathbf{C}, \dots, \mathbf{A}^{k-1}\mathbf{C}\right)$ in Eq. (5) is identifiable up to permutation and scaling of columns when the p_{e_j} (distribution of the e_{tj}) are all non-Gaussian:

PROPOSITION 1. *Suppose that all p_{e_j} are non-Gaussian. Given a known subsampling factor k and subsampled data $\tilde{X}$ generated according to Eq. (5), $\mathbf{L}$ may be determined up to permutation and scaling of columns.*

The proof closely follows the proof of Proposition 1 of [1] and follows from the following foundational result in the field of ICA [23].

LEMMA 2. *Let $\hat{e} = Jr$ and $\hat{e} = Ms$ be two representations of the n-dimensional random vector $\hat{e}$, where J and M are constant matrices of orders $n \times l$ and $n \times m$, respectively, and $r = (r_1, \dots, r_l)^T$ and $s = (s_1, \dots, s_m)^T$ are random vectors with independent components. Then the following assertions hold.*

1. *If the ith column of J is not proportional to any column of M, then r_i is Gaussian.*
2. *If the ith column of J is proportional to the jth column of M, then the logarithms of the characteristic functions of r_i and s_j differ by a polynomial in a neighborhood of the origin.*

Intuitively, this result states that if r is non-Gaussian, its elements are independent, and if Jr = Ms, it must be that M

and J are equal up to permutation and scaling of columns. This implies that one may estimate M from only observations of $\hat{e}$ and that the estimate of M should be equal up to permutations and scalings of the true data generating M.

To apply Lemma 2 to Proposition 1, note that $\mathbf{A}^k$ is identifiable by linear regression. Thus, the error component $\hat{e} = \tilde{x}_t - \mathbf{A}^k \tilde{x}_{t-1} = \mathbf{L}\tilde{e}_t$ satisfies the conditions of (i) of Lemma 2 and $\mathbf{L}$ is idendifiable up to permutations and scalings since $\tilde{e}_t$ are non-Gaussian.

3.3.2 Identifiability under $\mathbf{C} = I$

Using the identifiability result for $\mathbf{L}$ in Proposition 1 allows us to derive identifiability statements and conditions for $\mathbf{C}$ and $\mathbf{A}$ for the subsampled SVAR. We first require a few mild assumptions.

A1 x_t is stationary so that all singular values of $\mathbf{A}$ have modulus less than one.

A2 The distributions p_{e_i} are distinct for each i after rescaling by any non-zero scale factor, their characteristic functions are all analytic (or they are all non-vanishing), and none of them has an exponent factor with polynomial of degree at least 2.

Assumption A1 is standard in time series modeling [12] and A2 is also common in non-Gaussian, ICA-type models. Gong et al [1] provide identifiability results under A1 and A2 for the subsampled VAR with no instantaneous correlations, $\mathbf{C} = I$. We restate their result in our framework, both for comparison with the SVAR results and for use in Section 4, where we consider the mixed frequency sampling.

THEOREM 3. *(Gong et al. 2015) Suppose all of e_{ti} are non-Gaussian, and that the data $\tilde{x}_t$ are generated by Eq. (2.1) with $\mathbf{C} = I_p$ and that it also admits another kth order subsampling representation $(\mathbf{A}', I_p, e', k)$. If assumptions A1 and A2 hold the following statements are true.*

1. *$\mathbf{A}'$ can be represented as $\mathbf{A} = \mathbf{A}D_1$, where D_1 is a diagonal matrix with 1 or -1 on the diagonal. If we constrain the self influences, represented by the diagonal entries to be positive, then $\mathbf{A}' = \mathbf{A}$.*
2. *If each p_{e_i} is asymmetric, we have $\mathbf{A}' = \mathbf{A}$.*

3.3.3 Identifiability for general SVAR model

For identifiability of the full SVAR model under subsampling we require two additional assumptions:

A3 The variance of each p_{e_i} is equal to one, i.e., $\Lambda = I_p$.

