國立臺灣大學理學院數學研究所

碩士論文

Department of Mathematics

College of Science

National Taiwan University

Master Thesis

向量自迴歸模型的費雪訊息與超調和先驗分布

Fisher Information and Superharmonic Priors for the Vector Autoregression Model

吳雋毅

Jiun-Yi Wu

指導教授: 楊鈞澔 博士

Advisor: Chun-Hao Yang, Ph.D.

中華民國 112 年 6 月

June, 2023

doi:10.6342/NTU202301697

doi:10.6342/NTU202301697

致謝

感謝我的指導教授楊鈞澔教授這兩年來悉心的指導，在我徬徨時給予我鼓勵及方向，協助我完成我的碩士論文研究。感謝我的父母支持我進行數學研究，讓我能走跟大部分人不一樣的路。感謝崔茂培教授從我大一以來的教導，在百忙之間，仍經常關心我的學業與未來方向。也謝謝我的女朋友鍾沛純，在我遇到困難時為我加油打氣，給予我求學前進的信心。最後感謝在天文數學館這六年來遇到的同學、學長姐、學弟妹們，陪我一起度過了在數學系的這些日子，充實了我的每一天。

doi:10.6342/NTU202301697

doi:10.6342/NTU202301697

摘要

向量自迴歸模型是一種常用在總體經濟及自然科學上的線性統計模型，本文聚焦在該模型的參數估計問題上。在缺乏對模型參數先驗資訊的情況下，無信息先驗分布是經常被考慮的選項，而傑佛瑞先驗分布是最常使用的分布之一。Komaki (1999) 提出了在滿足特定條件下，一種超調和先驗分布將存在並在參數估計上有比傑佛瑞先驗分布更好的表現，而Tanaka (2018) 成功的計算出了在一維的自迴歸模型上的超調和先驗分布。本篇論文沿襲了相同的思路，將該理論沿用在向量自迴歸模型上，並嘗試去計算出該模型的超調和先驗分布。本文依序介紹了定義超調和先驗分布所需要的背景知識，包含譜密度、費雪信息及信息幾何學，並給出了在向量自迴歸模型的模型流形上，明確的黎曼流形距離，為進一步的計算奠定了幾何基礎。最後我們提出了在自迴歸模型上，維度對流形距離的影響，並總結了計算該模型的超先驗分布時將遇到的困難及其中幾個可能的解決方法。

關鍵字：向量自迴歸模型、無信息先驗分布、超調和先驗分布、信息幾何學、費雪信息

doi:10.6342/NTU202301697

doi:10.6342/NTU202301697

Abstract

The vector autoregressive model (VAR) is a common choice when studying macroeconomics and natural science. In this thesis, we focus on estimating the parameters of the VAR model. When estimating without prior knowledge of the parameters, we often apply a non-informative prior, and Jeffreys prior is one of the most common choices. Komaki (1999) proposed that under certain conditions, a superharmonic prior exists and outperforms the estimation of the Jeffreys prior. Tanaka (2018) successfully derive the superharmonic prior for the autoregressive model. Our research applies this approach to the VAR model and aims to calculate the superharmonic prior for the VAR model. In the following thesis, we first introduce the necessary knowledge to define the superharmonic prior, including spectral density matrix, Fisher information, and information geometry. We compute the explicit form of the Riemannian metric of the VAR model manifold and establish the necessary geometry foundation for further computation. To conclude, we highlight the significant differences between the AR and VAR models, the obstacles when calculating the superharmonic prior, and some possible solutions.

Keywords: vector autoregression model, non-informative prior, superharmonic prior, information geometry, Fisher information matrix

doi:10.6342/NTU202301697

Contents

doi:10.6342/NTU202301697

doi:10.6342/NTU202301697

Chapter 1 Introduction

1.1 Bayesian Estimation for VAR model

Vector autoregression (VAR) is a linear multivariate statistical model to describe the joint relations between multiple time series. It provides a framework for data description, forecast, and decision-making. VARs have been applied to different fields, such as Medicine (Wild et al., 2010), Epidemiology (Langley et al., 2012), Economics (Stock and Watson, 2001), and Biology (Opgen-Rhein and Strimmer, 2007), to examine the dynamic relationships between variables that interact with one another. When applying VARs to Macroeconomic models, Sims (1980) proposed that a Bayesian approach could have improved upon previous frequentist ones in parameter estimation. Also, the number of parameters to be estimated in VARs increases significantly as time lags and dimension increase. Therefore, various priors for VARs have been studied and applied to different fields to overcome these difficulties, such as horseshoe priors (Prüser, 2021), shrinkage priors (Choi and Mullhaupt, 2015), Minnesota-prior (Kadiyala and Karlsson, 1997) and global-local priors (Cross et al., 2020), each with its advantages and disadvantages.

1.2 Spectral Density and Information Geometry

There are two main approaches for the analysis of physical signals: the time-domain approach and the frequency-domain approach. The time-domain analysis is based on the joint density of the observations. In contrast, the frequency-domain analysis is based on the Fourier transform of the joint density, i.e., the spectral density. Under certain assump-

doi:10.6342/NTU202301697

tions, the two approaches are equivalent. For example, under the stationarity assumption, the two approaches are equivalent for the VAR model (Klein, 2000). In this thesis, we follow the frequency-domain approach.

We adopt an information-theoretic point of view to perform statistical inference for spectral densities. More specifically, we consider the set of all spectral densities, which is a submanifold of the Euclidean space, and equip the submanifold with a Riemannian metric. This approach is called the Information Geometrical approach and will be described in more detail in Chapter 2. This approach allows us to derive priors using geometric notions. Jeffreys prior is one example: the volume measure on the submanifold. Another example is the superharmonic prior, which is the main focus of this thesis.

1.3 Superharmonic Prior for AR model

When selecting a prior for the VAR model, a non-informative prior is often considered if we only have vague or general information about a parameter. Jeffreys prior is one of the most common choices for this purpose. Two advantages of Jeffreys prior are (i) that it is invariant to reparametrization and (ii) it depends only on the model. However, Komaki (2006) showed that if there exists a positive superharmonic function on the model manifold of a parametric statistical model for i.i.d. random variables, the corresponding prior asymptotically dominates the Jeffreys prior when estimating the parameters. Tanaka and Komaki (2008) continued this idea and calculated the superharmonic prior of the autoregressive model with dimension two.

