The EAS approach for graphical selection consistency in vector autoregression models

Jonathan P Williams[(1)], Yuying Xie[(2)], Jan Hannig[(1)]

University of North Carolina at Chapel Hill[(1)]
Michigan State University[(2)]

Abstract

As evidenced by various recent and significant papers within the frequentist literature, along with numerous applications in macroeconomics, genomics, and neuroscience, there continues to be substantial interest to understand the theoretical estimation properties of high-dimensional vector autoregression (VAR) models. To date, however, while Bayesian VAR (BVAR) models have been developed and studied empirically (primarily in the econometrics literature) there exist very few theoretical investigations of the repeated sampling properties for BVAR models in the literature. In this direction, we construct methodology via the ε-*admissible* subsets (EAS) approach for posterior-like inference based on a generalized fiducial distribution of relative model probabilities over all sets of active/inactive components (graphs) of the VAR transition matrix. We provide a mathematical proof of *pairwise* and *strong* graphical selection consistency for the EAS approach for stable VAR(1) models which is robust to model misspecification, and demonstrate numerically that it is an effective strategy in high-dimensional settings.

Keywords: empirical Bayes; generalized fiducial inference; graph selection; high-dimensional model selection; large sample properties

Running title: EAS for graph selection

Corresponding Author: Jonathan P Williams, `jwilli27@ncsu.edu`

1 Introduction

Despite the lack of theoretical investigations of the repeated sampling properties for BVAR models, Bayesian methodology can surely offer important contributions to the high-dimensional VAR model literature, beyond what could be developed in a frequentist framework. One notable such contribution is the construction of posterior distributions over the set of all relative model probabilities. This framework of posterior inference has been widely exploited over the last decade in the high-dimensional linear regression literature, and we anticipate it will see comparable success for high-dimensional VAR models in the near future.

Our constructed EAS methodology allows for such posterior-like inference of relative model probabilities for all graphs, and additionally we provide an algorithm which is self-tuning (i.e., no cross-validation is needed for calibration to data sets). Such Bayesian model selection approaches are very useful for learning important relationships among the various components (univariate time-series) in the VAR model. The EAS methodology is an entirely new perspective on model selection which was originally developed to effectively account for linear dependencies among subsets of covariates in the high-dimensional linear regression setting in Williams & Hannig (2019).

To the best of our knowledge, our established *pairwise* and *strong* model selection consistency results are the first of their kind in the BVAR literature. This type of result is sure to be followed by similar results in the high-dimensional BVAR literature, analogous to the emergence of model selection strong consistency results in the high-dimensional Bayesian linear regression literature such as Johnson & Rossell (2012), Narisetty & He (2014), Williams & Hannig (2019).

Further, we demonstrate how to construct an alternative framework for posterior-like inference in the VAR(1) model setting which eliminates prior choice and specification. We avoid the necessity of prior distributions altogether by implementing a generalized fiducial inference (GFI) approach (see Hannig et al. 2016). And while our model selection consistency results derive from a Gaussian assumption on the VAR(1) model errors, they are actually the first ever results about a fiducial distribution under model misspecification. This is due to the fact that all of the supporting theorems and lemmas we contribute are non-asymptotic, and rely on a collection of explicit fourth moment bounds given in Section 3.3. Consequently, as long as the VAR(1) model errors are independent within and across time and there exist bounded fourth moments, our generalized fiducial consistency results (which assume Gaussian data) still hold even if the true data is not Gaussian.

We validate our methods empirically in low and high-dimensional settings on both synthetic and real data, and provide Python code for implementing our algorithm. This code/workflow for reproducing all numerical results can be found at `https://jonathanpw.github.io/research`.

Fiducial inference has a long history, but in the last decade there has been a renewed interest in the topic with a large number of authors contributing fundamental insights (Edlefsen et al. 2009, Berger et al. 2009, Xie & Singh 2013, Taraldsen & Lindqvist 2013, Veronese & Melilli 2015, Martin & Liu 2015, Schweder & Hjort 2016, Fraser 2019). A gentle introduction to technical

aspects of GFI is provided in Section 2.

Recent theoretical work on VAR models is largely comprised of considerations of regularized estimation procedures, most notably Basu et al. (2015). The Bayesian literature has not yet caught up. There do exist numerous papers on BVAR methodology, especially in the econometric literature, but on predominantly empirical investigations, see for example Bańbura et al. (2010), Korobilis (2013), Giannone et al. (2015), Ahelegbey et al. (2016). The primary tool of the BVAR literature has been implementations of the Minnesota (shrinkage) prior and its variants (Litterman 1986).

It has been found that BVAR with shrinkage priors is effective for large VAR models of economic time-series, but little has been provided in the way of theoretical guarantees (a notable exception is Ghosh et al. 2018) or even uncertainty quantification of competing model choices (a notable exception is Korobilis 2013). To the best of our knowledge, Ghosh et al. (2018) is the first in the literature to establish posterior parameter estimation consistency in the "large p large n" BVAR setting with $p = o(n)$, where p is the dimension of the VAR model and n is the number of observed time instances. While their consistency results are about the posterior behavior of the transition matrix coefficients under various prior specifications, our consistency results are about the posterior-like behavior of all relative model probabilities (akin to Bayes factors) under the prior-free GFI framework.

We loosely adopt notation for multivariate time-series from Lütkepohl (2005). The time-series $X^{(1)}, \ldots, X^{(n)} \in \mathbb{R}^p$ is taken to denote data from a VAR(1) model with no serial correlation, and so is generated as

$$\mathcal{Y} = A\mathcal{X} + \Sigma^{\frac{1}{2}}\mathcal{U}, \tag{1}$$

where $\mathcal{Y} := \begin{pmatrix} X^{(1)} & \cdots & X^{(n)} \end{pmatrix}$ and $\mathcal{X} := \begin{pmatrix} X^{(0)} & \cdots & X^{(n-1)} \end{pmatrix}$ are $p \times n$ matrices, $\mathcal{U} := \begin{pmatrix} U^{(1)} & \cdots & U^{(n)} \end{pmatrix}$ is a $p \times n$ matrix with $U^{(t)} \overset{\text{iid}}{\sim} \mathrm{N}_p(0, I_p)$ for $t \in \{1, \ldots, n\}$, A is a $p \times p$ matrix of coefficients, and $\Sigma := \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_p^2\}$. Assume $X^{(0)}$ is the p-dimensional zero vector. Further, let $G \subseteq \{1, \ldots, p^2\}$ be a set of indices denoting a graph of active (i.e., nonzero) components of A, and take A_g to be the $p \times p$ matrix A with active components corresponding to the graph G (all other components are zero).

We extend the high-dimensional linear regression EAS methodology developed in Williams & Hannig (2019) to this VAR(1) setting. This extension, particularly in the context of GFI, arises a variety of additional difficulties. Namely, the mathematical details of the generalized fiducial distribution for a VAR model are highly nontrivial and have never been worked out before. In particular, in a linear regression model, uncorrelated predictors can be omitted from the model with no effect on the data-generating system. However, if components in a VAR model are omitted, then the data-generating system is altered. These details make our paper substantially different from Williams & Hannig (2019).

The idea behind the EAS procedure is to efficiently make inference on the set of 2^{p^2} graphs, G, by discriminating on graphs which contain redundant active components. Our notion of redundancy is defined rigorously by the 'h-function' given later in (4). However, the basic intu-

ition is to assign negligible posterior-like probability to all A_g that can be closely approximated, predictively, by a graph containing fewer active components. This can occur for a variety of reasons, namely, correlated time-series in the VAR system of equations, and too small signal-to-noise coefficient magnitudes. For example, suppose $G = \{1, 2, 3, 4\}$ with $A_g = \left(\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right)$. Then the coefficient matrix A_g is not ε-*admissible* if, for instance, for some well-calibrated precision, $\varepsilon > 0$,

$$\left\| \left(\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right) \mathcal{X} - \left(\begin{smallmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{smallmatrix}\right) \mathcal{X} \right\| < \varepsilon,$$

where $\| \cdot \|$ is some measure of distance. In this case, predictions from the graph $\{1, 2, 4\}$ approximate that of A_g within ε precision, and so A_g is said to contain redundant information.

Note that in finite sample, and particularly high-dimensional, settings with highly-correlated data the EAS framework has the intuition that the oracle graph itself may not be ε-*admissible*. In these settings, the EAS methodology re-defines the notion of the 'true' graph to be some non-redundant subgraph of the oracle graph, at least non-asymptotically. This idea is important because it suggests that to develop inherently scalable methodology the key may be to re-define the notion of what one should hope to recover from a 'true' data-generating model in high-dimensional settings. Additional intuition for the EAS methodology is provided in Williams & Hannig (2019) in the context of linear regression.

The remainder of the paper is organized as follows. Section 2 defines the notion of ε-*admissibility* as well as constructs the generalized fiducial distribution for the EAS approach, and describes the Markov chain Monte Carlo (MCMC)-based computations. The main theoretical results are presented in Section 3, and numerical results are provided in Sections 4 and 5. The majority of the proofs are moved to the supplementary materials.

2 Methodology

To adapt ideas more smoothly from the linear regression setting of Williams & Hannig (2019), re-express the VAR(1) model in (1) in the form

$$Y = \mathcal{Z}_{G_o} \boldsymbol{\alpha}^0_{G_o} + (\mathcal{W}^0)^{\frac{1}{2}} \mathrm{vec}(\mathcal{U}), \tag{2}$$

where $Y := \mathrm{vec}(\mathcal{Y})$, $\mathcal{Z} := \mathcal{X}' \otimes I_p$, $\mathcal{W}^0 := I_n \otimes \Sigma^0$, $\boldsymbol{\alpha} := \mathrm{vec}(A)$, and G_o (as well as g_o seen later) denotes the oracle graph. Here and throughout, the superscript-zero notation denotes the true fixed values of the corresponding quantities. The subscript notation, $\mathcal{Z}_{G_o}$ (or $\boldsymbol{\alpha}_{G_o}$), refers to the sub-matrix (or sub-vector) with columns (or components) corresponding to the active components given by the index set G_o. The $\mathrm{vec}(\cdot)$ operator transforms an $n \times p$ matrix into an $np \times 1$ vector by stacking columns in descending order, from left to right. For example, $\mathrm{vec}(\mathcal{Y}) = (\, X^{(1)\prime} \, \ldots \, X^{(n)\prime} \,)'$. This linear model representation is also more convenient for expressing the

likelihood function,

$$f\big(Y|\boldsymbol{\alpha}_{G_o},\{\sigma_j\}\big)=\frac{1}{(2\pi)^{\frac{np}{2}}\big(\sigma_1^2\cdots\sigma_p^2\big)^{\frac{n}{2}}}e^{-\frac{1}{2}(Y-\mathcal{Z}_{G_o}\boldsymbol{\alpha}_{G_o})'\mathcal{W}^{-1}(Y-\mathcal{Z}_{G_o}\boldsymbol{\alpha}_{G_o})}, \tag{3}$$

which will be needed later on. For conciseness, the notation $\{\sigma_j\}$ is used as shorthand for $\{\sigma_1,\ldots,\sigma_p\}$.

Additional notation used for the remainder of the paper includes the following. For a scalar-valued argument $|\cdot|$ represents the absolute value, but for a set-valued argument it represents the cardinality. The norms $\|\cdot\|$ and $\|\cdot\|_0$ denote the vector L_2 and L_0 norms, respectively, while for a matrix A, $\|A\|_2:=\sqrt{\lambda_{\max}(A'A)}$ and $\|A\|_F:=\sqrt{\mathrm{tr}(A'A)}$ represent the matrix spectral and Frobenius norms, respectively. Additionally, the quantities $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimum and maximum eigenvalues of a given matrix, A, respectively. The notation $P(\cdot)$ and $E(\cdot)$ refer, respectively, to the probability measure and expectation with respect to the joint generalized fiducial distribution of A_g and Σ. Conversely, the notation $P_x(\cdot)$ and $E_x(\cdot)$ refer, respectively, to the probability measure and expectation associated with the uncertainty from the VAR(1) process, rather than the probability measure for the generalized fiducial distribution of the unknown parameters.