A4 All p_{e_i} are asymmetric.

A5 $\mathbf{C}$ is full rank.

Assumption $A3$ is common in structural VAR models due to the inherent non-identifiability between scaling the e_is and scaling the columns of $\mathbf{C}$. Assumption A5 is mild, and contains the more restrictive assumptions in non-Gaussian SVAR models (necessary to infer the instantaneous structural DAG [24]), that $\mathbf{C}$ may be row and column permuted to a lower triangular matrix with non-zeros on the diagonal. Under these assumptions, we have the following identifiability result for general subsampled SVAR models:

THEOREM 4. *Suppose e_{ti} are all non-Gaussian and independent, and that the data are generated by a SVAR(1) process with representation $(\mathbf{A}, \mathbf{C}, e, k)$ that also admits another mixed frequency subsampling representation $(\mathbf{A}', \mathbf{C}', e', k)$. If assumptions A1-A3 hold, the following statements are true:*

1. *$\mathbf{C}$ is equal to $\mathbf{C}'$ up to permutation of columns and scaling of columns by 1 or -1, i.e. $\mathbf{C}' = \mathbf{C}P$ where P is a scaled permutation matrix with 1 or -1 elements. This implies $\Sigma = \mathbf{C}\mathbf{C}^T = \mathbf{C}'\mathbf{C}'^T = \Sigma^T$.*
2. *$\mathbf{A}$ is equal to $\mathbf{A}'$ if A4 and A5 also hold.*

The requirement that $\mathbf{C}$ be full rank is due to the structure of $\mathbf{L}$. Since one may identify $\mathbf{C}$ as the first p columns of $\mathbf{L}$, to obtain $\mathbf{A}$ we must premultiply the second set of p columns of $\mathbf{L}$ by $\mathbf{C}^{-1}$. The asymmetry assumption is needed since the scaling of the columns of $\mathbf{C}$ and $\mathbf{AC}$ by 1s or -1s is ambiguous and may be different if the distributions are symmetric; the asymmetry assumption ensures that the scalings are the same. See the Appendix for a full proof.

3.3.4 Identifying the structural DAG

If the instantaneous causal effects follow a directed acyclic graph (DAG), we may identify the DAG structure without any prior information about causal ordering of the variables in the DAG.

COROLLARY 5. *If $\mathbf{C}$ is lower triangular with positive diagonals,i.e. it follows a DAG G, then $\mathbf{C} = \mathbf{C}'$, implying the structure of G is identifiable without prior specification of the causal ordering of G under Assumptions A1-A3.*

(*proof sketch*) This result follows from the fact that $\mathbf{C}$ may be identified up to column permutations. Based on the identifiability results of [24], if $\mathbf{C}$ follows a DAG structure this implies it may be row and column permuted to a unique lower triangular matrix. The row permutations identify the causal ordering, and the nonzero elements below the diagonal identify the edges in G. See [24] for more details on identifiability and estimation of the DAG from $\mathbf{C}$.

Taken together, the results of Theorem 4 and Corollary 5 imply that when the shocks, e_t, are independent and non-Gaussian, a complete causal diagram of the lagged effects and the instantaneous effects are fully identifiable from the subsampled time series, $\tilde{\mathbf{X}}$.

4. MIXED FREQUENCY SVAR

Mixed frequency (MF) time series are common in econometrics, where the canonical example is in combined modeling of quarterly GDP with other monthly economic indicators. Estimation and forecasting of mixed frequency time series are commonly approached using both standard VAR and SVAR models [25, 26]. Typically, the VAR model is fit at the same scale as the fastest sampled time series. Due to costly data collection, this scale is generally arbitrary and may not reflect the true causal dynamics, leading to confounded Granger and instantaneous causality judgements [5, 7]. Our identifiability results from Section 3 may be generalized to the mixed frequency case with arbitrary levels of subsampling for each time series, implying accurate estimation of Granger and instantaneous causality is possible. We note that identifiability conditions for mixed frequency VAR