The pth order autoregression model, AR(p), is commonly used in univariate time-series analysis. It assumes the data $\{x_t\}$ satisfies

$$x_t = -\sum_{i=1}^{p} a_i x_{t-i} + \epsilon_t$$

where ϵ_t are i.i.d. Gaussian white noise with mean 0 and variance σ^2. The estimation of parameters $\{a_i\}$ is well studied from a frequentist point of view. However, the Bayesian approach for AR models remains challenging. Komaki (2006) has proved that a superhar-

doi:10.6342/NTU202301697

monic prior asymptotically dominates the Jefferys prior under Kullback-Leibler (KL) loss. Other studies (Brown, 1971) on the prediction of random Gaussian vectors also showed that superharmonic prior is a sufficient condition for the resulting estimator to be minimax. Last but not least, when studying a higher-dimensional time series, the large number of parameters is one of the obstacles ahead. We often need some sparse or shrinkage properties from the prior to reduce the number of significant covariates and better reveal the relations between each variable. In Tanaka's work, a superharmonic prior is biased toward parameters close to zero, so it is an ideal choice when analyzing an AR or VAR model.

Note that there are many non-informative priors besides Jeffreys prior and superharmonic priors, for example, the reference prior proposed by Berger and Bernardo (1992) and the probability matching priors proposed by Welch and Peers (1963). For a more detailed review of various non-informative priors, see Ghosh (2011) and the references therein. Our work follows Tanaka and Komaki (2008)'s approach toward the superharmonic priors of AR models and extends it to VARs.

doi:10.6342/NTU202301697

doi:10.6342/NTU202301697

Chapter 2 Theoretical Preliminary

In Section 2.1, we recall the basics of VAR models. We also introduce the lag operator and define the stationary condition for the VAR model. Throughout the rest of this thesis, we assume that all the models are stationary. In Section 2.2, we follow the frequency approach and introduce the spectral density of the general time series. The main goal then is to estimate the spectral density of the model, and we consider the Kullback-Leibler divergence when evaluating the performance. In Section 2.3, we review the general concept of information geometry and define the model manifold of the VAR model. In Section 2.4, we introduce the superharmonic function on a model manifold and the resulting superharmonic prior. Superharmonic priors have a shrinkage effect that can improve estimation or prediction. For example, Komaki (2006) showed that for Bayesian prediction, superharmonic priors asymptotically outperform the Jeffreys prior. Lastly, in Section 2.5, we review some results regarding the Fisher information matrix of the spectral density of VAR models.

2.1 Vector Autoregression Model of Order p

The vector autoregression model of order p is used to describe the discrete n-dimensional data $\{y_t\}$ and its dependence structure,

$$y_t = -C_1 y_{t-1} - C_2 y_{t-2} - \ldots - C_p y_{t-p} + \epsilon_t, \quad t \in \mathbb{N} \tag{2.1}$$

where p is the number of lags, $\{C_i : i = 1, \ldots, p\}$ are $n \times n$ real matrix parameters to be estimated. The error terms ϵ_t are independent and follow a multivariate Gaussian distribution with mean 0 and a positive semi-definite covariance matrix Λ. We will refer to the

doi:10.6342/NTU202301697

above model with p lags as a VAR(p) model for the rest of the article. For $n = 1$, we will refer to the model as an AR(p) model.

A common approach when analyzing regression models is to introduce the lag operator or back shift operator L, which operates on an element of the time series to produce the previous element, i.e. $Ly_t = y_{t-1}$ for all elements of the model. Then, a VAR(p) model can be rewritten as

$$y_t + C_1 L y_t + C_2 L^2 y_t + \ldots + C_p L^p y_t = A(L) y_t = \epsilon_t \ , \quad t \in \mathbb{N}$$

where

$$A(z) = I_n + C_1 z + C_2 z^2 + \ldots + C_p z^p$$

is the lag polynomial that follows similar rules as a regular polynomial. The use of this polynomial will be shown later in the thesis.

For the models in this thesis, we further assume that the model is stationary. A stochastic process is stationary if the unconditional joint probability distribution is invariant when time shifts. In such a case, the data will not have a particular trend and remain stable over the long term. For a VAR(p) model to be stationary, the zeros of the corresponding lag polynomial need to be outside the unit circle. That is

$$\forall \ |z| \leq 1, \quad \det\{A(z)\} \neq 0.$$

An important implication of this assumption is that $A(z)^{-1}$ exists, and can be written as a power series of z in this case.

A VAR(p) model can also be written as a VAR(1) model by stacking the vectors of p

doi:10.6342/NTU202301697

consecutive data, that is

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ \vdots \\ y_{t-p+1} \end{bmatrix} = - \begin{bmatrix} C_1 & C_2 & \cdots & C_{p-1} & C_p \\ -I_n & 0_n & \cdots & 0_n & 0_n \\ 0_n & -I_n & \cdots & 0_n & 0_n \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0_n & 0_n & \cdots & -I_n & 0_n \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ \vdots \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \vec{0} \\ \vdots \\ \vdots \\ \vec{0} \end{bmatrix}$$

$$\iff Y_t = -C\,Y_{t-1} + \hat{\epsilon}_t, \quad t \in \mathbb{N}$$

where $Y_t = \text{vec}\{y_t, y_{t-1}, \dots, y_{t-p+1}\}$, and $\hat{\epsilon}_t = \text{vec}\{\epsilon_t, \vec{0}, \dots, \vec{0}\}$. In this expression, the model is stationary if the eigenvalues of C are all inside the unit circle.

2.2 Spectral Density Matrix of Time Series

When analyzing a time series, the probability density function (pdf) of the joint distribution consists of infinitely many inputs. So it's useful to consider the spectral density matrix of the series rather than its pdf. The spectral density matrix of a stationary discrete-time series is defined as

$$S(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} r_{XX}(k) e^{-i\omega k}, \quad \omega \in [-\pi, \pi] \tag{2.2}$$

where $r_{XX}(k) = \mathbb{E}\left[X(t)X(t-k)\right]$ is the autocorrelation function of the time series $X(t)$. The function can also be interpreted as the Fourier transform of the autocorrelation function. Through this definition, the spectral density matrix $S(\omega \mid \theta)$ of the VAR model (2.1) is

$$S(\omega \mid \theta) = \left(\frac{1}{2\pi}\right) A^{-1}\left(\mathrm{e}^{\mathrm{i}\omega}\right) \Lambda A^{-\mathrm{T}}\left(\mathrm{e}^{-\mathrm{i}\omega}\right), \quad \omega \in [-\pi, \pi] \tag{2.3}$$

where $A(z) = I_n + C_1 z + C_2 z^2 + \ldots + C_p z^p$, and $\theta^\top = \text{vec}\left(C_1, C_2, \cdots, C_p\right)$ is a $n^2 p \times 1$ vector representing all the parameters of the model (Den Haan and Levin, 1998). It is known that a stationary VAR model has a one-to-one correspondence to its spectral

doi:10.6342/NTU202301697

density matrix (Whittle, 1963), so the estimation of a VAR model can be achieved by estimating its spectral density.