The centerpiece of the EAS model selection approach is a definition of model redundancy, as made rigorous by our notion of *ε-admissibility* and the h-function, presented next. As described in Section 1, the main intuition is that $\boldsymbol{\alpha}_G$ is considered non-redundant, or *ε-admissible*, if and only if there does not exist a close fitting graph with strictly fewer active components. However, there are also two additional constraints embedded in the h-function for *ε-admissibility*.

Definition 2.1. *Assume $\varepsilon,d>0$ and $c\in(0,1)$. A given coefficient matrix A_g, equivalently $\boldsymbol{\alpha}_G$, for some graph G is said to be ε-admissible if and only if $h\big(\boldsymbol{\alpha}_G,\{\sigma_j\}\big)=1$, where*

$$h\big(\boldsymbol{\alpha}_G,\{\sigma_j\}\big):=1\left\{\frac{1}{2}\|\mathcal{Z}_G'\mathcal{W}^{-1}\mathcal{Z}_G(\boldsymbol{\alpha}_G-b_{\min})\|^2\geq\varepsilon,\min_{1\leq j\leq p}\{m_j^g\}\geq d,\|A_g\|_2\leq c\right\} \tag{4}$$

where $b_{\min}$ solves $\min_{b\in\mathbb{R}^{|G|}}\frac{1}{2}\|\mathcal{Z}_G'\mathcal{W}^{-1}\mathcal{Z}_G(\boldsymbol{\alpha}_G-b)\|^2$ subject to $\|b\|_0\leq|G|-1$,

$$\{m_1^g,\ldots,m_p^g\}=diag\big\{(\mathcal{Y}-\widehat{A}_g\mathcal{X})(\mathcal{Y}-\widehat{A}_g\mathcal{X})'\big\}, \tag{5}$$

and $\widehat{A}_g:=\mathcal{Y}\mathcal{Z}_G'(\mathcal{Z}_G\mathcal{Z}_G')^{-1}$ is the least squares estimator for graph G.

To begin to understand the behavior of the h-function, first note that

$$\|\mathcal{Z}_G'\mathcal{W}^{-1}\mathcal{Z}_G(\boldsymbol{\alpha}_G-b_{\min})\|^2=\|\mathcal{Z}_G'\mathcal{W}^{-1}(\mathcal{Z}_G\boldsymbol{\alpha}_G-\mathcal{Z}_Gb_{\min})\|^2,$$

is analogous to a noiseless version of the Dantzig selector (Candes & Tao 2007) where $\mathcal{Z}_G$ is the design matrix for the linear model representation (2). One reason to use $\mathcal{Z}_G'\mathcal{W}^{-1}\mathcal{Z}_G$

versus simply $\mathcal{Z}_G$ is that the former is scale-invariant to the $\{\sigma_j\}$ and invariant to orthogonal transformations of the data. Second, note that if $\mathcal{Z}_G$ contains linearly dependent columns, then for any coefficients $\boldsymbol{\alpha}_G$, the linear prediction $\mathcal{Z}_G\boldsymbol{\alpha}_G$ can be exactly recovered by $\mathcal{Z}_G b_{\min}$ (since $\|b_{\min}\|_0 \leq |G|-1$). This immediately implies that since $\mathcal{Z}_G$ is an $np \times |G|$ matrix, for all G with $|G| > np$, $h\big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big) = 0$ by definition. For high-dimensional settings where $p > n$, then by construction, considering only ε-*admissible* graphs reduces the model selection problem from 2^{p^2} candidate graphs to only 2^{np}. This fact makes the EAS methodology inherently scalable.

The quantities c, d, and ε will now be described in alphabetical order. The component, $\|A_g\|_2 \leq c$, in the h-function concentrates the distribution of A_g to only allow for stable VAR(1) models with $c \in (0,1)$ (see Negahban et al. (2011), Loh & Wainwright (2012), Han et al. (2015)). In practice, since $\|A^0\|_2$ is typically not known the constraint $\|A_g\|_2 \leq c$ is replaced by $\|A_g\|_2 < 1$. The second component in the h-function is the expression $\min_{1\leq j\leq p}\{m_j^g\} \geq d$, where m_j^g for $j \in \{1,\ldots,p\}$ is understood as the residual sum-of-squares (RSS) for the j^{th} component of the VAR system. The basic idea is that the data-dependent quantity $d = d(\mathcal{Y}, \mathcal{X}, G_o)$ should be calibrated to $\min_{1\leq j\leq p}\{m_j^{g_o}\}$ which corresponds to the oracle graph, and so any graphs which have a better fit than the oracle will be excluded from consideration via the h-function. Accordingly, this device is designed to eliminate graphs which over-fit the data, and is important for establishing our asymptotic consistency results. However, in practice d can be set to a small value and left alone; more will be said about this in Section 4 with the numerical results.

For $\mathcal{Z}_G$ which have full column rank, the degree to which the features associated with graph G are redundant depends on the correlations between the p components of the VAR model, the distribution of the coefficients $\boldsymbol{\alpha}_G$ (i.e., the transition matrix A_g), scale matrix components $\{\sigma_j\}$, and the specified level of precision, ε. Our proposed default choice of ε, formulated from theoretical investigations (based on the Gaussian contemporaneous errors assumption), is for some $\rho \in (0, \frac{1}{2})$,

$$\varepsilon = \Lambda_g \cdot \max\Big\{1, n^{1-\rho}p^2\Big(.5\log(\log(n))|G| - |G_o|\Big)\Big\}. \tag{6}$$

There are predominantly two components to ε; the quantity $\Lambda_g := \|\mathcal{W}^{-\frac{1}{2}}\mathcal{Z}_G\|_F^2$ is particularly calibrated to the observed data since it originates from a tight concentration inequality for the transition matrix A_g, and the term $n^{1-\rho}p^2\log(\log(n))|G|$ is necessary asymptotically for managing the accumulating data and rapidly growing number of candidate graphs as $n, p \to \infty$. The basic idea is that Λ_g will always contribute, and the remaining terms will contribute for sufficiently large n or for $|G|$ which exceeds the number of active components in the oracle model. However, as is demonstrated in Section 4, for observed data Λ_g is so well-calibrated that it suffices to set $\varepsilon = \Lambda_g$, and thus also eliminating the need for a tuning parameter. More details about Λ_g are given in Section 3.2, particularly its expectation in (12).

With the EAS methodology now developed a framework of statistical inference is required for implementing it. A suitable such framework is GFI because it will allow us to construct posterior-like inference over the 2^{p^2} candidate graphs without having to specify any prior dis-

tributions. The intuition for GFI is to begin with a data-generating equation such as (2) and invert the equation on the data to solve for the unknown parameters. The resulting quantity is defined as the generalized fiducial distribution of the unknown parameters. Precise details for the construction of this approach are provided in Hannig et al. (2016). The generalized fiducial probability density function for the parameters in the VAR(1) model (2) has the form

$$r(\boldsymbol{\alpha}_G, \{\sigma_j\} \mid Y) = \frac{f\big(Y \mid \boldsymbol{\alpha}_G, \{\sigma_j\}\big) \cdot J\Big(Y, \big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big)\Big) \cdot h\big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big)}{\int\int f\big(Y \mid \boldsymbol{\alpha}_G, \{\sigma_j\}\big) \cdot J\Big(Y, \big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big)\Big) \cdot h\big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big)\; d\boldsymbol{\alpha}_G\; d\{\sigma_j\}}, \tag{7}$$

where the multiplication by the h-function appears as an infusion of the EAS methodology into the GFI framework, and the Jacobian term,

$$J\Big(Y, \big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big)\Big) := D\bigg(\nabla_{(\boldsymbol{\alpha}_G, \{\sigma_j\})} V\big(u, (\boldsymbol{\alpha}_G, \{\sigma_j\})\big)\Big|_{u=V^{-1}\left(Y, (\boldsymbol{\alpha}_G, \{\sigma_j\})\right)}\bigg)$$

with $D(A) = (\det A'A)^{\frac{1}{2}}$ and V denoting the data-generating equation (2). The Jacobian term results from inverting the data-generating equation on the unknown parameters. Note that the $\{\sigma_j\}$ are also dependent on the the particular graph G, but this dependence is suppressed in the notation for conciseness.

The likelihood function in (7) is given by (3), the h-function is given by (4), and the derivation of the Jacobian term is presented in the supplementary material. From the generalized fiducial density of $\boldsymbol{\alpha}_G$ and $\{\sigma_j\}$, the generalized fiducial mass function for a graph G is proportional to the normalizing constant in (7). In Bayesian theory, this constant of proportionality is understood as the marginal density of the data. Evaluating the integral in the denominator of (7) gives,

$$r(G \mid Y) \propto \frac{E\Big(h\big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big)|\widetilde{\mathcal{D}}_g'\widetilde{\mathcal{D}}_g|^{\frac{1}{2}}\Big) \prod_{j=1}^{p} \big(\frac{m_j^g}{2}\big)^{-\frac{n-|r_j^g|}{2}} \Gamma\Big(\frac{n-|r_j^g|}{2}\Big)}{\big(\frac{n}{2\pi}\big)^{\frac{|G|}{2}} \prod_{j=1}^{p} \Big|\sum_{t=1}^{n} X_{r_j^g}^{(t-1)} X_{r_j^g}^{(t-1)'}\Big|^{\frac{1}{2}}}, \tag{8}$$

where r_j^g is the set of active row indices of A_g for column $j \in \{1, \ldots, p\}$, and $\widetilde{\mathcal{D}}_g$ is a data-dependent and parameter-free $(np) \times (|G| + p)$ matrix defined in the supplementary material as part of the Jacobian term. Note that the inner expectation is with respect to the $\mathrm{N}_{|G|}\big(\widehat{\boldsymbol{\alpha}}_g, (\mathcal{Z}_G'\mathcal{W}^{-1}\mathcal{Z}_G)^{-1}\big)$ distribution, conditional on $\{\sigma_j^2\}$, and for each σ_j^2, is taken with respect to the inv-gamma$\big(\frac{1}{2}(n-|r_j^g|), \frac{1}{2}m_j^g\big)$ distribution. To ensure that $r(G \mid Y)$ defines a proper probability mass function, the normalizing constant in (8) is scaled so that $\sum_{i=1}^{p^2} \sum_{G:|G|=i} r(G \mid Y) = 1$.

Lastly, the relative model probabilities (8) can be computed via psuedo-marginal MCMC algorithms. Traditional MCMC is not feasible because the expected value appearing in (8) is not available in closed form. We implement the grouped independence Metropolis-Hastings

(GIMH) algorithm described in (Andrieu & Roberts 2009), which replaces the expected value with the empirical mean of importance samples at each step of the MCMC algorithm. In the case of (8), efficient importance samples are easily drawn from the $\mathrm{N}_{|G|}\big(\widehat{\boldsymbol{\alpha}}_g, (\mathcal{Z}_G'\mathcal{W}^{-1}\mathcal{Z}_G)^{-1}\big)$ and inv-gamma$\big(\frac{1}{2}(n-|r_j^g|), \frac{1}{2}m_j^g\big)$ distributions for $\boldsymbol{\alpha}_G$ and σ_j, respectively. The GIMH algorithm we construct is a Markov chain on the set of graphs $G \subseteq \{1,\ldots,p^2\}$, and proposals are made by either adding, removing, or replacing a component index in the current iterate of G in the chain.

A point of caution about the GIMH algorithm is that the mixing conditions are usually particularly sensitive to the number of importance samples taken to estimate an expectation at each step of the algorithm. However, the algorithm mixed well enough to yield very encouraging numerical results for the high-dimensional linear regression setting in (Williams & Hannig 2019), and Sections 4 and 5, here, serve to demonstrate that the algorithm is not only computationally feasible but also favorable for graph selection in the VAR(1) model setting. Further discussion of the algorithm is provided in (Williams & Hannig 2019), and a detailed pseudo-code description of the algorithm is provided at `https://jonathanpw.github.io/research`.

3 Theoretical results

The problem of graphical selection is difficult because the number of candidate graphs to choose among grows super-exponentially in the dimension of the VAR(1) model, 2^{p^2}. Accordingly, the utility of the EAS procedure is its inherent ability to effectively manage a very large number of candidate graphs by assigning negligible posterior-like probability to redundant graphs. The meaning of this assertion is made precise in Theorem 3.12 which states that the generalized fiducial distribution obtained from the EAS methodology exhibits pairwise graph selection consistency as both n and p are taken to infinity, and as a corollary, strong selection consistency for fixed p. The necessary mathematical conditions are discussed next.