models with no subsampling at the fastest scale (Figure 1B) was an open problem for many years [27] . Anderson et al. [28] recently showed the mixed frequency VAR (MF-VAR) of type B in Figure 1 is *generically* identifiable from the first two observed moments of the MF-VAR, meaning that unidentifiable models make up at most a set of measure zero of the parameter space. However, no explicit identifiability conditions of the VAR process were given. In contrast, we use the results from Section 3 to provide explicit identifiability conditions for MF-SVAR models under arbitrary subsampling with non-Gaussian error assumptions.

4.1 Mixed Frequency SVAR

For simplicity of presentation, we assume each time series in $x_t \in \mathbb{R}^p$ is sampled at one of two sampling rates, slow subsampling rates k_s and fast subsampling rates k_f, so that $x_t = (x_t^s, x_t^f)$ where x_t^s are those series subsampled at k_s and x_t^f are those subsampled at k_f. Let $\mathbf{k} \in \{k_s, k_f\}^p$ be the list of subsampling rates for each time series. In Figure 1B, $k_f = 1$ and $k_s = 2$, whereas in Figure 1C, $k_f = 2$ and $k_s = 4$. Analogous to the subsampled case, we refer to a parameterization of a MF-SVAR model as $(\mathbf{A}, \mathbf{C}, e, \mathbf{k})$. Let k^* be the smallest multiple of both k_s and k_f; for example, in Figure 1C, $k^* = 4$.

We may derive a similar representation to Eq. (5) for MF series. Fix a time point t such that all series are observed. Let $I^{(q)}$ be a modified $p \times p$ identity matrix where all rows i such that x_{ti} is not observed at time $t - q$ are set to zero. Further, let $I^{(\bar{q})} = I - I^{(q)}$, $A^{(q)} = I^{(q)}A$, and $A^{(\bar{q})} = I^{(\bar{q})}A$. Then

$$\begin{aligned} x_t &= \mathbf{A}x_{t-1} + \mathbf{C}e_t \\ &= \mathbf{A}I^{(1)}x_{t-1} + \mathbf{A}I^{(\bar{1})}x_{t-1} + \mathbf{C}e_t \\ &= \mathbf{A}I^{(1)}x_{t-1} + \mathbf{A}(\mathbf{A}^{(\bar{1})}x_{t-2} + \mathbf{C}^{(\bar{1})}e_{t-1}) + \mathbf{C}e_t \\ &= \dots \\ &= \mathbf{F}\tilde{x}_{t-1} + \mathbf{L}\tilde{e}_t, \end{aligned} \tag{7}$$

where

$$\begin{aligned} \mathbf{F} &= (\mathbf{A}, \mathbf{A}\mathbf{A}^{(\bar{1})}, \dots, \mathbf{A}\mathbf{A}^{(\bar{1})} \dots \mathbf{A}^{(\overline{k^*-1})}) \\ \mathbf{L} &= (\mathbf{C}, \mathbf{A}\mathbf{C}, \mathbf{A}\mathbf{A}^{(\bar{1})}\mathbf{C}, \dots, \mathbf{A}\mathbf{A}^{(\bar{1})} \dots \mathbf{A}^{(\overline{k^*-1})}\mathbf{C}) \\ \tilde{x}_{t-1} &= (I^{(1)}x_{t-1}, \dots, I^{(k)}x_{t-k^*}) \\ \tilde{e}_t &= (e_t, I^{(1)}e_{t-1}, \dots, I^{(k^*-1)}e_{t-k^*-1}) \end{aligned}$$

Eq. (7) takes the same form as Eq. (5), namely some matrix $\mathbf{F}$ times the observed time series samples $\tilde{x}_{t-1}$ plus a matrix $\mathbf{L}$ times a vector of non-Gaussian errors $\tilde{e}_t$. This intuitively suggests that similar identifiability results will hold.