The performance of the estimate $\hat{S}(\omega)$ can be evaluated by the Kullback-Leibler divergence,

$$D\left(S\left(\omega \mid \theta\right) || \hat{S}(\omega)\right) := \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{4\pi} \left\{ \frac{S\left(\omega \mid \theta\right)}{\hat{S}(\omega)} - 1 - \log\left(\frac{S\left(\omega \mid \theta\right)}{\hat{S}(\omega)}\right) \right\} \tag{2.4}$$

where $S(\omega \mid \theta)$ is the true spectral density matrix. With a given prior $\pi(\theta)$, the average risk is then

$$\begin{aligned} &\mathbb{E}^{\Theta}\mathbb{E}^{X}[D(S(\omega \mid \theta) \| \hat{S}(\omega))] \\ &:= \int \mathrm{d}\theta \pi(\theta) \int p_n\left(x_1, \ldots, x_n \mid \theta\right) D(S(\omega \mid \theta) \| \hat{S}(\omega)) \mathrm{d}x_1 \cdots \mathrm{d}x_n. \end{aligned}$$

To minimize the average risk under the given prior $\pi(\theta)$, the Bayesian spectral density is then $S_\pi(\omega) := \int S(\omega \mid \theta)\pi(\theta \mid x)\mathrm{d}\theta$. Therefore, our main objective is to find a suitable prior for the VAR(p) model that outperforms the estimation of the usual noninformative prior, i.e. the Jeffreys prior.

2.3 Information Geometry and Model Manifold

Information geometry is a method of studying statistical models with modern geometry. The key observation is that the Fisher information matrix can be regarded as a Riemannian metric for a parametric family of statistical models. The modern theory of information geometry is formalized by Shun'ichi Amari. We will briefly review the basic construction of a statistical manifold and we refer the readers to Amari (2016) for more details and examples.

Let $\mathcal{M}$ be a family of statistical models indexed by the parameter set $\Theta \subseteq \mathbb{R}^d$, i.e. $\mathcal{M} = \{f_\theta \mid \theta \in \Theta\}$. The simplest example is the family of Gaussian distributions $\mathcal{M} = \{\mathcal{N}(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$. For $P \in \mathcal{M}$, we write θ_P as the parameter corresponding to P. In other words, θ_P can be viewed as the coordinates of P in $\mathcal{M}$. On such a family, we can define a *divergence*, which measures the dissimilarity between two models.

doi:10.6342/NTU202301697

Definition 2.3.1 (Amari (2016))**.** A function $D : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ is called a *divergence* if it satisfies

(i) $D(P\|Q) \geq 0$ for all $P, Q \in \mathcal{M}$,

(ii) $D(P\|Q) = 0$ if and only if $P = Q$, and

(iii) When P and Q are sufficiently close, by denoting their coordinates by θ_P and $\theta_Q = \theta_P + d\theta$, the Taylor expansion of D is written as

$$D(P\|Q) = \frac{1}{2} \sum g_{ij}(\theta_P) d\theta_i d\theta_j + O(|d\theta|^3),$$

and matrix $\mathbf{G} = (g_{ij})$ is positive-definite, depending on θ_P.

Note that divergence or its square root is not a distance since it need not be symmetric or satisfy the triangle inequality. The most common example of divergence is the Kullback-Leibler divergence (KL-divergence)

$$D_{KL}(p(x)\|q(x)) = \int p(x) \log \frac{p(x)}{q(x)} dx = \mathbb{E}_P \left[\log \frac{p(X)}{q(X)} \right]$$

where $p(x)$ and $q(x)$ are two probability density functions. Given a family $\mathcal{M}$ of models and a divergence $D(\cdot\|\cdot)$, various geometric structures can be induced on $\mathcal{M}$.

In this thesis, we mainly focus on the Riemannian metric of the model manifold. To find and define a suitable metric for the model manifold, we consider two distributions with different parameters, $p(x, \xi)$ and $p(x, \xi')$. The KL-divergence between these two distributions is then

$$\begin{aligned} D(p(x, \boldsymbol{\xi})\|p(x, \boldsymbol{\xi} + d\boldsymbol{\xi})) =& \frac{1}{2} \mathbb{E}_\xi \left[\partial_i \log p(x, \boldsymbol{\xi}) \partial_j \log p(x, \boldsymbol{\xi}) \right] d\xi^i d\xi^j \\ =& \frac{1}{2} F_{ij} d\xi^i d\xi^j. \end{aligned}$$

Here, F is the Fisher information matrix of the VAR(p) model. With this expression, it's natural to consider the Fisher information matrix to be the Riemannian metric of the model manifold.

doi:10.6342/NTU202301697

We know that any stationary VAR(p) model corresponds one to one to its spectral density matrix $S(\omega|\theta)$. So we substitute the pdf of the VAR(p) model with its spectral density matrix. Then, the spectral density matrix of the parametric family forms a manifold

$$\mathcal{M} := \{S(\omega \mid \theta) : \theta \in \Theta\}$$

This manifold is referred to as the model manifold of VAR(p). As mentioned earlier, we naturally choose the Fisher information matrix to be the Riemannian metric on the model manifold.

With the concept of model manifold, any prior of the model is essentially a probability distribution function of the parameters $\theta^\top = \text{vec}\,(C_1, C_2, \cdots, C_p)$, which is a positive function defined on the model manifold. For example, the Jeffreys prior, $\pi_J(\theta) \propto \sqrt{|\det(F(\theta))|}$, corresponds to the volume element of the manifold. With the above background knowledge, we are prepared to define the superharmonic prior of the VAR(p) model.

2.4 Superharmonic Prior

In the Bayesian framework, if the prior knowledge is vague, it is common to choose a non-informative prior. Jeffreys prior is one of the most common choices in most scenarios since it is invariant under reparametrization. However, Komaki (2006) gave sufficient conditions for the existence of shrinkage predictive distributions that asymptotically dominate the Jeffreys predictive distribution. More specifically, if we can obtain a positive superharmonic function on the model manifold, we can then define the corresponding superharmonic prior, which outperforms the Jeffreys prior under the Kullback-Leibler divergence (Komaki, 2006).

A superharmonic prior of a model is defined through the respective model manifold. For a model manifold $\mathcal{M} := \{S(\omega \mid \theta) \mid \theta \in \Theta\}$ with a Riemannian metric $F(\theta)$, the

doi:10.6342/NTU202301697

Laplace-Beltrami operator on the model manifold is,

$$\Delta\phi := \frac{1}{\sqrt{F}}\frac{\partial}{\partial\theta^i}\left(\sqrt{F}F^{ij}\frac{\partial}{\partial\theta^j}\phi\right) \tag{2.5}$$

where ϕ is any scalar function defined on the manifold, and F^{ij} is the (i, j) entry of F^{-1}. Note that we adopt Einstein's summation convention in the above equation. With the Laplace-Beltrami operator, any scalar function ϕ defined on the model manifold is called a superharmonic function if $\Delta\phi \leq 0$ for all θ. For any given positive superharmonic function ϕ, the corresponding superharmonic prior can then be defined as $\pi_H(\theta) := \pi_J(\theta)\phi(\theta)$, where $\pi_J(\theta) \propto \sqrt{|\det(F(\theta))|}$ is the Jeffreys prior of the VAR model.