3.1 Conditions

The first two conditions presented are related to the identifiability of the true data-generating graph, G_o. The expected ramifications of these conditions are verified empirically on data in Sections 4 and 5. Since this manuscript focuses on developing (i) novel methodology for graphical selection in the context of VAR models, (ii) the theory for GFI in the context of VAR models, and (iii) asymptotic theory related to BVAR models, we consider only a stable VAR(1) model for simplicity. We understand that, out of context, restriction to the stable VAR(1) setting is limiting. However, with respect to the three theoretical milestones we address, it would be overly ambitious to consider a more general setting in this first investigation. We adopt our notion of stability from Negahban et al. (2011), Loh & Wainwright (2012) and Han et al. (2015) that the true transition matrix satisfies $\|A^0\|_2 \leq c$ for some $c \in (0,1)$. It is assumed throughout that a valid c has been fixed a-priori. However, we expect the EAS procedure to

perform meaningful graph selection even if the stability condition is violated. The whole idea about the EAS approach is to find a meaningful model that is *non-redundant* as defined by our h function. So if h includes a stability constraint, then we are finding the "best fitting" distribution over models that are *non-redundant*. We are intentionally dismissing the notion that we hope to find the "true model" in finite data sets.

Condition 3.1 arises in the proof of Lemma S2.4 which is a necessary result for Theorem 3.10. It guarantees that the Jacobian term for the oracle graph in (8) will be lower bounded away from zero in probability. The quantity δ represents an approximation to $\lambda_{\min}\big(\Omega - E_x(\Omega)\big)$ (via Lemma 3.15) which manages the uncertainty resulting from the minimum eigenvalue of the Jacobian matrix $\widetilde{\mathcal{D}}'_{g_o}\widetilde{\mathcal{D}}_{g_o}$, where $\Omega := \frac{1}{n}\left(\begin{smallmatrix} \mathcal{X}\mathcal{X}' & \mathcal{X}\mathcal{U}' \\ \mathcal{U}\mathcal{X}' & \mathcal{U}\mathcal{U}' \end{smallmatrix}\right)$ and $E_x(\Omega) = \left(\begin{smallmatrix} \Gamma_n(0) & \\ & I_p \end{smallmatrix}\right)$. It is also assumed that a valid $\delta > 0$ has been fixed a-priori.

Condition 3.1. *The true transition matrix satisfies* $\|A^0\|_2 \leq c < 1$, $\lambda_{\max}\big(\Gamma_n(0)\big)$ *is bounded from above by a fixed constant, and*

$$\sqrt{n}\left[\lambda_{\min}\begin{pmatrix}\Gamma_n(0) & \\ & I_p\end{pmatrix} - \delta\right] > 4(1+c^2),$$

where $\delta > 0$, *and*

$$\Gamma_n(0) := \frac{1}{n}E_x(\mathcal{X}\mathcal{X}') = \frac{1}{n}\sum_{t=1}^{n}\sum_{k=0}^{t-2}(A^0)^k\Sigma^0(A^0)^{k'}.$$

Observe that this condition also implies that $\lambda_{\min}\big(\Gamma_n(0)\big) > \delta$.

Note that Lemma 3.15 guarantees $\lambda_{\min}\big(\Omega - E_x(\Omega)\big) \xrightarrow{P_x} 0$ as $n \to \infty$, assuming the p versus n relationship given by Condition 3.4. Thus, the condition can reasonably be verified on real data by assuming $\delta > 0$ is arbitrarily small and comparing the value of $\sqrt{n}\lambda_{\min}\left(\begin{smallmatrix} \frac{1}{n}\mathcal{X}\mathcal{X}' & \\ & I_p \end{smallmatrix}\right)$ to $4(1+c^2)$, where $\frac{1}{n}\mathcal{X}\mathcal{X}'$ is the obvious sample analogue to the population quantity considered in Condition 3.1. Since c is unknown in practice, for the purposes of checking this condition on real data evaluate $4(1+c^2) = 8$ for the worst case with c replaced by 1. We demonstrate on synthetic data in Section 4 that this verifiable condition is indeed meaningful for practical applications.

Condition 3.2, which originates from the proof of Theorem 3.10, is also well calibrated to real data. This condition states the maximum rate at which ε can be allowed to grow as a function of n, p, and Λ_{g_o}, whilst the oracle model G_o remains identifiable (i.e., no faster than $n^{1-\rho}p^2\Lambda_{g_o}$). The fixed quantity $\rho \in (0, \frac{1}{2})$ represents the 'gap' between how fast ε must grow (stated in Condition 3.4) to effectively manage the set of all 2^{p^2} candidate graphs under consideration, and how slow it must grow to not eliminated the oracle graph from consideration. Namely, $\varepsilon \propto n^{1-\rho}p^2\Lambda_g$ simultaneously satisfies Conditions 3.2 and 3.4 for any $\rho \in (0, \frac{1}{2})$. It is assumed throughout that a valid ρ has been fixed a-priori. The quantities on the left side of the inequality in Condition 3.2 are expected values of the corresponding quantities on the left side of the first constraint in the h-function (4).

Condition 3.2. *The oracle graph, G_o, satisfies* $\min_{1\leq j\leq p}\{m_j^{g_o}\}\geq d$,

$$\frac{1}{18}\left\|(\Gamma_n(0)\otimes(\Sigma^0)^{-1})_{G_o,G_o}(\boldsymbol{\alpha}^0_{G_o}-\widetilde{b})\right\|^2\geq\frac{\varepsilon}{n^{1-\rho}p^2\Lambda_{g_o}},$$

where $\rho\in(0,\frac{1}{2})$, $\widetilde{b}$ *solves* $\min_{b\in\mathbb{R}^{|G_o|}}\left\|(\Gamma_n(0)\otimes(\Sigma^0)^{-1})_{G_o,G_o}(\boldsymbol{\alpha}^0_{G_o}-b)\right\|^2$ *subject to* $\|b\|_0\leq|G_o|-1$, *and* $\varepsilon=\Lambda_{g_o}\cdot\widetilde{\varepsilon}$ *for some* $\widetilde{\varepsilon}$ *not depending on* Σ *or* A_{g_o}.

Unless the oracle model is known, Condition 3.2 is not verifiable on real data, but in Section 4 we are able to demonstrate the varying performance of the EAS procedure on simulated data when this condition is and is not satisfied. Note, that the coefficient of $\frac{1}{18}$ is a constant more pertinent to asymptotic considerations (and our proof technique), and should be understood as closer to the value of $\frac{1}{2}$ (which appears in the h-function).

The next condition is a component in the proof of Theorem 3.9 for guaranteeing that the h-function will drive the EAS procedure to assign negligible posterior-like probability to non-ε-*admissible* graphs, G, via the mass function $r(G\mid Y)$ in (8).

Condition 3.3. *For any G with $G\not\subseteq G_o$,*

$$\frac{9}{2}\left\|\left(E_x(\mathcal{Z}_G'\mathcal{Z}_G)\right)^{-1}E_x(\mathcal{Z}_G'Y)-\widetilde{b}\right\|^2<\frac{\varepsilon}{n^{1+\frac{\rho}{2}}p^3\Lambda_g},$$

where $\widetilde{b}$ *solves* $\min_{b\in\mathbb{R}^{|G|}}\left\|\left(E_x(\mathcal{Z}_G'\mathcal{Z}_G)\right)^{-1}E_x(\mathcal{Z}_G'Y)-b\right\|^2$ *subject to* $\|b\|_0\leq|G|-1$, *and* $\varepsilon=\Lambda_g\cdot\widetilde{\varepsilon}$ *for some* $\widetilde{\varepsilon}$ *not depending on* Σ *or* A_g.

The intuition for Condition 3.3 is that for graphs containing redundant active components the central tendency of the least squares estimator $\widehat{\boldsymbol{\alpha}}_g$ can be closely approximated by a vector of fewer active components. Notice that $\left(E_x(\mathcal{Z}_G'\mathcal{Z}_G)\right)^{-1}E_x(\mathcal{Z}_G'Y)$ is an approximation to $E_x(\widehat{\boldsymbol{\alpha}}_g)$. Since the least squares estimator is asymptotically well behaved for Gaussian VAR models, this condition is not particularly interesting and is easily satisfied in numerical experiments. Furthermore, it will hold trivially, for instance, if the columns $\mathcal{Z}_G$ are linearly dependent.

The final condition in this section is Condition 3.4, which simply states the asymptotic rate at which ε and d from the definition of h in (4) must increase as $n,p\to\infty$ for our main result, Theorem 3.12, to be established. In fact, the previous three conditions were all for establishing non-asymptotic bounds of concentration.

Condition 3.4. *For some fixed* $\rho\in(0,\frac{1}{2})$, $p^{\max\left\{\frac{14}{\rho},\frac{2}{1-2\rho}\right\}}=o(n)$. *For the positive constant K_1 specified in (14), as $n\to\infty$ or $n,p\to\infty$, ε satisfies*

$$\frac{\varepsilon}{9\Lambda_g}-K_1\left(\frac{p\|Y\|^2}{\sqrt{n}}+p^2\log(n)+\frac{n}{q}\cdot p^2\sqrt{n}\right)\xrightarrow{P_x}\infty,$$

d satisfies

$$\frac{d \cdot n^{\frac{\rho}{2}} p^2}{4\lambda_{\max}(\mathcal{X}\mathcal{X}'/n)} - \frac{np}{2} - K_1\left(\frac{p\|Y\|^2}{\sqrt{n}} + p^2\log(n) + \frac{n}{q}\cdot p^2\sqrt{n}\right) \xrightarrow{P_x} \infty,$$

and $n = O_p(q)$*, where* $q := \min_{1\leq j\leq p}\{m_j\}$ *with* $m_1, \ldots, m_p$ *corresponding to the full model (i.e., all components active), and* $\varepsilon = \Lambda_g \cdot \widetilde{\varepsilon}$ *for some* $\widetilde{\varepsilon}$ *not depending on* Σ *or* A_g*.*

An important attribute of Condition 3.4 is the requirement that while the dimension of the VAR(1) model, p, can be taken to infinity, it must be exceeded polynomially by the number of observed time instances, n. This is in contrast to the model selection consistency result established for the high-dimensional linear regression setting in (Williams & Hannig 2019), where p was allowed to grow sub-exponentially in n. The primary difference here is that we derive model selection consistency results for the multivariate VAR model setting which are robust to model misspecification, namely the assumption of Gaussian VAR model errors. Such a robust generalized fiducial result requires (to the best of our understanding) non-asymptotic second moment concentration bounds. High-dimensional ($p > n$) consistency results require exponential tail bounds when establishing concentration of data-dependent quantities such as in Lemma 3.8 in the next section, and exponential tail bounds here are intimately related to the assumption of Gaussianity.

Note that no assumption of sparsity is made in any of the conditions. This section concludes with a definition of various quantities that will be referenced in the next section, and throughout the proofs.