4.2 Identifiability of MF-SVAR

We provide generalizations of both Theorems 3 and 4 to the mixed frequency case.

THEOREM 6. *Suppose all of the e_{ti} are non-Gaussian, and that the data $\tilde{x}_t$ are generated by Eq. (2.1) with $\mathbf{C} = I_p$. Further suppose that the process also admits another mixed frequency subsampling representation $(\mathbf{A}', I_p, e', \mathbf{k})$. If assumptions A1 and A2 hold, the following statements are true.*

1. *$\mathbf{A}'$ can be represented as $\mathbf{A} = \mathbf{A}D_1$, where D_1 is a diagonal matrix with 1 or -1 on the diagonal.*
2. *If any multiple of k_i is 1 smaller than some multiple of k_j then $\mathbf{A}_{ij} = \mathbf{A}'_{ij}$. If $\mathbf{A}_{ij} \neq 0$ this implies that $(D_1)_{jj} = 1$, i.e. the jth columns of $\mathbf{A}$ and $\mathbf{A}'$ are equal: $\mathbf{A}_{:j} = \mathbf{A}'_{:j}$.*
3. *If each p_{e_i} is asymmetric, we have $\mathbf{A}' = \mathbf{A}$.*

(*proof*) Points 1. and 3. follow since we may further subsample all series in x_t to a subsampling rate of k^*. This gives a subsampled $\tilde{\mathbf{X}}$ with representation $(\mathbf{A}, I, e, k^*)$. Applying Theorem 3 gives the result. Furthermore, we note that if some multiple of k_i is one less than some multiple of k_j then there exists a set of ts for Eq. (7), where series i is observed at time $t-1$ and series j is observed at time t. By identifiability of linear regression, $A'_{ij} = A_{ij}$. This resolves the sign ambiguity of the columns in 1. so that $\mathbf{A}_{:j} = \mathbf{A}'_{:j}$

THEOREM 7. *Suppose all of the e_{ti} are non-Gaussian and independent, and that the data is generated by a structural VAR(1) process with representation $(\mathbf{A}, \mathbf{C}, e, \mathbf{k})$ that also admits another mixed frequency subsampling representation $(\mathbf{A}', \mathbf{C}', e', \mathbf{k})$. If assumptions A1-A3 hold, the following statements are true:*

1. *$\mathbf{C}$ is equal to $\mathbf{C}'$ up to permutation of columns and scaling of columns by 1 or -1, ie $\mathbf{C} = \mathbf{C}P$ where P is a scaled permutation matrix with 1 or -1 elements. This implies $\Sigma = \mathbf{C}\mathbf{C}^T = \mathbf{C}'\mathbf{C}'^T = \Sigma^T$.*
2. *If $\mathbf{C}$ is lower triangular with positive diagonals, i.e. the instantaneous interactions follow a DAG, and if for all i there exists a j such that any multiple of k_i is 1 smaller than some multiple of k_j with $A_{j:}C_{:i} \neq 0$, then $\mathbf{A} = \mathbf{A}'$.*
3. *$\mathbf{A} = \mathbf{A}'$ if A4 and A5 also hold.*

The proof of points 1. and 3. follows the same subsampling logic as the proof given for Theorem 6. The proof of 2 is given in the Appendix.

Taken together, Theorems 6 and 7 demonstrate that identifiability of SVAR models still holds for MF series with subsampling under non-Gaussian errors.

5. DISCUSSION

Our results provide sufficient conditions for identifiability of structural VAR models for both subsampled and mixed frequency time series. Importantly, the complete causal diagram of both lagged effects and instantaneous causal effects is fully identifiable under arbitrary subsampling schemes and non-Gaussian errors. Our identifiability results hold for non-Gaussian errors, but it would be interesting to specify what order moments of the process are required for identifiability. This line of work may aide in developing spectral estimation methods for this problem.