With this definition, we then examine the existence of the superharmonic function. Komaki (2006) proved that for a complete simply connected model manifold endowed with the Fisher metric, superharmonic functions exist if the model manifold has strictly negative curvature ($d = 2$) or has negative curvature ($d \geq 3$). Here, the curvature is considered negative if the sectional curvature is negative for all tangent planes at any point. This theorem provides us with a sufficient condition for the existence of superharmonic functions.

It is natural to apply the above sufficient condition to verify whether a superharmonic function exists on VAR(p) models. However, Tanaka (2003) showed that the sectional curvature of the autoregressive model manifold (lag ≥ 3) is strictly positive for some planes and at some point. Then, the above sufficient condition could not be applied to the autoregressive model. Since the autoregressive model is a sub-model of the VAR(p) model, the sectional curvature of the VAR(p) model manifold is also unlikely to satisfy the sufficient condition. The existence of superharmonic functions will have to rely on other conditions or through our construction.

Tanaka (2018) concluded an all-around result of the autoregressive model and one of the superharmonic prior of the model. In Tanaka (2018), the Fisher information metric and the model manifold were established. Also, one of the superharmonic functions of the autoregressive model is obtained through computation. The result is significant but hard to be replicated and applied to the VAR(p) model. However, the result for the autoregressive

doi:10.6342/NTU202301697

model gives a hint for predicting and computing the case of the VAR(p) model.

2.5 The Fisher Information Matrix of VAR(p)

To obtain the Riemannian metric of the model manifold, we need to compute the Fisher information matrix of the VAR(p) model. Generally, the Fisher information matrix $F(\theta)$ is defined as

$$[F(\theta)]_{i,j} = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta_i}\log f(X;\theta)\right)\left(\frac{\partial}{\partial\theta_j}\log f(X;\theta)\right) \mid \theta\right]$$

where $f(x;\theta)$ is the probability density function. Here, we illustrate two different approaches to calculating the Fisher information matrix of the VAR(p) model.

Using the lag polynomial, we can rewrite the model as

$$\epsilon_t = A(L)y_t \ , \quad t \in \mathcal{N}.$$

Here, ϵ_t are identically independent Gaussian distributions with zero mean and covariance matrix Λ. The corresponding Fisher information matrix of the Gaussian distribution is given by

$$F(\theta) = \mathbb{E}\left\{\frac{\partial\varepsilon^*}{\partial\theta^{\mathrm{T}}}\Lambda^{-1}\frac{\partial\varepsilon}{\partial\theta^{\mathrm{T}}}\right\}. \tag{2.6}$$

With this expression, the Fisher information of the VAR(p) model is

$$F(\theta) = \frac{1}{2\pi\mathrm{i}}\oint_{|z|=1}\left(\frac{\partial\,\mathrm{vec}\,A\left(z^{-1}\right)}{\partial\theta^{\mathrm{T}}}\right)^{\mathrm{T}}\left(A^{-1}\left(z^{-1}\right)\Lambda\,A^{-\mathrm{T}}(z)\otimes\Lambda^{-1}\right)\left(\frac{\partial\,\mathrm{vec}\,A(z)}{\partial\theta^{\mathrm{T}}}\right)\frac{\mathrm{d}z}{z}. \tag{2.7}$$

Since a stationary VAR(p) model corresponds one-to-one to its spectral density matrix, another approach to calculating the Fisher information matrix is through its spectral density matrix. The Fisher information matrix of a spectral density matrix is given by

$$F_{ij}(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\mathrm{Tr}\left(\frac{\partial S(\omega\mid\theta)}{\partial\theta_i}S^{-1}(\omega\mid\theta)\frac{\partial S(\omega\mid\theta)}{\partial\theta_j}S^{-1}(\omega\mid\theta)\right)\mathrm{d}\omega.$$

doi:10.6342/NTU202301697

This result can also be rewritten as

$$F(\theta) = \frac{1}{4\pi \mathrm{i}} \oint_{|z|=1} \left(\frac{\partial \operatorname{vec} S\left(z^{-1}\right)}{\partial \theta^{\mathrm{T}}} \right)^{\mathrm{T}} \left(S^{\mathrm{T}}(z) \otimes S(z) \right)^{-1} \left(\frac{\partial \operatorname{vec} S(z)}{\partial \theta^{\mathrm{T}}} \right) \frac{\mathrm{d}z}{z}. \tag{2.8}$$

Klein (2000) has shown that the two methods above agree on the VARMA model, which consists of the VAR(p) model. We will use the two formulas for different cases in this thesis.

doi:10.6342/NTU202301697

doi:10.6342/NTU202301697

Chapter 3 Main Results

In Section 3.1, we give a closed-form expression for the Fisher information matrix for the VAR(1) model. This result is not only a particular case but a helpful result, because all the autoregression models, including AR(p) and VAR(p), can be reduced to a VAR(1) model. The closed-form expression is also crucial for calculating the Laplace-Beltrami operator on the model manifold. Without the closed-form expression, the determinant and inverse of the Fisher information matrix in the operator can not be obtained.

In Section 3.2, we show how the Fisher information matrix of the VAR(p) model can be derived from the case of VAR(1). To further simplify the result, Komaki (1999) applied a change of coordinate when studying the AR(p) model. In Section 3.3, we verify that the AR(p) model can be derived from our result and our calculation matches with the previous work from Komaki (1999).

In Section 3.4, we consider the covariance matrix of the white noise, Λ, as one of the unknown parameters to be estimated. Then, the model manifold would consist of additional n^2 parameters. We compare the AR(p) and VAR(p) models and verify that the Fisher information matrix in both cases follows the same formula.

Finally, in Section 3.5, we give an example of a VAR(1) model with $n = 2$ and calculate the results under different assumptions. These examples show that the multi-dimensional case model is fundamentally more complex and different from the one-dimensional case.

doi:10.6342/NTU202301697

3.1 Fisher Matrix for VAR(1) model

To begin with, we first calculate the Fisher information matrix of the VAR(1) model,

$$y_t = -Cy_{t-1} + \epsilon_t, \quad t \in \mathbb{N}.$$

We follow equation 2.7 which is

$$F(\theta) = \frac{1}{2\pi \mathrm{i}} \oint_{|z|=1} \left(\frac{\partial \operatorname{vec} A\left(z^{-1}\right)}{\partial \theta^{\mathrm{T}}} \right)^{\mathrm{T}} \left(A^{-1}\left(z^{-1}\right) \Lambda A^{-\mathrm{T}}(z) \otimes \Lambda^{-1} \right) \left(\frac{\partial \operatorname{vec} A(z)}{\partial \theta^{\mathrm{T}}} \right) \frac{\mathrm{d}z}{z}.$$