Definition 3.5. *N_1 is any positive constant such that $n \geq N_1$ implies*

$$1 - \frac{1-c^{2n}}{n(1-c^2)} \leq 1.$$

N_2 is any positive constant such that $n \geq N_2$ implies

$$1 + c^2 - 2\frac{c^2 - (c^2)^{n+1}}{n(1-c^2)} \leq 1 + c^2.$$

Additionally, N_3 is defined as in (S3).

$$V_1 := 16(\sigma^0_{\max})^4\left[\frac{p^6 n^{1-\frac{3\rho}{2}}}{\xi}\cdot\left(\frac{\|\Gamma_n(0)\|_2^2}{(\sigma^0_{\max})^4 p} + \frac{(3+c^4)}{(1-c^2)^3 n}\right) + \frac{\delta^{-2}p^2}{(1-c^2)^3 n^{1-2\rho}} + \frac{(3+c^4)p^6 n^{\frac{\rho}{2}}}{(1-c^2)^3\xi}\right], \tag{9}$$

with $\xi = \frac{2\delta^2}{9\Lambda_g}\varepsilon$. *The alternate* $\widetilde{V}_1$ *denotes* V_1 *with* ε *replaced by* $c^2 \cdot \frac{9n^{1+\frac{\rho}{2}}p^3\Lambda_{g_o}}{2}$.

$$V_2 := 4\delta^{-2}\frac{(\sigma^0_{\max})^4(1+c^2)}{(1-c^2)^3} \cdot \frac{2\min\{|G_o|,p\}^2}{n}. \tag{10}$$

$$V_3 := \frac{V_2}{4} + \delta^{-2}\left[\frac{2p(\sigma^0_{\max})^2\min\{|G_o|,p\}}{n(1-c^2)} + \frac{p(p+1)}{n}\right]. \tag{11}$$

3.2 Results

Our strategy for establishing graph selection consistency in Theorem 3.12 is largely composed of the contents of Lemmas 3.6 and 3.8 and Theorems 3.9 and 3.10. Lemmas 3.6 and 3.8 describe, respectively, the generalized fiducial concentration of the VAR(1) transition matrix around its least squares estimate and the concentration of the least squares estimate around an approximation to its expectation. The probability bounded in Lemma 3.6 is with respect to the joint generalized fiducial distribution of A_g and Σ. In contrast, Lemma 3.8 is a concentration inequality with respect to the data-generating mechanism (2) which derives its distribution from the errors $U^{(t)} \overset{\text{iid}}{\sim} \mathrm{N}_p(0, I_p)$ for $t \in \{1,\ldots,n\}$. In what follows we chose $\varepsilon = \Lambda_g \cdot \widetilde{\varepsilon}$ for some $\widetilde{\varepsilon}$ not depending on Σ or A_g.

Lemma 3.6. *For any* G *with* $|G| \leq np$,

$$P\Big(\|\mathcal{Z}_G'\mathcal{W}^{-1}\mathcal{Z}_G(\boldsymbol{\alpha}_G - \widehat{\boldsymbol{\alpha}}_g)\|^2 \geq \varepsilon\Big) \leq \frac{|G|\sqrt{2\Lambda_g}}{\sqrt{\pi\varepsilon}}e^{-\frac{\varepsilon}{2\Lambda_g}},$$

where $\widehat{\boldsymbol{\alpha}}_g := \left(\mathcal{Z}_G'\mathcal{Z}_G\right)^{-1}\mathcal{Z}_G'Y$, *and* $\Lambda_g := \|\mathcal{W}^{-\frac{1}{2}}\mathcal{Z}_G\|_F^2$.

Recall that Λ_g, which comes from the proof of this lemma, is a key component of our suggested default ε in (6) and of Condition 3.4. This results from the fact that ε must control for Λ_g in order to establish the well-behaved concentration of the generalized fiducial distribution of $\boldsymbol{\alpha}_G$ which is exhibited by this lemma. The $\mathcal{W}^{-\frac{1}{2}}$ plays the role of appropriately scaling the design matrix $\mathcal{Z}_G$. Observe that for the full model $G = \{1,\ldots,p^2\}$,

$$\Lambda = \|\mathcal{W}^{-\frac{1}{2}}\mathcal{Z}\|_F^2 = \mathrm{tr}(\mathcal{Z}'\mathcal{W}^{-1}\mathcal{Z}) = \mathrm{tr}\big((\mathcal{X}\mathcal{X}') \otimes \Sigma^{-1}\big) = \mathrm{tr}(\mathcal{X}\mathcal{X}') \cdot \mathrm{tr}(\Sigma^{-1}),$$

which gives

$$E_x(\Lambda) = n \cdot \mathrm{tr}(\Gamma_n(0)) \cdot \mathrm{tr}(\Sigma^{-1}). \tag{12}$$

Thus, for a given graph G, Λ_g is a combined measure of the covariance or dependence among the p univariate time-series in the VAR model, the contemporaneous error precision matrix, and the number of observed instances of the time-series. This is what makes Λ_g effective as apart of ε in the h-function for determining the *ε-admissibility* of a given $\boldsymbol{\alpha}_G$. Lemma 3.7 gives a probabilistic bound on Λ_g as a function of n and p, given the h-function constraint that

$\min_{1\le j\le p}\{m_j^g\} \ge d$.

Lemma 3.7. *For any G,*

$$P\Bigg(\Lambda_g \ge n^{1+\frac{\rho}{2}}p^3, \min_{1\le j\le p}\{m_j^g\} \ge d\Bigg) \le e^{-\left(\frac{d\cdot n^{\frac{\rho}{2}}p^2}{4\lambda_{\max}(\mathcal{X}\mathcal{X}'/n)}-\frac{np}{2}\right)}2^{-\frac{|G|}{2}},$$

where $\Lambda_g := \|\mathcal{W}^{-\frac{1}{2}}\mathcal{Z}_G\|_F^2$.

Next, consider the concentration of the least squares estimate.

Lemma 3.8. *Assume Condition 3.1 holds. Then for all $n \ge \max\{N_1, N_2\}$, and for any G with $|G| \le np$,*

$$P_x\Bigg(\|\widehat{\boldsymbol{\alpha}}_g - \big(E_x(\mathcal{Z}_G'\mathcal{Z}_G)\big)^{-1}E_x(\mathcal{Z}_G'Y)\|^2 \ge \frac{2\varepsilon}{9n^{1+\frac{\rho}{2}}p^3\Lambda_g}\Bigg) \le V_1,$$

where V_1 is as in (9).

Materially, the three preceding lemmas are needed in the proofs of Theorems 3.9 and 3.10, presented next. These theorems are results about the behavior of the EAS methodology coupled with the generalized fiducial distribution (i.e., the Jacobian term); they are analogous to studying the behavior of given priors for a (Bayesian) posterior distribution. Theorem 3.9 is a non-asymptotic concentration inequality which yields an upper bound on the rate at which the expected value (w.r.t. the joint generalized fiducial distribution of A_g and Σ) of the h-function times the Jacobian term diverges for non-ε-*admissible* graphs, G.

Theorem 3.9. *Take any G with $G \not\subseteq G_o$ and $|G| \le np$, and assume Conditions 3.1 and 3.3 hold. Then for all $n \ge \max\{N_1, N_2\}$,*

$$\begin{aligned}E\Big(h\big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big)|\widetilde{\mathcal{D}}_g'\widetilde{\mathcal{D}}_g|^{\frac{1}{2}}\Big) \le\ & e^{\frac{1}{2}(1-c)^{-2}\left(r_{\max}^g+(1+c)^2\right)\frac{\|Y\|^2}{\sqrt{n}}-\frac{|G|+p}{2}} \\ & \times\left(\frac{3|G|\sqrt{\Lambda_g}}{\sqrt{\pi\varepsilon}}e^{-\frac{\varepsilon}{9\Lambda_g}} + e^{-\left(\frac{d\cdot n^{\frac{\rho}{2}}p^2}{4\lambda_{\max}(\mathcal{X}\mathcal{X}'/n)}-\frac{np}{2}\right)}2^{-\frac{|G|}{2}+1}\right)\end{aligned}$$

with probability exceeding $1 - V_1$, where V_1 is as in (9), $r_{\max}^g := \max_{1\le j\le p}|r_j^g|$.

Conversely, Theorem 3.10 is a non-asymptotic lower bound on the h-function times the Jacobian term for the oracle graph, G_o.

Theorem 3.10. *Assume Conditions 3.1, 3.2, and 3.4 hold. Then for all $n \ge \max\{N_1, N_2, N_3\}$, with N_3 and the fixed $K_3 \in (0,1)$ defined by (S3),*

$$P_x\Bigg(E\Big(h\big(\boldsymbol{\alpha}_{G_o}, \{\sigma_j\}\big)|\widetilde{\mathcal{D}}_{g_o}'\widetilde{\mathcal{D}}_{g_o}|^{\frac{1}{2}}\Big) \ge (1-K_3)e^{\frac{|G_o|+p}{4}}\Bigg) \ge 1 - V_1 - \widetilde{V}_1 - 2V_2 - 2e^{-\frac{np}{4}} - V_3,$$

where V_1 and $\widetilde{V}_1$, V_2, and V_3 are as in (9), (10), and (11), respectfully.

Before stating the main result of this paper one final condition, Condition 3.11, is needed. In its absence a less strong, yet still meaningful statement of posterior-like graphical consistency holds; we formulate this alternative statement as Corollary 3.13. The importance of Condition 3.11 is that it covers the gap left open in Theorem 3.9 since the theorem only bounds the generalized fiducial probability of non-ε-*admissible* graphs (i.e., $G \not\subseteq G_o$).

Condition 3.11. *For the positive constant K_2 specified in (13),*

$$\max_{G:G\subset G_o} \left\{ e^{K_2\left(\frac{p\|Y\|^2}{\sqrt{n}}+p^2\log(n)\right)} \prod_{j=1}^{p} \left[\frac{(m_j^{g_o})^{\frac{n-|r_j^{g_o}|}{2}}}{(m_j^{g})^{\frac{n-|r_j^{g}|}{2}}} \right] \right\} \xrightarrow{P_x} 0$$

as $n \to \infty$ or $n, p \to \infty$.

Recall from (5) that m_j^g is the univariate RSS, corresponding to graph G, for the j^{th} component of the VAR(1) model. Hence, this condition is a statement that the product of the ratio of RSS components for the true graph over that of any strict sub-graph, taken to a power on the order of n, will vanish at a rate of $\exp\{\frac{p\|Y\|^2}{\sqrt{n}}\} = O_p(\exp\{p^2\sqrt{n}\})$. This is not unreasonable to expect since for each $j \in \{1,\ldots,p\}$, $m_j^g = O_p(n)$, $m_j^{g_o} = O_p(n)$, $m_j^{g_o} \leq m_j^g$ for $G \in \{G : G \subset G_o\}$, and an explicit condition about the oracle model being sufficiently better fitting than all sub-models is typical of model consistency results.

The main result of our paper, a statement of pairwise graphical selection consistency for the constructed EAS methodology, is now presented. This result demonstrates that the generalized fiducial probability of the oracle graph will asymptotically dominate that of all other graphs. Note that there is no assumption of sparsity.

Theorem 3.12 (pairwise selection consistency)**.** *Given Conditions 3.1-3.11, for any $G \subseteq \{1,\ldots,p^2\} \setminus G_o$,*

$$\frac{r(G \mid Y)}{r(G_o \mid Y)} \xrightarrow{P_x} 0$$

as $n \to \infty$ or $n, p \to \infty$.

If Condition 3.11 is violated, Corollary 3.13 demonstrates that the generalized fiducial mass function $r(G \mid Y)$ will concentrate asymptotically on the subset of graphs $\{G : G \subseteq G_o\}$. In practice, for sufficiently large n, this means that there will be a few graphs which the algorithm visits frequently, and the largest one (in cardinality) likely contains the greatest number of the oracle components.

Corollary 3.13 (pairwise selection consistency)**.** *Relaxing Condition 3.11 in Theorem 3.12 gives, for any $G \subseteq \{1,\ldots,p^2\} \setminus \{G : G \subseteq G_o\}$,*

$$\frac{r(G \mid Y)}{r(G_o \mid Y)} \xrightarrow{P_x} 0$$

as $n \to \infty$ or $n, p \to \infty$.

The additional corollary stated next demonstrates that the EAS methodology will concentrate all generalized fiducial mass on the true model, asymptotically, for fixed p.

Corollary 3.14 (strong selection consistency, fixed p)**.** *Given Conditions 3.1-3.11 and fixed p,*

$$r(G_o \mid Y) \xrightarrow{P_x} 1$$

as $n \to \infty$.

Note the following short remark about the meaning of the difference between *pairwise* and *strong* model selection consistency. The statement of *strong* graph selection consistency is essentially a statement that the true model will be assigned large probability and all other models will be assigned small probabilities. Conversely, the implication of *pairwise* graph selection consistency is that the probability assigned to the true model will be large relative to each of the other model probabilities, individually, but that all models (including the true model) may have small probabilities. Such a phenomenon is common for model selection paradigms in which the set of candidate models grows very fast with dimension (i.e., like 2^{p^2} in the case of a VAR(1) model).

The next subsection illustrates the additional attribute that our model selection consistency results are robust to model misspecification, namely, the assumption of Gaussian VAR model errors.