While we have focused on identifiability in this paper, ongoing work is focused around estimation of the lagged and instantaneous causal effects under subsampling. Gong et al. [1] model the non-Gaussian errors using a mixture of Gaussians and develop a full EM algorithm for inference in subsampled VAR models with identity error covariance. Unfortunately, due to the combinatorial nature of their missing data EM algorithm, each iteration of the algorithm scales as $m^{(k+1)p}$ where m is the number of mixture components for

each error distribution. We are currently pursuing approximate inference algorithms for this framework to scale to more than the two-time-series examples in [1]. Furthermore, Gong et al.'s algorithm only works for time series subsampled at the same rate, not mixed frequency data. Work to develop a more scalable inference method for arbitrary mixed frequency, subsampled data is currently ongoing.

6. APPENDIX

6.1 Proof of Theorem 4

We prove it for the subsampled case. The structural VAR model can be decomposed as:

$$\tilde{x}_t = \mathbf{A}^k \tilde{x}_{t-k} + \mathbf{L}\tilde{e}_t \tag{8}$$

$$= \mathbf{A}^k \tilde{x}_{t-1} + \vec{e}_t, \tag{9}$$

where $\mathbf{L} = (\mathbf{C}, \mathbf{AC}, \ldots, \mathbf{A}^{k-1}\mathbf{C})$ and $\vec{e}_t = \mathbf{L}\tilde{e}_t$. We may determine A^k uniquely by linear regression and thus determine the distribution of $\vec{e}_t$. Proposition 1 states that each column of L' is a scaled version of a column of L. Denote by $\mathbf{L}_{lp+i}$, $l = 0, \ldots, k-1$, $i = 1, \ldots, p$ the $(lp+i)$th column of $\mathbf{L}$, and similarly for L'_{lp+i}. From the Uniqueness Theorem in Erikson and Koivunen 2004 [23], we know that under condition A2, for each i, there exists one and only j such that the distribution of $e_{(t-l)i}, l = 1, \ldots, k-1$ is the same as the disitrbution of $e'_{(t-l)j}$, $l = 1, \ldots, k-1$ up to changes in location and scale. This implies that each column in $\mathbf{L}_{lp+i}$, $l = 0, \ldots, k-1$, is proportional to at least one of the nonzero columns in $\mathbf{L}_{lp+j}$, $l = 1, \ldots, k-1$, and vice versa. The proportionality must be either 1 or -1 since we have standardized the p_e to have unit variance. Furthermore, it must be the case that ${}_{lp+i}$ is proportional to column $\mathbf{L}'_{lp+j}$ for j and $i = 1, \ldots, p$ since the columns are ordered in magnitude in both $\mathbf{L}$ and $\mathbf{L}'$, ie $||\mathbf{L}_{lp+i}||_2 > ||\mathbf{L}_{(l+1)p+i}||_2$,

$$||\mathbf{L}_{(l+1)p+i}||_2 = ||\mathbf{A}\mathbf{A}^l\mathbf{C}_{:i}||_2 \tag{10}$$

$$< ||\mathbf{A}||_2||\mathbf{A}^l\mathbf{C}_{:i}||_2 \tag{11}$$

$$< ||\mathbf{A}^l\mathbf{C}_{:i}||_2 \tag{12}$$

$$= ||\mathbf{L}_{lp+i}||_2. \tag{13}$$

This implies that $\mathbf{L}'$ may be written as:

$$\mathbf{L}' = \mathbf{L}P \tag{14}$$

$$= \left(\mathbf{C}P_0, \mathbf{AC}P_1, \ldots \mathbf{A}^{k-1}\mathbf{C}P_{k-1}\right), \tag{15}$$

where P_i is a scaled permutation matrix with either 1s or -1 scaling factors where P_i and P_j have the same permutation pattern but potentially different scaling factors. This proves the first assertion, ie $\mathbf{C}' = \mathbf{C}P_0$ and $\Sigma' = \mathbf{C}'\mathbf{C}'T = \mathbf{C}P_0P_0^T\mathbf{C}^T = \mathbf{C}\mathbf{C}^T = \Sigma$. Now, if the p_e are restricted to be nonsymmetric then the scaling factors must all be 1 so that all the P_i are equal.