Here, $A(z) = I_n + Cz$ is the lag polynomial of the VAR(1) model. Since we assume the VAR models are stationary, the roots of $\det(A(z)) = \det(I_n + Cz) = 0$ are outside the unit disc. So the elements of $A^{-1}(z)$ can be written as power series in z for $|z| \leq 1$. On the other hand, the roots of $\det A(z^{-1}) = 0$ are inside the unit circle, which suggests that $A^{-1}(z^{-1})$ exists and can be written as power series in z for $|z| \geq 1$. Therefore, both $A^{-1}(z)$ and $A^{-1}(z^{-1})$ can be written as a power series in z on the unit circle $|z| = 1$ (Higham, 2008). So on the unit circle $|z| = 1$, we have the two expansions

$$A^{-1}(z) = (I_n + Cz)^{-1} = \sum_{k=0}^{\infty} \left(-Cz^{-1}\right)^k$$

$$A^{-1}(z^{-1}) = (I_n + Cz^{-1})^{-1} = \sum_{k=0}^{\infty} (-Cz)^k .$$

Now we can start simplifying the formula. By arranging the parameters as $\theta^{\top} = \operatorname{vec}(C)$, the partial derivatives become apparent, since

$$\operatorname{vec} A(z) = \operatorname{vec} I_n + \operatorname{vec}(C) z, \quad \frac{\partial \operatorname{vec} A(z)}{\partial \theta^{\mathrm{T}}} = I_n z$$

and

$$\operatorname{vec} A\left(z^{-1}\right) = \operatorname{vec} I_n + \operatorname{vec}(C) z^{-1}, \quad \frac{\partial \operatorname{vec} A\left(z^{-1}\right)}{\partial \theta^{\mathrm{T}}} = I_n z^{-1}.$$

doi:10.6342/NTU202301697

Therefore,

$$
\begin{aligned}
F(\theta) &= \frac{1}{2\pi i}\oint_{|z|=1} [A^{-1}\left(z^{-1}\right)\Lambda A^{-\mathrm{T}}(z)] \otimes \Lambda^{-1}\frac{dz}{z} \\
&= \frac{1}{2\pi i}\oint_{|z|=1} \left(I_n + Cz^{-1}\right)^{-1}\Lambda\left(I_n + C^{\mathrm{T}}z\right)^{-1}\frac{dz}{z}\otimes\Lambda^{-1} \\
&= \frac{1}{2\pi i}\oint_{|z|=1} \sum_{i=0}^{\infty}\left(-Cz^{-1}\right)^{i}\Lambda\sum_{j=0}^{\infty}\left(-C^{\mathrm{T}}z\right)^{j}\frac{dz}{z}\otimes\Lambda^{-1} \\
&= \sum_{k=0}^{\infty}(-C)^{k}\Lambda\left(-C^{\mathrm{T}}\right)^{k}\otimes\Lambda^{-1} \\
&= f(\theta)\otimes\Lambda^{-1}.
\end{aligned}
$$

Here, $f(\theta)=\sum_{k=0}^{\infty}(-C)^{k}\Lambda(-C^{\mathrm{T}})^{k}=\sum_{k=0}^{\infty}(C)^{k}\Lambda(C^{\mathrm{T}})^{k}$.

Note that,

$$C(\sum_{k=0}^{N}(C)^{k}\Lambda\left(C^{\mathrm{T}}\right)^{k})C^{\mathrm{T}}-\sum_{k=0}^{N}(C)^{k}\Lambda\left(C^{\mathrm{T}}\right)^{k}=C^{N+1}\Lambda(C^{\mathrm{T}})^{N+1}-\Lambda.$$

The stationarity of the model ensures that $\lim_{N\to\infty}C^{N}=0_n$. So $f(\theta)$ satisfies the following equation

$$Cf(\theta)C^{\mathrm{T}}-f(\theta)+\Lambda=0$$

which is the Lyapunov equation (Hammarling, 1982). The solution can be expressed in matrix form via the vectorization operator

$$
\begin{aligned}
&(I_{n^2}-C\otimes C)\,\mathrm{vec}(f(\theta))=\mathrm{vec}(\Lambda) \\
&\Rightarrow \mathrm{vec}(f(\theta))=(I_{n^2}-C\otimes C)^{-1}\,\mathrm{vec}(\Lambda).
\end{aligned}
\tag{3.1}
$$

To omit the vectorization operator, we can apply the inverse vectorize operator. For a $n^2\times 1$ vector $\vec{x}$,

$$\mathrm{vec}^{-1}(\vec{x})=\left((\mathrm{vec}\,I_n)^{\mathrm{T}}\otimes I_n\right)(I_n\otimes\vec{x}).$$

With the above calculations and expression, we have computed the explicit form of the Fisher information matrix of the VAR(1) model.

doi:10.6342/NTU202301697

Theorem 3.1.1. For a stationary VAR(1) model with C as its parameter

$$y_t = -Cy_{t-1} + \epsilon_t, \quad t \in \mathbb{N},$$

the corresponding Fisher information matrix is

$$\begin{aligned} F(\theta) &= f(\theta) \otimes \Lambda^{-1} \\ &= \left\{ \left[(\mathrm{vec}\, I_n)^{\mathrm{T}} \otimes I_n \right] \left[I_n \otimes ((I_{n^2} - \mathrm{C} \otimes \mathrm{C})^{-1} \mathrm{vec}(\Lambda)) \right] \right\} \otimes \Lambda^{-1}. \end{aligned}$$

3.2 VAR(p) model as a VAR(1) model

To tackle the Fisher information matrix of VAR(p) models, we rewrite the recursive equation $y_t = -C_1 y_{t-1} - C_2 y_{t-2} - \ldots - C_p y_{t-p} + \epsilon_t$ as

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ \vdots \\ y_{t-p+1} \end{bmatrix} = - \begin{bmatrix} C_1 & C_2 & \cdots & C_{p-1} & C_p \\ -I_n & 0 & & 0 & 0 \\ 0 & -I_n & & 0 & \vdots \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & & -I_n & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ \vdots \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}. \tag{3.2}$$

Then, it's obvious that this is a VAR(1) model

$$Y_t = -C\, Y_{t-1} + \hat{\epsilon}_t, \text{ with}$$

$$\hat{\epsilon}_t = e_p \otimes \epsilon_t, \;\; e_p = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix}.$$

Here, Y_t is a $np \times 1$ vector, and C is a $np \times np$ matrix with only $n^2 p$ parameters to be estimated.