3.3 Standalone supporting results

This subsection provides five lemmas which were foundational to our proof techniques for establishing our theory for the EAS methodology. Non-asymptotic moment bounds on products of $\mathcal{X}$ and $\mathcal{U}$ (in the VAR model formulation (1)), with respect to n and p, are the building blocks for any theoretical pursuit of understanding high-dimensional, multivariate VAR models. These results are essentially a collection of second moment bounds of the quantities and cross-quantities in $\Omega := \frac{1}{n}\left(\begin{smallmatrix} \mathcal{X}\mathcal{X}' & \mathcal{X}\mathcal{U}' \\ \mathcal{U}\mathcal{X}' & \mathcal{U}\mathcal{U}' \end{smallmatrix}\right)$, and establish the notion that our preceding fiducial consistency results will remain true under model misspecification. This is due to the fact that as long as the VAR(1) model errors are independent within and across time and there exist bounded fourth moments (i.e., components appearing in $E_x(\Omega^2)$), the following collection of lemmas will remain true (up to some constants of proportionality). And as a consequence, our generalized fiducial consistency results (which assume Gaussian data) will still hold even if the true data is not Gaussian.

Lemma 3.15. *Assume $\|A^0\|_2 \leq c$. Then for all $n \geq \max\{N_1, N_2\}$,*

$$P_x\Big(\big[\lambda_{\min}\big(\Omega - E_x(\Omega)\big)\big]^2 > \delta^2\Big) \leq V_3,$$

where V_3 is as in (11), $\Omega := \frac{1}{n}\left(\begin{smallmatrix} \mathcal{X}\mathcal{X}' & \mathcal{X}\mathcal{U}' \\ \mathcal{U}\mathcal{X}' & \mathcal{U}\mathcal{U}' \end{smallmatrix}\right)$, and $E_x(\Omega) = \left(\begin{smallmatrix} \Gamma_n(0) & \\ & I_p \end{smallmatrix}\right)$.

Lemma 3.16. *Assume* $\|A^0\|_2 \leq c$. *Then for all* $n \geq N_1$,

$$\frac{1}{n^2} tr\Big(E_x(\mathcal{X}\mathcal{U}'\mathcal{U}\mathcal{X}')\Big) \leq \frac{p(\sigma^0_{\max})^2 \min\{|G_o|, p\}}{n(1-c^2)}.$$

Lemma 3.17. *Assume* $\|A^0\|_2 \leq c$. *Then for all* $n \geq N_2$,

$$tr\Big(\frac{1}{n^2} E_x\big((\mathcal{X}\mathcal{X}')^2\big) - \Gamma_n^2(0)\Big) \leq \frac{\delta^2}{4} V_2.$$

Lemma 3.18. *Assume* $\|A^0\|_2 \leq c$. *Then,*

$$\frac{1}{n^2} tr\Big(E_x(\mathcal{X}\mathcal{X}'\mathcal{X}\mathcal{U}'A^0)\Big) \leq \frac{2(\sigma^0_{\max})^3 c^2 \min\{|G_o|, p\}^2}{(1-c^2)^2 n}.$$

Lemma 3.19. *Assume that Condition 3.1 holds. Then for all* $n \geq N_2$,

$$P_x\Big(\lambda_{\min}(\mathcal{X}\mathcal{X}'/n) \geq \delta/2\Big) \geq 1 - V_2,$$

where V_2 *is as in (10).*

4 Simulation results

While the theoretical pursuits of this paper have been focused on the conditions and supporting lemmas/theorems needed for the EAS procedure to assign the highest probability to the oracle graph with probability converging to 1 as $n, p \to \infty$, we ultimately designed the EAS approach with more practical intuitions in mind. In applications, the true data-generating model, G_o, may itself contain redundant information (i.e., unnecessary active components), and through our h-function methodology we are able to focus on recovering only the necessary active components. In doing so, at least for finite samples the EAS approach re-defines what is meant by the true graph. The purpose of our asymptotic considerations was to illustrate the conditions needed for our re-defined notion of the true graph to correspond precisely to the oracle graph.

In this section, we demonstrate on synthetic data that when the theoretical conditions are satisfied the EAS procedure performs as our asymptotic theory suggests, and is also able to perform as well as or better than existing methods in high-dimensional settings with respect to out-of-sample prediction error and estimation error. In fact, we find and present evidence to suggest that Conditions 3.1 and 3.2 are useful for high-dimensional settings. Moreover, Condition 3.1 is a simple and verifiable condition for actual observed data which informs of the sample size needed for competitive performance and is so well calibrated that we demonstrate deteriorating performance when it is not satisfied.

Furthermore, the EAS algorithm does not require any tuning parameter to achieve at or better than the out-of-sample predictive performance of competing methods such as LASSO or elastic net. The latter, more conventional methods, require cross-validation over a grid of

tuning parameters, and the appropriateness of the grid depends on the scaling of the data (i.e., Σ). On the contrary, via our Λ_g component in $\varepsilon = \Lambda_g \cdot \max\left\{1, n^{.51}p^2\Big(.5\log(\log(n))|G| - |G_o|\Big)\right\}$ (see (6) with $\rho = .49$) the EAS algorithm is scale invariant.

For numerical results, the component d in the h-function is set at $d = \min_{1\leq j\leq p}\{m_j^{g_{\text{enet}}}\}/10$, where G_{enet} are the active components estimated by elastic net. And as discussed previously, the constraint $\|A_g\|_2 \leq c$ in the h-function is replaced with $\|A_g\|_2 < 1$ since c is not available on real data.

In the following two subsections we present both low ($p = 4, n = 120$) and high ($p = 10, n = 20$ and $p = 30, n = 180$) dimensional simulation studies on synthetic data generated according to model (2). For each of 100 random data-generating seeds, the transition matrix is randomly generated according to each of the five patterns described in Han et al. (2015). In each instance of a transition matrix A^0 the p diagonal components are active, and for patterns with additional randomly assigned active/inactive components the probability of each component being generated as active is .01. Values of each diagonal component are assigned by sampling from the $N(\pm 12, 1)$ distribution, while off-diagonal component values are assigned by sampling from the $N(\pm 3, 1)$ distribution. As is common practice (e.g., Han et al. 2015), after a given A^0 is randomly generated it is rescaled so that $\|A^0\|_2 = .5 =: c$, and as in Han et al. (2015) the contemporaneous error covariance matrix $\Sigma^0 := I_p$.

In all simulation designs, the performance of the EAS algorithm is compared to that of LASSO and elastic net implementations, and to a recent "direct estimation of high-dimensional stationary VAR" estimation procedure proposed by Han et al. (2015) which is formulated as a linear program (we denote this procedure by DELP for "direct estimation linear program"). The LASSO and elastic net routines are implemented from the *Python* module `scikit-learn` Pedregosa et al. (2011), along with their builtin cross-validation procedures for time-series data. For the DELP routine, the authors of Han et al. (2015) were kind enough to provide their *R* code. However, we had to supplement their provided code by writing code to implement the cross-validation procedure they propose in Han et al. (2015) for selecting their tuning parameter. Note that we generate synthetic data consistent with that described in Han et al. (2015) so that the scaling of the data is appropriate for their default grid of tuning parameters for cross-validation.

The entirety of the simulation study was computed in parallel on a computing cluster, and completed in approximately one day of run time. The code/workflow for reproducing all numerical results presented in this paper can be found at `https://jonathanpw.github.io/research`.

4.1 Definitions of performance metrics

A variety of metrics are considered for evaluating performance across procedures. For each random generator seed for each simulation design, $2n$ instances of the time-series are generated with $X^{(0)} = 0_{p\times 1}$. The first n are used for estimation, and the last n are set aside as an out-of-sample test set. As in Han et al. (2015), on the out-of-sample test set we compute the L_2

prediction error, $\frac{1}{n}\|\mathcal{Y} - \widehat{A}\mathcal{X}\|_2$, and the L_F prediction error, $\frac{1}{n}\|\mathcal{Y} - \widehat{A}\mathcal{X}\|_F$, where $\widehat{A}$ represents the estimated transition matrix on the first n, in-sample, time instances. As in Basu et al. (2015) and Ghosh et al. (2018), we also calculate the estimation error, $\|\widehat{A} - A^0\|_F / \|A^0\|_F$. For the EAS procedure, $\widehat{A}$ is computed analogously to Bayesian model averaging, with least squares estimates used for every visited graph in the MCMC chain.

Random pattern transition matrix

	p = 4, n = 120				
	oracle	eas	delp	lasso	enet
L2	1.27	1.28	1.31	1.3	1.3
	(0.11)	(0.11)	(0.12)	(0.12)	(0.12)
LF	2.01	2.02	2.03	2.03	2.03
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
est err	0.17	0.21	0.32	0.32	0.33
	(0.06)	(0.1)	(0.11)	(0.1)	(0.1)
$\lvert G_{\text{MAP}}\rvert$	4.12	4.0	7.84	7.94	8.39
	(0.35)	(0.32)	(3.34)	(2.97)	(3.24)
FPR		0.01	0.32	0.33	0.36
		(0.02)	(0.28)	(0.25)	(0.27)
FNR		0.04	0.01	0.01	0.01
		(0.09)	(0.05)	(0.05)	(0.05)
$\widehat{r}(G_o \mid Y)$		0.7			
		(0.32)			
$\#\{G_{\text{MAP}} = G_o\}$		0.81	0.08	0.11	0.11
	r.h.s. Condition 3.1 = 10.10 (s.e. 0.90) vs 5				
	prop data sets Condition 3.2 satisfied = 0.83				

Table 1: See Section 4.1 for definitions of each performance metric, except for the last two which are described in Section 4.2. All metrics are quantities averaged over 100 generated data sets, and standard errors are in parentheses. The 'oracle' column displays corresponding characteristics in the case that the oracle graph, G_o, is known, and using the least squares estimate of A^0. Note that for Condition 3.1, $4(1 + c^2) = 5$. Recall that a new set of active components G_o are generated for each data set, which gives the variability for $|G_{\text{MAP}}|$ in the 'oracle' column.

Additionally, we report $|G_{\text{MAP}}|$ as the number of nonzero (or active) components in the estimated graph for the frequentist LASSO, elastic net, and DELP procedures, and as the number of active components in the most frequently visited graph (i.e., maximum a-posteriori probability or MAP) for the MCMC-based EAS algorithm. The false positive rate (FPR) is computed as the number of the p^2 components in the estimated transition matrix incorrectly set active, as a proportion of the number of truly inactive components. Conversely, the false negative rate (FNR) is computed as the number of the p^2 components in the estimated transition matrix incorrectly set inactive, as a proportion of the number of truly active components. For the EAS procedure, the FPR and FNR are computed based on the estimated G_{MAP}.

4.2 Low-dimensional setting

This first simulation design serves to demonstrate that the EAS procedure performs consistently with what the theory in Section 3 suggests for data with $p^2 < n$. For this simulation we present two additional performance metrics, $\widehat{r}(G_o \mid Y)$ and $\#\{G_{\text{MAP}} = G_o\}$. The former is the estimated generalized fiducial probability of the oracle model, calculated as the number of times the MCMC algorithm visited G_o divided by the number of steps of the chain. This metric is only available within the EAS framework because relative model probabilities are computed. The latter metric, $\#\{G_{\text{MAP}} = G_o\}$, is the proportion, over all 100 generated data sets, of instances in which the estimated G_{MAP} corresponds precisely to G_o.

Observe from Table 1 that the EAS procedure performs very competitively with these existing methods; better average performance metric values across the board, but all routines are within about one standard error of each other. Furthermore, the EAS algorithm selected a G_{MAP} with 3-4 fewer active components, on average, with $G_{\text{MAP}} = G_o$ for 81 of the 100 of the data sets. This is far better graph selection than the competing methods which consistently over-select active components. Note that based on the proportion of data sets in which Condition 3.2 is satisfied, the oracle model is only identifiable for the EAS algorithm in 83 percent of the data sets. In other words, our theory would suggest that the EAS procedure should identify the true model in 83 of the 100 data sets considered, and in actuality the EAS algorithm identified the true model in 81 of the 100 data sets.