$$\mathbf{A}'\mathbf{C}' = \mathbf{A}'\mathbf{C}P \tag{16}$$

$$= \mathbf{AC}P \tag{17}$$

and since $\mathbf{C}$ is full rank, $\mathbf{C}P$ is full rank so that $\mathbf{A}' = \mathbf{A}$, as desired.

6.2 Theorem 7 part 2

If $\mathbf{C}$ is lower triangular then $\mathbf{C} = \mathbf{C}'$. Now, $\mathbf{AC} = \mathbf{A}'\mathbf{C}'P_1 = \mathbf{A}'\mathbf{C}D$ where D is diagonal with either 1 or -1 on the diagonal. This implies that $\mathbf{L}'_{p+1:2p} = \mathbf{AC}D$. We procceed by induction. Since the last column of $\mathbf{C}$, $\mathbf{C}_{:p}$, is zeros everywhere except the last element, we must have that $\mathbf{C}_{pp}\mathbf{A}_{:p}D_{pp} = \mathbf{L}'_{2p} = \mathbf{C}_{pp}\mathbf{A}'_{:p}$, so that $\mathbf{A}_{:p}D_{pp} = \mathbf{A}'_{:p}$. Following the same logic as the proof to item 2 of Theorem 6, if there exists some j such that a multiple of k_p is one less than a multiple of k_j and $\mathbf{A}_{pj} \neq 0$, then we can identify $\mathbf{A}_{pj}$, and hence its sign, implying $\mathbf{A}_{:p} = \mathbf{A}'_{:p}$.

Assume that $\mathbf{A}_{:i} = \mathbf{A}'_{:i}$ for $i > j$. Since $\mathbf{C}$ is lower diagonal we must have that

$$\mathbf{L}'_{p+j} = \left(\mathbf{C}_{jj}\mathbf{A}'_{:j} + \sum_{i>j}\mathbf{C}_{ij}\mathbf{A}_{:i}\right) \tag{18}$$

$$= D_{jj}\left(\mathbf{C}_{jj}\mathbf{A}_{:j} + \sum_{i>j}\mathbf{C}_{ij}\mathbf{A}_{:i}\right). \tag{19}$$

Since $\mathbf{A}_{lj} = \mathbf{A}'_{lj}$ with $\mathbf{C}_{jj}\mathbf{A}_{lj} + \sum_{i>j}\mathbf{C}_{ij}\mathbf{A}_{li} \neq 0$ for some l, this implies $D_{jj} = 1$, so that $\mathbf{A}_{:j} = \mathbf{A}'_{:j}$. Taken together, $\mathbf{A} = \mathbf{A}'$.

7. REFERENCES

[1] Mingming Gong, Kun Zhang, Bernhard Schölkopf, Dacheng Tao, and Philipp Geiger. Discovering temporal causal relations from subsampled data. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, pages 1898–1906, 2015.

[2] Filippo Moauro and Giovanni Savio. Temporal disaggregation using multivariate structural time series models. *The Econometrics Journal*, 8(2):214–234, 2005.

[3] Daniel O Stram and William WS Wei. A methodological note on the disaggregation of time series totals. *Journal of Time Series Analysis*, 7(4):293–302, 1986.

[4] John CG Boot, Walter Feibes, and Johannes Hubertus Cornelius Lisman. Further methods of derivation of quarterly figures from annual data. *Applied Statistics*, pages 65–75, 1967.

[5] Douglas Zhou, Yaoyu Zhang, Yanyang Xiao, and David Cai. Analysis of sampling artifacts on the granger causality analysis for topology extraction of neuronal dynamics. *Frontiers in computational neuroscience*, 8, 2014.