However, in this VAR(1) model, the covariance matrix of $\hat{\epsilon}_t$ is of the form $(e_p e_p^{\mathrm{T}}) \otimes \Lambda$ which is not of full rank. Essentially, the model (3.2) is a degenerate case of the general VAR(1) model because only the first np rows of the data contain random variables. As a

doi:10.6342/NTU202301697

result, the spectral density matrix

$$S(\omega) = \left(\frac{1}{2\pi}\right) A^{-1}\left(\mathrm{e}^{\mathrm{i}\omega}\right)\left((e_p e_p^{\mathrm{T}}) \otimes \Lambda\right) A^{-\mathrm{T}}\left(\mathrm{e}^{-\mathrm{i}\omega}\right), \quad \omega \in [-\pi, \pi]$$

is not Invertible. To tackle this problem, we need to consider the pseudo inverse of the spectral density matrix. That is to consider the pseudo inverse of $(e_p e_p^{\mathrm{T}}) \otimes \Lambda$ to be $(e_p e_p^{\mathrm{T}}) \otimes \Lambda^{-1}$.

Now, we can follow the same procedure as the VAR(1) model to obtain the result for VAR(p) models. The parameters will be arrange as $\theta^{\mathrm{T}} = \operatorname{vec}(C_1, C_2, \cdots, C_p)$, $A(z) = I_{np} + Cz$, with

$$\frac{\partial \operatorname{vec} A(z)}{\partial \theta^{\mathrm{T}}} = z I_{np} \otimes e_p \otimes I_n, \quad \left(\frac{\partial \operatorname{vec} A\left(z^{-1}\right)}{\partial \theta^{\mathrm{T}}}\right)^{\mathrm{T}} = z^{-1} I_{np} \otimes e_p^{\mathrm{T}} \otimes I_n.$$

So the Fisher information matrix of the VAR(p) model is as follows

$$\begin{aligned} F(\theta) = \frac{1}{2\pi i} \oint_{|z|=1.} &\left(I_{np} \otimes e_p^{\mathrm{T}} \otimes I_n\right) \Big\{ \left[\left(I_{np} + Cz^{-1}\right)^{-1}\left((e_p e_p^{\mathrm{T}}) \otimes \Lambda\right)\left(I_{np} + C^{\mathrm{T}} z\right)^{-1}\right] \\ &\otimes \left[(e_p e_p^{\mathrm{T}}) \otimes \Lambda^{-1}\right] \Big\} \left(I_{np} \otimes e_p \otimes I_n\right) \frac{dz}{z} \\ = \left(I_{np} \otimes e_p^{\mathrm{T}} \otimes I_n\right) &\frac{1}{2\pi i} \oint_{|z|=1.} \Big\{ \left[\left(I_{np} + Cz^{-1}\right)^{-1}\left((e_p e_p^{\mathrm{T}}) \otimes \Lambda\right)\left(I_{np} + C^{\mathrm{T}} z\right)^{-1}\right] \\ &\otimes \left[(e_p e_p^{\mathrm{T}}) \otimes \Lambda\right] \Big\} \frac{dz}{z} \left(I_{np} \otimes e_p \otimes I_n\right) \\ = \left(I_{np} \otimes e_p^{\mathrm{T}} \otimes I_n\right) &\left[g(\theta) \otimes (e_p e_p^{\mathrm{T}} \otimes \Lambda^{-1})\right]\left(I_{np} \otimes e_p \otimes I_n\right), \end{aligned}$$

where

$$\begin{aligned} g(\theta) &= \frac{1}{2\pi i} \oint_{|z|=1.} \left(I_{np} + Cz^{-1}\right)^{-1}\left((e_p e_p^{\mathrm{T}}) \otimes \Lambda\right)\left(I_{np} + C^{\mathrm{T}} z\right)^{-1} \frac{dz}{z} \\ &= \frac{1}{2\pi i} \oint_{|z|=1} \sum_{i=0}^{\infty}\left(-Cz^{-1}\right)^i \left((e_p e_p^{\mathrm{T}}) \otimes \Lambda\right) \sum_{j=0}^{\infty}\left(-C^{\mathrm{T}} z\right)^j \frac{dz}{z} \\ &= \sum_{k=0}^{\infty}(-C)^k \left((e_p e_p^{\mathrm{T}}) \otimes \Lambda\right)\left(-C^{\mathrm{T}}\right)^k. \end{aligned}$$

doi:10.6342/NTU202301697

Similar to Section 3.1, $g(\theta)$ satisfies the Lyapunov equation

$$
\begin{aligned}
&Cg(\theta)C^{\mathrm{T}} - g(\theta) + (e_p e_p^{\mathrm{T}}) \otimes \Lambda = 0 \\
\Rightarrow &\operatorname{vec}(g(\theta)) = (I_{n^2p^2} - C \otimes C)^{-1} \operatorname{vec}((e_p e_p^{\mathrm{T}}) \otimes \Lambda)) \\
\Rightarrow &g(\theta) = \left[(\operatorname{vec} I_{np})^{\mathrm{T}} \otimes I_{np}\right] \left[I_{np} \otimes \left(I_{n^2p^2} - C \otimes C\right)^{-1} \operatorname{vec}\left((e_p e_p^{\mathrm{T}}) \otimes \Lambda\right)\right]
\end{aligned}
$$

Plug in $g(\theta)$ into the previous equation, and we have the Fisher information matrix of the VAR(p) model.

Theorem 3.2.1. For a stationary VAR(p) model with $C_1 \dots C_p$ as its parameter

$$
y_t = -C_1 y_{t-1} - C_2 y_{t-2} - \dots - C_p y_{t-p} + \epsilon_t, \quad t \in \mathbb{N},
$$

the corresponding Fisher information matrix is

$$
\begin{aligned}
F(\theta) &= f(\theta) \otimes \Lambda^{-1} \\
\operatorname{vec} f(\theta) &= (I_{n^2p^2} - \mathrm{C} \otimes \mathrm{C})^{-1} \operatorname{vec}((e_p e_p^{\mathrm{T}}) \otimes \Lambda)
\end{aligned}
$$

where

$$
C = \begin{bmatrix}
C_1 & C_2 & \cdots & C_{p-1} & C_p \\
-I_n & 0 & & 0 & 0 \\
0 & -I_n & & 0 & \vdots \\
\vdots & \vdots & & \vdots & \vdots \\
0 & 0 & & -I_n & 0
\end{bmatrix}.
$$

3.3 Relation with AR(p) Model under Change of Coordinate

The main obstacle ahead is that when determining the superharmonicity of a function ϕ, we need to apply the Laplace-Beltrami operator (2.5), and the computation becomes significantly complicated. When computing the Fisher information matrix of the autoregressive model, Komaki (1999) presented an approach to apply a change of coordinate on

doi:10.6342/NTU202301697

the AR(p) parameters, and the results were very promising.