4.3 High-dimensional setting

Band pattern transition matrix

	p = 10, n = 20					p = 30, n = 180				
	oracle	eas	delp	lasso	enet	oracle	eas	delp	lasso	enet
L2	3.04	3.29	4.13	2.94	2.9	1.92	2.02	2.03	2.02	2.02
	(0.68)	(0.65)	(5.9)	(0.49)	(0.45)	(0.09)	(0.09)	(0.1)	(0.1)	(0.1)
LF	3.46	3.54	3.55	3.4	3.39	5.53	5.64	5.67	5.65	5.65
	(0.22)	(0.24)	(0.71)	(0.18)	(0.18)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)
est err	1.07	1.24	1.15	0.97	0.94	0.34	0.63	0.68	0.64	0.65
	(0.17)	(0.18)	(0.77)	(0.04)	(0.05)	(0.03)	(0.06)	(0.05)	(0.06)	(0.06)
$\lvert G_{\text{MAP}}\rvert$	28.0	11.64	8.89	3.43	19.63	88.0	22.32	43.43	49.78	60.85
	(0.0)	(2.88)	(23.38)	(4.56)	(18.44)	(0.0)	(2.92)	(25.39)	(11.55)	(29.62)
FPR		0.1	0.08	0.02	0.17		0.0	0.01	0.02	0.03
		(0.03)	(0.24)	(0.04)	(0.18)		(0.0)	(0.03)	(0.01)	(0.03)
FNR		0.83	0.88	0.93	0.74		0.75	0.63	0.59	0.57
		(0.06)	(0.23)	(0.07)	(0.21)		(0.03)	(0.05)	(0.05)	(0.07)
	r.h.s. Condition 3.1 = 0.7112 (s.e. 0.2384) vs 5 prop data sets Condition 3.2 satisfied = 0					r.h.s. Condition 3.1 = 5.7756 (s.e. 0.4147) vs 5 prop data sets Condition 3.2 satisfied = 0				

Table 2: See caption for Table 1.

Cluster pattern transition matrix

	p = 10, n = 20					p = 30, n = 180				
	oracle	eas	delp	lasso	enet	oracle	eas	delp	lasso	enet
L2	2.66	3.52	10.35	3.22	3.16	1.91	1.96	2.03	2.01	2.01
	(0.37)	(0.85)	(44.59)	(0.51)	(0.48)	(0.1)	(0.12)	(0.11)	(0.11)	(0.11)
LF	3.28	3.6	4.05	3.51	3.48	5.5	5.55	5.63	5.61	5.61
	(0.16)	(0.24)	(2.06)	(0.18)	(0.18)	(0.06)	(0.07)	(0.07)	(0.06)	(0.06)
est err	0.48	1.08	1.42	0.95	0.92	0.17	0.34	0.5	0.46	0.46
	(0.12)	(0.16)	(1.49)	(0.05)	(0.06)	(0.03)	(0.09)	(0.05)	(0.05)	(0.05)
$\lvert G_{\text{MAP}}\rvert$	10.39	12.16	18.51	4.33	21.05	31.24	27.64	42.28	47.76	48.14
	(0.68)	(2.28)	(34.53)	(5.52)	(19.45)	(1.26)	(2.47)	(23.75)	(8.28)	(8.4)
FPR		0.09	0.17	0.03	0.18		0.0	0.01	0.02	0.02
		(0.03)	(0.35)	(0.04)	(0.19)		(0.0)	(0.03)	(0.01)	(0.01)
FNR		0.63	0.68	0.8	0.55		0.14	0.04	0.04	0.04
		(0.14)	(0.33)	(0.19)	(0.29)		(0.08)	(0.04)	(0.04)	(0.04)
	r.h.s. Condition 3.1 = 0.6941 (s.e. 0.2265) vs 5 prop data sets Condition 3.2 satisfied = 0					r.h.s. Condition 3.1 = 6.0315 (s.e. 0.4613) vs 5 prop data sets Condition 3.2 satisfied = 0				

Table 3: See caption for Table 1. Recall that a new set of active components G_o are generated for each data set, which gives the variability for $|G_{\text{MAP}}|$ in the 'oracle' column.

The tables in this section display the results of two high-dimensional simulation designs in which $p^2 > n$, and for all five transition matrix patterns. An important distinction to observe between the two designs, for all transition matrix patterns, is that for the $p = 10, n = 20$ case Condition 3.1 is never satisfied, while it is always satisfied for the $p = 30, n = 180$ case. This occurrence is by design to demonstrate the deteriorated performance of the EAS algorithm when this important, well-calibrated, and verifiable condition is not satisfied. In the $p = 30, n = 180$ case the EAS algorithm performs just as well, or better than the competing methods, with respect to all metrics.

Notice also that the high-dimensional numerical results presented in this section do not list $\widehat{r}(G_o \mid Y)$ nor $\#\{G_{\text{MAP}} = G_o\}$ as performance metrics. For each of the estimation methods, the metric $\#\{G_{\text{MAP}} = G_o\}$ (and $\widehat{r}(G_o \mid Y)$ for EAS) produces zeros in almost all cases. For the EAS algorithm, this is due to the fact that Condition 3.2 is never satisfied for these high-dimensional simulation designs, and so the oracle model is *not* identifiable for the EAS procedure. However, we would not necessarily expect Condition 3.2 to be satisfied when $p^2 > n$ since our theory does not apply.

Moreover, recall that in finite samples, and particularly high-dimensional, settings with highly-correlated data the EAS framework was developed with the intuition that the oracle graph itself may not be ε-*admissible*. In these settings, the EAS methodology re-defines the notion of the 'true' graph to be some non-redundant subgraph of the oracle graph, at least non-asymptotically. This is validated empirically in the tables that follow by observing that when Condition 3.1 is satisfied the EAS algorithm almost always requires fewer active components to achieve on par or better performance than the competing methods, with respect to all metrics.

Recall also that the EAS algorithm has no tuning parameter, while the competing methods use cross-validation to optimize out-of-sample prediction accuracy.

Hub pattern transition matrix

	p = 10, n = 20					p = 30, n = 180				
	oracle	eas	delp	lasso	enet	oracle	eas	delp	lasso	enet
L2	3.06	3.3	6.52	3.02	2.98	1.93	2.03	2.05	2.03	2.03
	(0.64)	(0.66)	(29.28)	(0.61)	(0.58)	(0.09)	(0.09)	(0.1)	(0.1)	(0.1)
LF	3.44	3.54	3.68	3.41	3.4	5.53	5.63	5.66	5.64	5.65
	(0.21)	(0.2)	(1.53)	(0.2)	(0.2)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
est err	1.06	1.23	1.26	0.98	0.96	0.33	0.63	0.69	0.65	0.66
	(0.17)	(0.17)	(1.15)	(0.03)	(0.05)	(0.03)	(0.05)	(0.05)	(0.06)	(0.05)
$\lvert G_{\text{MAP}}\rvert$	26.0	11.95	11.9	2.89	19.74	78.0	21.97	41.24	47.64	61.35
	(0.0)	(2.48)	(29.42)	(4.84)	(20.78)	(0.0)	(2.72)	(24.41)	(11.38)	(32.45)
FPR		0.1	0.11	0.02	0.17		0.0	0.01	0.02	0.03
		(0.03)	(0.3)	(0.04)	(0.2)		(0.0)	(0.03)	(0.01)	(0.03)
FNR		0.82	0.86	0.94	0.73		0.73	0.6	0.57	0.54
		(0.07)	(0.29)	(0.09)	(0.25)		(0.03)	(0.05)	(0.05)	(0.07)
	r.h.s. Condition 3.1 = 0.7 (s.e. 0.2323) vs 5 prop data sets Condition 3.2 satisfied = 0					r.h.s. Condition 3.1 = 5.667 (s.e. 0.4128) vs 5 prop data sets Condition 3.2 satisfied = 0				

Table 4: See caption for Table 1.

Random pattern transition matrix

	p = 10, n = 20					p = 30, n = 180				
	oracle	eas	delp	lasso	enet	oracle	eas	delp	lasso	enet
L2	2.6	3.36	5.21	3.06	3.03	1.91	1.99	2.03	2.01	2.01
	(0.4)	(0.8)	(9.99)	(0.49)	(0.5)	(0.09)	(0.1)	(0.1)	(0.1)	(0.1)
LF	3.24	3.55	3.71	3.45	3.43	5.51	5.58	5.64	5.62	5.62
	(0.17)	(0.26)	(0.95)	(0.21)	(0.22)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
est err	0.49	1.07	1.2	0.95	0.92	0.2	0.44	0.55	0.51	0.51
	(0.15)	(0.16)	(0.82)	(0.06)	(0.07)	(0.03)	(0.08)	(0.05)	(0.05)	(0.05)
$\lvert G_{\text{MAP}}\rvert$	10.9	12.49	14.51	3.69	19.28	38.72	25.65	45.45	50.07	50.55
	(0.96)	(2.68)	(30.43)	(4.68)	(20.07)	(3.13)	(2.68)	(30.98)	(9.95)	(10.89)
FPR		0.1	0.13	0.02	0.17		0.0	0.02	0.02	0.02
		(0.03)	(0.31)	(0.03)	(0.19)		(0.0)	(0.03)	(0.01)	(0.01)
FNR		0.63	0.72	0.82	0.58		0.35	0.21	0.19	0.19
		(0.15)	(0.31)	(0.18)	(0.31)		(0.09)	(0.07)	(0.07)	(0.07)
	r.h.s. Condition 3.1 = 0.6961 (s.e. 0.2523) vs 5 prop data sets Condition 3.2 satisfied = 0					r.h.s. Condition 3.1 = 5.7805 (s.e. 0.43) vs 5 prop data sets Condition 3.2 satisfied = 0				

Table 5: See caption for Table 1. Recall that a new set of active components G_o are generated for each data set, which gives the variability for $|G_{\text{MAP}}|$ in the 'oracle' column.

Scale–free pattern transition matrix

	p = 10, n = 20					p = 30, n = 180				
	oracle	eas	delp	lasso	enet	oracle	eas	delp	lasso	enet
L2	3.37	3.19	8.31	2.86	2.83	1.94	2.0	2.01	2.01	2.0
	(0.81)	(0.75)	(30.93)	(0.47)	(0.47)	(0.09)	(0.1)	(0.1)	(0.1)	(0.1)
LF	3.51	3.49	3.83	3.36	3.35	5.53	5.61	5.62	5.63	5.62
	(0.22)	(0.22)	(1.73)	(0.19)	(0.18)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)
est err	1.36	1.32	1.56	0.99	0.96	0.52	0.84	0.87	0.9	0.86
	(0.27)	(0.19)	(1.76)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.06)	(0.05)
$\lvert G_{\text{MAP}}\rvert$	28.0	11.8	15.88	2.07	16.97	88.0	15.14	23.22	17.66	77.21
	(0.0)	(2.85)	(34.04)	(3.76)	(18.88)	(0.0)	(2.24)	(4.07)	(10.06)	(35.83)
FPR		0.1	0.15	0.01	0.15		0.0	0.01	0.0	0.06
		(0.03)	(0.34)	(0.03)	(0.18)		(0.0)	(0.0)	(0.0)	(0.03)
FNR		0.84	0.82	0.96	0.78		0.86	0.79	0.83	0.63
		(0.07)	(0.34)	(0.07)	(0.22)		(0.03)	(0.04)	(0.09)	(0.11)
	r.h.s. Condition 3.1 = 0.6866 (s.e. 0.2609) vs 5 prop data sets Condition 3.2 satisfied = 0					r.h.s. Condition 3.1 = 5.3767 (s.e. 0.4072) vs 5 prop data sets Condition 3.2 satisfied = 0				

Table 6: See caption for Table 1.

5 Real data application

As a final exposition of the EAS methodology developed for the VAR(1) model, this section presents results of implementing the algorithm on real data. Monthly closing stock price data for eight well-known companies from 1995-2018 are downloaded from Yahoo Finance via the R package `BatchGetSymbols` (Perlin 2019). First-differences of the time-series are used for stationarity, and the data is split into two time periods, 1995-2006 and 2007-2018 (plots of each time-series for each time period are available with the code for reproducing our results at `https://jonathanpw.github.io/research`). It is verified that Condition 3.1 is satisfied for the time period 2007-2018 (9.33 versus $8 = 4(1 + 1^2)$), but not for 1995-2006 (2.05 versus $8 = 4(1 + 1^2)$). This occurrence is useful for observing the performance of the EAS procedure on real data when the condition is and is not satisfied.