[6] Andrea Silvestrini and David Veredas. Temporal aggregation of univariate and multivariate time series models: a survey. *Journal of Economic Surveys*, 22(3):458–497, 2008.

[7] Jörg Breitung and Norman R Swanson. Temporal aggregation and spurious instantaneous causality in multiple time series models. *Journal of Time Series Analysis*, 23(6):651–665, 2002.

[8] David Danks and Sergey Plis. Learning causal structure from undersampled time series. 2013.

[9] Sergey Pils, David Danks, and Jianyu Yang. Mesochronal structure learning. 2015.

[10] Sergey Plis, David Danks, Cynthia Freeman, and Vince Calhoun. Rate-agnostic (causal) structure learning. In *Advances in Neural Information Processing Systems*, pages 3285–3293, 2015.

[11] Antti Hyttinen, Sergey Plis, Matti Järvisalo, Frederick

Eberhardt, and David Danks. Causal discovery from subsampled time series data by constraint optimization. *arXiv preprint arXiv:1602.07970*, 2016.

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

[13] Andrew C Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge university press, 1990.

[14] Jodie B Ullman and Peter M Bentler. *Structural equation modeling*. Wiley Online Library, 2003.

[15] Kun Zhang and Aapo Hyvärinen. Causality discovery with additive disturbances: An information-theoretical perspective. In *Machine learning and knowledge discovery in databases*, pages 570–585. Springer, 2009.

[16] Markku Lanne, Mika Meitz, Pentti Saikkonen, et al. Identification and estimation of non-gaussian structural vector autoregressions. *CREATES, Arhus Univerisity, Technical report*, 2015.

[17] Aapo Hyvärinen, Kun Zhang, Shohei Shimizu, and Patrik O Hoyer. Estimation of a structural vector autoregression model using non-gaussianity. *The Journal of Machine Learning Research*, 11:1709–1731, 2010.

[18] Aapo Hyvärinen, Shohei Shimizu, and Patrik O Hoyer. Causal modelling combining instantaneous and lagged effects: an identifiable model based on non-gaussianity. In *Proceedings of the 25th international conference on Machine learning*, pages 424–431. ACM, 2008.

[19] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. Causal inference on time series using restricted structural equation models. In *Advances in Neural Information Processing Systems*, pages 154–162, 2013.

[20] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent component analysis*, volume 46. John Wiley & Sons, 2004.

[21] Steffen L Lauritzen. *Graphical models*. Clarendon Press, 1996.

[22] Ali Shojaie and George Michailidis. Penalized likelihood methods for estimation of sparse high-dimensional directed acyclic graphs. *Biometrika*, 97(3):519–538, 2010.

[23] Jan Eriksson and Visa Koivunen. Identifiability, separability, and uniqueness of linear ica models. *Signal Processing Letters, IEEE*, 11(7):601–604, 2004.

[24] Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear non-gaussian acyclic model for causal discovery. *The Journal of Machine Learning Research*, 7:2003–2030, 2006.

[25] Claudia Foroni and Massimiliano Giuseppe Marcellino. A survey of econometric methods for mixed-frequency data. *Available at SSRN 2268912*, 2013.

[26] Frank Schorfheide and Dongho Song. Real-time forecasting with a mixed-frequency var. *Journal of Business & Economic Statistics*, 33(3):366–380, 2015.

[27] Baoline Chen and Peter A Zadrozny. An extended yule-walker method for estimating a vector autoregressive model with mixed-frequency data. *Advances in Econometrics*, 13:47–74, 1998.

[28] Brian DO Anderson, Manfred Deistler, Elisabeth Felsenstein, Bernd Funovits, Lukas Koelbl, and Mohsen Zamani. Multivariate ar systems and mixed frequency data: G-identifiability and estimation. *Econometric Theory*, pages 1–34, 2015.