An AR(p) model has exactly p parameters to be estimated and is of the form

$$y_t = -a_1 y_{t-1} - a_2 y_{t-2} - \ldots - a_p y_{t-p} + \epsilon_t, \quad t \in \mathbb{N}.$$

Komaki (1999) considered the alternative coordinate $Z = (z_1, z_2, \cdots z_p)^\top$, which are the complex roots of the characteristic polynomial

$$\begin{aligned} H(z) &= z^p A(z^{-1}) \\ &= z^p + a_1 z^{p-1} + a_2 z^{p-2} + \ldots + a_p. \end{aligned}$$

Then, the Fisher information matrix of the AR(p) model is of the form

$$F(\theta)_{ij} = \frac{1}{1 - z_i z_j}. \tag{3.3}$$

With this expression, a superharmonic function of the AR(p) model can be obtained more conveniently. Tanaka (2018) successfully computed the superharmonic function of the general AR(p) model. From our perspective, an AR(p) model can also be rewritten as a VAR(1) model, so we apply our result with the change of coordinate to verify that our approach is indeed correct,

The autoregression model in this case is

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ \vdots \\ y_{t-p+1} \end{bmatrix} = - \begin{bmatrix} a_1 & a_2 & \cdots & a_{p-1} & a_p \\ -1 & 0 & & 0 & 0 \\ 0 & -1 & & 0 & \vdots \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & & -1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ \vdots \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}.$$

doi:10.6342/NTU202301697

The parameters $A = (a_1, a_2, \cdots a_p)^{\mathrm{T}}$ form a companion matrix

$$C = \begin{bmatrix} a_1 & a_2 & \cdots & a_{p-1} & a_p \\ -1 & 0 & & 0 & 0 \\ 0 & -1 & & 0 & \vdots \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & & -1 & 0 \end{bmatrix}.$$

Applying our result of the VAR(p) model (3.2.1), the corresponding Fisher information matrix in vectorized form is

$$\begin{aligned} &\mathrm{vec}\, F(\theta) = (I_{p^2} - C \otimes C)^{-1} \mathrm{vec}\,(e_p) \\ &\Rightarrow F(\theta) = \left((\mathrm{vec}\, I_p)^{\mathrm{T}} \otimes I_p \right) \left(I_p \otimes \left[(I_{p^2} - C \otimes C)^{-1} \mathrm{vec}\,(e_p) \right] \right). \end{aligned}$$

Note that if we apply the eigenvalue decomposition on C, the eigenvalues are exactly the roots of the characteristic polynomial (Chen and Louck, 1996), i.e

$$C = \begin{bmatrix} a_1 & a_2 & \cdots & a_{p-1} & a_p \\ -1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & \cdots & \vdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 0 \end{bmatrix} = QDQ^{-1}, \quad D = \mathrm{diag}\,(z_1, z_2, \cdots, z_p).$$

Since the Riemannian metric is a local property on the tangent plane, we can use the Jacobian matrix $\frac{\partial A}{\partial Z}$ to obtain the Riemannian metric after the change of coordinate.

Here, let $\tilde{F}_{z_i z_j}$ denotes the (i, j) entry of the Riemannian metric under the Z coordinate, and $F_{a_k a_l}$ denotes the (k, l) entry of the Riemannian metric under the A coordinate. Then, the two metric satisfies the equation

$$\tilde{F}_{z_i z_j} = \sum_{k=1}^{p} \sum_{l=1}^{p} \frac{\partial a_k}{\partial z_i} \frac{\partial a_l}{\partial z_j} F_{a_k a_l}.$$

doi:10.6342/NTU202301697

Consequently,

$$\widetilde{F} = \left(\frac{\partial A}{\partial Z}\right)^{\mathrm{T}} F \left(\frac{\partial A}{\partial Z}\right), \quad \operatorname{vec}(\tilde{F}) = \left(\left(\frac{\partial A}{\partial Z}\right)^{\mathrm{T}} \otimes \left(\frac{\partial A}{\partial Z}\right)^{\mathrm{T}}\right) \operatorname{vec}(F).$$

Using MATLAB, we verify our result indeed coincides with Komaki's result (3.3). However, direct proof of equality is still required. Only with more understanding of the choice of the change of variable can we apply it to the VAR(p) model.

3.4 VAR(p) Model with Unknown Noise Covariance

In Section 3.1 and Section 3.2, the Gaussian white noise covariance Λ is considered a known constant. However, this assumption is often impractical in most situations. For most scenarios, Λ is one of the unknown parameters to be estimated. In this section, we will view Λ as one of the unknown parameters to be estimated. As a result, the model manifold will now consist of a total of $n^2(p+1)$ coordinates. Namely $\theta^{\mathrm{T}} = \operatorname{vec}(\Lambda, C_1, C_2, \cdots, C_p)$. Since Λ is a real symmetric matrix. it actually only contains $n(n+1)/2$ free variables. However, we will still consider the model manifold with $n^2(p+1)$ parameters with $n(n-1)/2$ coordinates being extra copies.

The spectral density of the model with unknown Λ is

$$S(\omega) = \left(\frac{1}{2\pi}\right) A^{-1}\left(e^{i\omega}\right) \Lambda A^{-\mathrm{T}}\left(e^{-i\omega}\right), \quad \omega \in [-\pi, \pi].$$

To calculate the Fisher information matrix, we follow the formula (2.8)

$$F(\theta) = \frac{1}{4\pi i} \oint_{|z|=1} \left(\frac{\partial \operatorname{vec} S\left(z^{-1}\right)}{\partial \theta^{\mathrm{T}}}\right)^{\mathrm{T}} \left(S^{\mathrm{T}}(z) \otimes S(z)\right)^{-1} \left(\frac{\partial \operatorname{vec} S(z)}{\partial \theta^{\mathrm{T}}}\right) \frac{dz}{z}.$$

We first calculate the differential of the spectral density by following the differential rule

doi:10.6342/NTU202301697

$\mathrm{d}A^{-1} = -A^{-1}(\mathrm{d}A)A^{-1}$, and note that $\mathrm{d}\Lambda \neq 0$ in this context. Then,

$$
\begin{aligned}
S(z) &= \left(\frac{1}{2\pi}\right) A^{-1}(z)\Lambda A^{-\mathrm{T}}\left(z^{-1}\right) \\
2\pi\,\mathrm{d}S(z) &= -\,A^{-1}(z)[\,\mathrm{d}A(z)]A^{-1}(z)\Lambda A^{-\mathrm{T}}\left(z^{-1}\right) \\
&\quad + A^{-1}(z)\,\mathrm{d}\Lambda A^{-\mathrm{T}}\left(z^{-1}\right) \\
&\quad - A^{-1}(z)\Lambda A^{-\mathrm{T}}\left(z^{-1}\right)[\,\mathrm{d}A^{\mathrm{T}}\left(z^{-1}\right)]A^{-\mathrm{T}}\left(z^{-1}\right).
\end{aligned}
$$