The results are displayed graphically in Figures 1 and 2. Nodes represent individual company stocks, and each edge label represents the marginal generalized fiducial (or posterior-like) inclusion probability of a particular component of A. That is the proportion of graphs, G, (over all MCMC-sampled graphs) in which each component (i.e., edge) of A is active. Line widths are proportional to inclusion probabilities, and inclusion probabilities less than .05 are omitted.

Interpretation of the findings on these data should be restricted to the time period 2007-2018 in which Condition 3.1 is satisfied. However, the real data analysis conducted here is not a thorough investigation of these time-series, but rather a "proof of concept" for how the EAS methodology can be useful on real data. A well qualified study would require considerable additional analysis of the data which is beyond the scope of our paper.

Nonetheless, the results do appear sensible. From Table 2, it is observed that seemingly

redundant time-series in the system such as for the two oil companies, Chevron and Exxon, do not have simultaneous marginal inclusions to a large extent, and the system is dominated by relatively few strong links. Since many of the considered stocks correspond to consumer goods corporations, it is reasonable that the results suggest the system has numerous links to and from the massive retailer Walmart, with an especially high link from the pharmaceutical giant Pfizer. Additionally, we see that the somewhat surprisingly strong link in Figure 1 between what we would suspect are unrelated corporations/stock prices, Ford and Pfizer, vanishes in Figure 2.

Note that such a graphical representation of the results, with marginal inclusion probabilities for all components of A, is not possible via frequentist nor Bayesian point estimation based procedures. This is a major advantage of estimating relative model probabilities (i.e., $r(G \mid Y)$) versus simply coefficients. MCMC-based approaches are computationally more expensive, but they provide more information for uncertainty quantification. The code/workflow for obtaining the real data and reproducing these results can be found at `https://jonathanpw.github.io/research`.

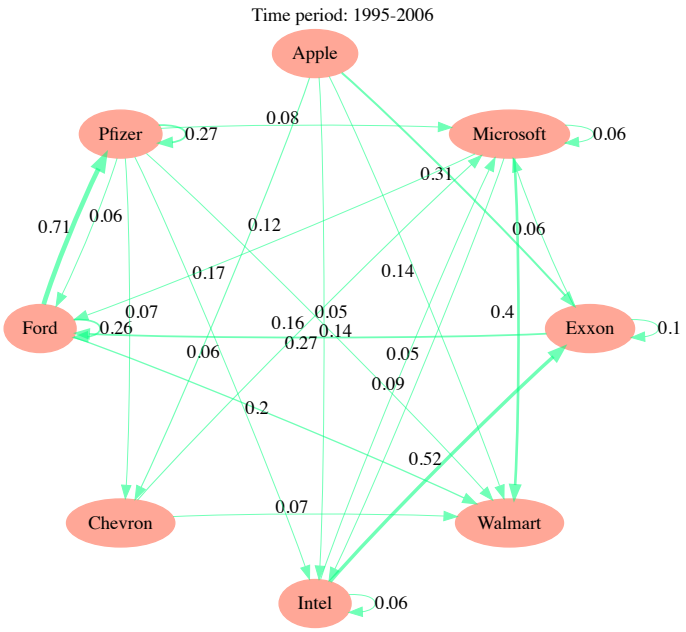

Figure 1: Directed graph of inclusion probabilities of components of the transition matrix, A, for monthly closing stock price of 8 companies. First differences of the data are used. Each edge label represents the marginal generalized fiducial (or posterior-like) inclusion probability of a particular component of A. That is, the proportion of graphs, G, (over all MCMC-sampled graphs) in which each component (i.e., edge) of A is active. Line widths are proportional to inclusion probabilities, and inclusion probabilities less than .05 are omitted.

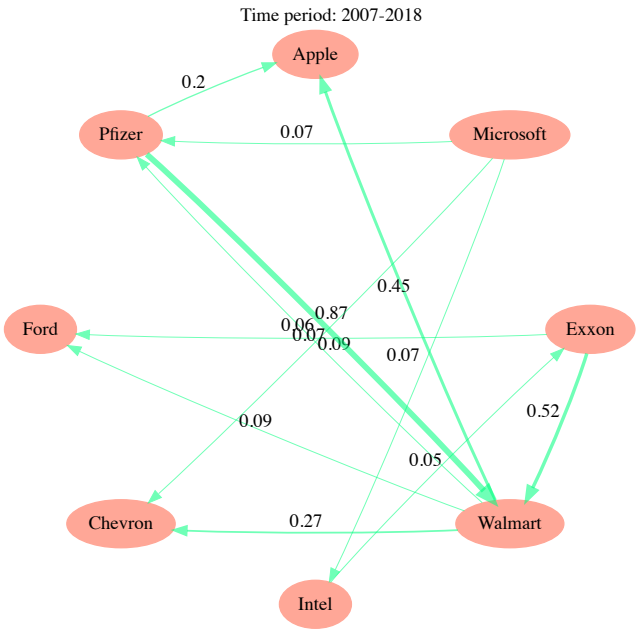

Figure 2: See description for Figure 1.

6 Concluding remarks

In summary, while BVAR models have been developed and explored empirically (primarily in the econometrics literature) there exist very few theoretical investigations of the repeated sampling properties for BVAR models in the literature. To the best of our knowledge, our established *pairwise* and *strong* model selection consistency results are the first of their kind in the BVAR literature. These types of results are sure to be followed by similar results in the high-dimensional BVAR literature, analogous to the emergence of model selection strong consistency results in the high-dimensional Bayesian linear regression literature such as Johnson & Rossell (2012), Narisetty & He (2014), Williams & Hannig (2019).

All things considered, while it is required for our theory that n exceeds some polynomial of p, consistent with our survey of the literature, it is claimed in Ghosh et al. (2018) that general posterior consistency results are not available for "large p small n" settings. Furthermore, our graphical selection consistency results provide a theoretical guarantee for model selection, which is stronger than establishing estimation consistency of a point estimator of the VAR model parameters, and our theory is robust to model misspecification.

Moreover, recall that in finite samples, and particularly high-dimensional, settings with highly-correlated data the EAS framework was developed with the intuition that the oracle graph itself may not be ε-*admissible*. In these settings, the EAS methodology re-defines the notion of the 'true' graph to be some non-redundant subgraph of the oracle graph, at least

non-asymptotically. Accordingly, with our EAS methodology, we hope to demonstrate the idea that to develop inherently scalable methodology the key may be to re-think what one should hope to recover for useful statistical inference from a data-generating model.

References

Ahelegbey, D. F., Billio, M. & Casarin, R. (2016), 'Sparse graphical vector autoregression: A bayesian approach', *Annals of Economics and Statistics/Annales d'Économie et de Statistique* **123/124**, 333–361.

Andrieu, C. & Roberts, G. O. (2009), 'The psuedo-marginal approach for efficient monte carlo computations', *The Annals of Statistics* **37**(2), 697–725.

Bańbura, M., Giannone, D. & Reichlin, L. (2010), 'Large bayesian vector auto regressions', *Journal of Applied Econometrics* **25**(1), 71–92.

Basu, S., Michailidis, G. et al. (2015), 'Regularized estimation in sparse high-dimensional time series models', *The Annals of Statistics* **43**(4), 1535–1567.

Berger, J. O., Bernardo, J. M. & Sun, D. (2009), 'The formal definition of reference priors', *The Annals of Statistics* **37**, 905–938.

Candes, E. & Tao, T. (2007), 'The dantzig selector: Statistical estimation when p is much greater than n', *The Annals of Statistics* **35**(6), 2313–2351.

Edlefsen, P. T., Liu, C. & Dempster, A. P. (2009), 'Estimating limits from Poisson counting data using Dempster–Shafer analysis', *The Annals of Applied Statistics* **3**, 764–790.

Fraser, D. A. S. (2019), 'The p-value Function and Statistical Inference', *The American Statistician* **73**(sup1), 135–147.

Ghosh, S., Khare, K. & Michailidis, G. (2018), 'High dimensional posterior consistency in bayesian vector autoregressive models', *Journal of the American Statistical Association* **0**(0), 1–14.

Giannone, D., Lenza, M. & Primiceri, G. E. (2015), 'Prior selection for vector autoregressions', *Review of Economics and Statistics* **97**(2), 436–451.

Han, F., Lu, H. & Liu, H. (2015), 'A direct estimation of high dimensional stationary vector autoregressions', *The Journal of Machine Learning Research* **16**(1), 3115–3150.

Hannig, J., Iyer, H., Lai, R. C. S. & Lee, T. C. M. (2016), 'Generalized fiducial inference: A reviev and new results', *Journal of American Statistical Association* **111**(515), 1346–1361.

Jameson, G. J. O. (2013), 'Inequalities for gamma function ratios', *American Math. Monthly* **120**, 936–940.

Johnson, V. E. & Rossell, D. (2012), 'Bayesian model selection in high-dimensional settings', *Journal of the American Statistical Association* **107**(498), 649–660.

Korobilis, D. (2013), 'Var forecasting using bayesian variable selection', *Journal of Applied Econometrics* **28**(2), 204–230.

Litterman, R. B. (1986), 'Forecasting with bayesian vector autoregressions?five years of experience', *Journal of Business & Economic Statistics* **4**(1), 25–38.

Loh, P.-L. & Wainwright, M. J. (2012), 'High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity', *The Annals of Statistics* **40**(3), 1637.

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

Martin, R. & Liu, C. (2015), *Inferential Models: Reasoning with uncertainty*, Vol. 145, CRC Press.

Narisetty, N. N. & He, X. (2014), 'Bayesian variable selection with shrinking and diffusing priors', *The Annals of Statistics* **42**(2), 789–817.

Negahban, S., Wainwright, M. J. et al. (2011), 'Estimation of (near) low-rank matrices with noise and high-dimensional scaling', *The Annals of Statistics* **39**(2), 1069–1097.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. (2011), 'Scikit-learn: Machine learning in Python', *Journal of Machine Learning Research* **12**, 2825–2830.

Perlin, M. (2019), *BatchGetSymbols: Downloads and Organizes Financial Data for Multiple Tickers*. R package version 2.4.
URL: *https://CRAN.R-project.org/package=BatchGetSymbols*

Schweder, T. & Hjort, N. L. (2016), *Confidence, likelihood, probability*, Vol. 41, Cambridge University Press.

Taraldsen, G. & Lindqvist, B. H. (2013), 'Fiducial theory and optimal inference', *The Annals of Statistics* **41**, 323–341.

Veronese, P. & Melilli, E. (2015), 'Fiducial and Confidence Distributions for Real Exponential Families', *Scandinavian Journal of Statistics* **42**, 471–484.

Williams, J. P. & Hannig, J. (2019), 'Nonpenalized variable selection in high-dimensional linear model settings via generalized fiducial inference', *Annals of Statistics* **47**(3), 1723–1753.

Xie, M. & Singh, K. (2013), 'Confidence distribution, the frequentist distribution estimator of a parameter: A review', *International Statistical Review* **81**, 3 – 39.

7 Appendix

This section provides proofs of the main theorem and its corollaries. See the supplementary material for proofs of all other results.

Proof of Theorem 3.12. Assume throughout this proof that $n \geq \max\{N_1, N_2, N_3\}$ (see Definition 3.5). From (8),

$$\frac{r(G \mid Y)}{r(G_o \mid Y)} = (2\pi)^{\frac{|G|-|G_o|}{2}} n^{\frac{|G_o|-|G|}{2}} \frac{E\Big(h\big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big)|\widetilde{\mathcal{D}}_g' \widetilde{\mathcal{D}}_g|^{\frac{1}{2}}\Big)}{E\Big(h\big(\boldsymbol{\alpha}_{G_o}, \{\sigma_j\}\big)|\widetilde{\mathcal{D}}_{g_o}' \widetilde{\mathcal{D}}_{g_o}|^{\frac{1}{2}}\Big)}$$
$$\times \prod_{j=1}^{p} \left[\frac{\big|(\mathcal{X}\mathcal{X}')_{r_j^{g_o}, r_j^{g_o}}\big|^{\frac{1}{2}}}{\big|(\mathcal{X}\mathcal{X}')_{r_j^{g}, r_j^{g}}\big|^{\frac{1}{2}}} \cdot \frac{\big(\frac{m_j^g}{2}\big)^{-\frac{n-|r_j^g|}{2}}}{\big(\frac{m_j^{g_o}}{2}\big)^{-\frac{n-|r_j^{g_o}|}{2}}} \cdot \frac{\Gamma\Big(\frac{n-|r_j^g|}{2}\Big)}{\Gamma\Big(\frac{n-|r_j^{g_o}|}{2}\Big)} \right].$$

From Jameson (2013),

$$\frac{\Gamma\Big(\frac{n-|r_j^g|}{2}\Big)}{\Gamma\Big(\frac{n-|r_j^{g_o}|}{2}\Big)} \leq \begin{cases} \Big(\frac{n-|r_j^{g_o}|}{2}\Big)\Big(\frac{n-|r_j^g|}{2}\Big)^{\frac{|r_j^{g_o}|-|r_j^g|}{2}-1} & \text{if } |r_j^{g_o}| - |r_j^g| \geq 1 \\ 1 & \text{if } |r_j^{g_o}| - |r_j^g| = 0 \\ \Big(\frac{n-|r_j^g|}{2} - 1\Big)^{\frac{-(|r_j^g|-|r_j^{g_o}|)}{2}} & \text{if } |r_j^{g_o}| - |r_j^g| \leq -1 \end{cases},$$

and so for $n - p \geq 4$,

$$\prod_{j=1}^{p} \frac{\Gamma\Big(\frac{n-|r_j^g|}{2}\Big)}{\Gamma\Big(\frac{n-|r_j^{g_o}|}{2}\Big)} \leq \prod_{j=1}^{p} \begin{cases} \big(\frac{n}{2}\big)^{\frac{|r_j^{g_o}|-|r_j^g|}{2}} & \text{if } |r_j^{g_o}| - |r_j^g| \geq 3 \\ \frac{n}{2} & \text{if } |r_j^{g_o}| - |r_j^g| = 2 \\ \frac{n}{2}\big(\frac{n-p}{2}\big)^{-\frac{1}{2}} & \text{if } |r_j^{g_o}| - |r_j^g| = 1 \\ 1 & \text{if } |r_j^{g_o}| - |r_j^g| = 0 \\ \big(\frac{n-p}{2} - 1\big)^{\frac{-(|r_j^g|-|r_j^{g_o}|)}{2}} & \text{if } |r_j^{g_o}| - |r_j^g| \leq -1 \end{cases}$$
$$\leq \prod_{j=1}^{p} \Big(\frac{n}{2}\Big)^{\max\left\{\frac{|r_j^{g_o}|}{2}, 1\right\}}$$
$$\leq \Big(\frac{n}{2}\Big)^{\frac{|G_o|}{2}+p}.$$

This bound, together with the simplification,

$$\prod_{j=1}^{p} \frac{\big(\frac{m_j^g}{2}\big)^{-\frac{n-|r_j^g|}{2}}}{\big(\frac{m_j^{g_o}}{2}\big)^{-\frac{n-|r_j^{g_o}|}{2}}} = 2^{\frac{|G_o|-|G|}{2}} \prod_{j=1}^{p} \frac{(m_j^{g_o})^{\frac{n-|r_j^{g_o}|}{2}}}{(m_j^g)^{\frac{n-|r_j^g|}{2}}}$$

gives

$$\frac{r(G \mid Y)}{r(G_o \mid Y)} \leq \Big(\frac{\pi}{n}\Big)^{\frac{|G|-|G_o|}{2}} \Big(\frac{n}{2}\Big)^{\frac{|G_o|}{2}+p} \frac{E\Big(h\big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big)|\widetilde{\mathcal{D}}'_g \widetilde{\mathcal{D}}_g|^{\frac{1}{2}}\Big)}{E\Big(h\big(\boldsymbol{\alpha}_{G_o}, \{\sigma_j\}\big)|\widetilde{\mathcal{D}}'_{g_o} \widetilde{\mathcal{D}}_{g_o}|^{\frac{1}{2}}\Big)} \times \prod_{j=1}^{p} \left[\frac{\big|(\mathcal{X}\mathcal{X}')_{r_j^{g_o}, r_j^{g_o}}\big|^{\frac{1}{2}}}{\big|(\mathcal{X}\mathcal{X}')_{r_j^{g}, r_j^{g}}\big|^{\frac{1}{2}}} \cdot \frac{(m_j^{g_o})^{\frac{n-|r_j^{g_o}|}{2}}}{(m_j^{g})^{\frac{n-|r_j^{g}|}{2}}} \right].$$

Further, by Lemmas S2.2 and S2.5,

$$\frac{r(G \mid Y)}{r(G_o \mid Y)} \leq \Big(\frac{\pi}{n}\Big)^{\frac{|G|-|G_o|}{2}} \Big(\frac{n}{2}\Big)^{\frac{|G_o|}{2}+p} \frac{E\Big(h\big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big)\Big) e^{\frac{1}{2}(1-c)^{-2}\left(r_{\max}^{g}+(1+c)^2\right)\frac{\|Y\|^2}{\sqrt{n}} - \frac{|G|+p}{2}}}{E\Big(h\big(\boldsymbol{\alpha}_{G_o}, \{\sigma_j\}\big)|\widetilde{\mathcal{D}}'_{g_o} \widetilde{\mathcal{D}}_{g_o}|^{\frac{1}{2}}\Big)} \times n^{\frac{|G_o|-|G|}{2}} e^{\frac{1}{2}\left(|G_o|[\delta+\lambda_{\max}(\Gamma_n(0))]+|G|2\delta^{-1}\right)} \cdot \prod_{j=1}^{p} \left[\frac{(m_j^{g_o})^{\frac{n-|r_j^{g_o}|}{2}}}{(m_j^{g})^{\frac{n-|r_j^{g}|}{2}}} \right]$$

with probability exceeding $1 - 2V_2$, where V_2 is as in (10). Then by Theorem 3.10, for the fixed $K_3 \in (0, 1)$,

$$\frac{r(G \mid Y)}{r(G_o \mid Y)} \leq \Big(\frac{\pi}{n}\Big)^{\frac{|G|-|G_o|}{2}} \Big(\frac{n}{2}\Big)^{\frac{|G_o|}{2}+p} \frac{E\Big(h\big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big)\Big) e^{\frac{1}{2}(1-c)^{-2}\left(r_{\max}^{g}+(1+c)^2\right)\frac{\|Y\|^2}{\sqrt{n}} - \frac{|G|+p}{2}}}{(1-K_3) e^{\frac{|G_o|+p}{4}}} \times n^{\frac{|G_o|-|G|}{2}} e^{\frac{1}{2}\left(|G_o|[\delta+\lambda_{\max}(\Gamma_n(0))]+|G|2\delta^{-1}\right)} \cdot \prod_{j=1}^{p} \left[\frac{(m_j^{g_o})^{\frac{n-|r_j^{g_o}|}{2}}}{(m_j^{g})^{\frac{n-|r_j^{g}|}{2}}} \right]$$

with probability exceeding $1 - V_1 - \widetilde{V}_1 - 4V_2 - 2e^{-\frac{np}{4}} - V_3$. Gathering terms, for some positive constant K_2 (not depending on n nor p),

$$\frac{r(G \mid Y)}{r(G_o \mid Y)} \leq E\Big(h\big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big)\Big) e^{K_2 \cdot \left(\frac{p\|Y\|^2}{\sqrt{n}} + p^2 \log(n)\right)} \prod_{j=1}^{p} \left[\frac{(m_j^{g_o})^{\frac{n-|r_j^{g_o}|}{2}}}{(m_j^{g})^{\frac{n-|r_j^{g}|}{2}}} \right]$$

with probability exceeding $1 - V_1 - \widetilde{V}_1 - 4V_2 - 2e^{-\frac{np}{4}} - V_3$. At this point, $E\Big(h\big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big)\Big)$ can be bounded as in the following two cases.

Case 1: $G \subset G_o$ with $|G| \in \{1, \ldots, |G_o| - 1\}$. In this case, Theorem 3.9 does not apply, so since $h\big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big) \leq 1$ uniformly,

$$\frac{r(G \mid Y)}{r(G_o \mid Y)} \leq e^{K_2 \cdot \left(\frac{p\|Y\|^2}{\sqrt{n}} + p^2 \log(n)\right)} \prod_{j=1}^{p} \left[\frac{(m_j^{g_o})^{\frac{n - |r_j^{g_o}|}{2}}}{(m_j^{g})^{\frac{n - |r_j^{g}|}{2}}} \right], \tag{13}$$

with probability exceeding $1 - V_1 - \widetilde{V}_1 - 4V_2 - 2e^{-\frac{np}{4}} - V_3$. Then by Condition 3.11, $\frac{r(G|Y)}{r(G_o|Y)} \xrightarrow{P_x} 0$ as $n \to \infty$ or $n, p \to \infty$.

Case 2: $G \not\subseteq G_o$ and $|G| \in \{1, \ldots, p^2\}$. By Lemma S2.6, and for some positive constant K_1 (not depending on n nor p),

$$\begin{aligned} \frac{r(G \mid Y)}{r(G_o \mid Y)} &\leq E\Big(h\big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big)\Big) e^{K_2 \cdot \left(\frac{p\|Y\|^2}{\sqrt{n}} + p^2 \log(n)\right)} \cdot \big((\sigma^0_{\max})^2 3n\big)^{\frac{p^2}{2}} e^{(\sigma^0_{\max})^2 p^2 \sqrt{n} \frac{n}{2q}} \\ &\leq E\Big(h\big(\boldsymbol{\alpha}_G, \{\sigma_j\}\big)\Big) e^{K_1 \cdot \left(\frac{p\|Y\|^2}{\sqrt{n}} + p^2 \log(n)\right)} \cdot e^{K_1 \cdot \left(p^2 \log(n) + \frac{n}{q} p^2 \sqrt{n}\right)} \end{aligned}$$

with probability exceeding $1 - V_1 - \widetilde{V}_1 - 5V_2 - 3e^{-\frac{np}{4}} - V_3 - \frac{2(\sigma^0_{\max})^2}{\delta(1-c^2)\sqrt{n}}$. Therefore, by Theorem 3.9 and Condition 3.4,

$$\frac{r(G \mid Y)}{r(G_o \mid Y)} \leq \left(e^{-\frac{\varepsilon}{9\Lambda_g}} + e^{-\left(\frac{d \cdot n^{\frac{\rho}{2}} p^2}{4\lambda_{\max}(\mathcal{X}\mathcal{X}'/n)} - \frac{np}{2}\right)} 2^{-\frac{|G|}{2}+1} \right) e^{K_1 \cdot \left(\frac{p\|Y\|^2}{\sqrt{n}} + p^2 \log(n) + \frac{n}{q} p^2 \sqrt{n}\right)} \tag{14}$$

with probability exceeding $1 - 2V_1 - \widetilde{V}_1 - 5V_2 - 3e^{-\frac{np}{4}} - V_3 - \frac{2(\sigma^0_{\max})^2}{\delta(1-c^2)\sqrt{n}}$. Thus, by Condition 3.4, $\frac{r(G|Y)}{r(G_o|Y)} \xrightarrow{P_x} 0$ as $n \to \infty$ or $n, p \to \infty$. ■

Proof of Corollary 3.13. Omit case 1 in the proof of Theorem 3.12. ■

Proof of Corollary 3.14. Observe that

$$r(G_o \mid Y) = \frac{r(G_o \mid Y)}{\sum_{j=1}^{p^2} \sum_{G:|G|=j} r(G \mid Y)} = \frac{1}{1 + \sum_{j=1}^{p^2} \sum_{G \neq G_o : |G|=j} \frac{r(G|Y)}{r(G_o|Y)}}.$$

Since p is fixed Theorem 3.12 gives,

$$\sum_{j=1}^{p^2} \sum_{G \neq G_o : |G|=j} \frac{r(G \mid Y)}{r(G_o \mid Y)} \xrightarrow{P_x} 0$$

as $n \to \infty$, which proves the desired result. ■