Following the vectorization rule $\operatorname{vec}(ABC) = (C^{\mathrm{T}} \otimes A)\operatorname{vec}(B)$, we obtain

$$
\begin{aligned}
(2\pi)\frac{\partial \operatorname{vec} S(z)}{\partial \theta^{\mathrm{T}}} &= -\left[\left(A^{-1}\left(z^{-1}\right)\Lambda A^{-\mathrm{T}}(z)\right) \otimes A^{-1}(z)\right]\frac{\partial \operatorname{vec} A(z)}{\partial \theta^{\mathrm{T}}} \\
&\quad + \left[A^{-1}\left(z^{-1}\right) \otimes A^{-1}(z)\right]\frac{\partial \operatorname{vec} \Lambda}{\partial \theta^{\mathrm{T}}} \\
&\quad - \left[A^{-1}\left(z^{-1}\right) \otimes \left(A^{-1}(z)\Lambda A^{-\mathrm{T}}(z^{-1})\right)\right]\frac{\partial \operatorname{vec} A^{\mathrm{T}}\left(z^{-1}\right)}{\partial \theta^{\mathrm{T}}} \\
&= 2\pi(P(z) + Q(z) + R(z)),
\end{aligned}
$$

where

$$
\begin{aligned}
P(z) &= -\frac{1}{2\pi}\left[\left(A^{-1}\left(z^{-1}\right)\Lambda A^{-\mathrm{T}}(z)\right) \otimes A^{-1}(z)\right]\frac{\partial \operatorname{vec} A(z)}{\partial \theta^{\mathrm{T}}} \\
Q(z) &= \frac{1}{2\pi}\left[A^{-1}\left(z^{-1}\right) \otimes A^{-1}(z)\right]\frac{\partial \operatorname{vec} \Lambda}{\partial \theta^{\mathrm{T}}} \\
R(z) &= -\frac{1}{2\pi}\left[A^{-1}\left(z^{-1}\right) \otimes \left(A^{-1}(z)\Lambda A^{-\mathrm{T}}(z^{-1})\right)\right]\frac{\partial \operatorname{vec} A^{\mathrm{T}}\left(z^{-1}\right)}{\partial \theta^{\mathrm{T}}}.
\end{aligned}
$$

The Fisher information matrix $F_{\Lambda,C}(\theta)$ with unknown Λ now consists of 4 parts, i.e.

$$
\begin{aligned}
F_{\Lambda,C}(\theta) =& \frac{1}{4\pi i}\oint_{|z|=1}\left[P\left(z^{-1}\right) + Q\left(z^{-1}\right) + R\left(z^{-1}\right)\right]^{\mathrm{T}}\left(S^{\mathrm{T}}(z) \otimes S(z)\right)^{-1}[P(z) + Q(z) + R(z)]\frac{dz}{z} \\
=& \frac{1}{4\pi i}\oint_{|z|=1}\left[P\left(z^{-1}\right) + R\left(z^{-1}\right)\right]^{\mathrm{T}}\left(S^{\mathrm{T}}(z) \otimes S(z)\right)^{-1}[P(z) + R(z)]\frac{dz}{z} \\
&+ \frac{1}{4\pi i}\oint_{|z|=1} Q\left(z^{-1}\right)^{\mathrm{T}}\left(S^{\mathrm{T}}(z) \otimes S(z)\right)^{-1} Q(z)\frac{dz}{z} \\
&+ \frac{1}{4\pi i}\oint_{|z|=1}\left[P\left(z^{-1}\right) + R\left(z^{-1}\right)\right]^{\mathrm{T}}\left(S^{\mathrm{T}}(z) \otimes S(z)\right)^{-1} Q(z)\frac{dz}{z} \\
&+ \frac{1}{4\pi i}\oint_{|z|=1} Q\left(z^{-1}\right)^{\mathrm{T}}\left(S^{\mathrm{T}}(z) \otimes S(z)\right)^{-1}[P(z) + R(z)]\frac{dz}{z}.
\end{aligned}
$$

doi:10.6342/NTU202301697

The first line of the equation corresponds to the result in Section 3.2 where Λ is considered a known constant. The second line can be further simplified as

$$
\begin{aligned}
&\frac{1}{4\pi i}\oint_{|z|=1} Q\left(z^{-1}\right)^{\mathrm{T}}\left(S^{\mathrm{T}}(z)\otimes S(z)\right)^{-1} Q(z)\frac{dz}{z}\\
=&\frac{1}{4\pi i}\oint_{|z|=1}\left(\frac{\partial \operatorname{vec}\Lambda}{\partial\theta^{\mathrm{T}}}\right)^{\mathrm{T}}\left[A^{-\mathrm{T}}(z)\otimes A^{-\mathrm{T}}\left(z^{-1}\right)\right]\left[\left(A^{-1}\left(z^{-1}\right)\Lambda A^{-\mathrm{T}}(z)\right)\otimes\left(A^{-1}(z)\Lambda A^{-\mathrm{T}}\left(z^{-1}\right)\right)\right]^{-1}\\
&\left[A^{-1}(z^{-1})\otimes A^{-1}(z)\right]\left(\frac{\partial \operatorname{vec}\Lambda}{\partial\theta^{\mathrm{T}}}\right)\frac{dz}{z}\\
=&\frac{1}{4\pi i}\oint_{|z|=1}\left(\frac{\partial \operatorname{vec}\Lambda}{\partial\theta^{\mathrm{T}}}\right)^{\mathrm{T}}\left(\Lambda^{-1}\otimes\Lambda^{-1}\right)\left(\frac{\partial \operatorname{vec}\Lambda}{\partial\theta^{\mathrm{T}}}\right)\frac{dz}{z}\\
=&\frac{1}{4\pi i}\oint_{|z|=1}\left(e_p\otimes I_n\right)\left(\Lambda^{-1}\otimes\Lambda^{-1}\right)\left(e_p^{\mathrm{T}}\otimes I_n\right)\frac{dz}{z}\\
=&\frac{1}{2}\begin{bmatrix}\Lambda^{-1}\otimes\Lambda^{-1} & O_{n^2\times n^2p}\\ O_{n^2p\times n^2} & O_{n^2p\times n^2p}\end{bmatrix}.
\end{aligned}
$$

For the third and fourth lines of the equation, we briefly demonstrate that both of them are equal to zero. The first part of line three can be simplified as

$$
\begin{aligned}
&\frac{1}{4\pi i}\oint_{|z|=1} P\left(z^{-1}\right)^{\top}\left[S^{\top}(z)\otimes S(z)\right]^{-1} Q(z)\frac{dz}{z}\\
=&\frac{1}{4\pi i}\oint_{|z|=1}\left(\frac{\partial \operatorname{vec}A\left(z^{-1}\right)}{\partial\theta^{\top}}\right)^{\top}\left(A^{-1}\left(z^{-1}\right)\otimes\Lambda^{-1}\right)\frac{\partial \operatorname{vec}\Lambda}{\partial\theta^{\top}}\frac{dz}{z}\\
=&\frac{1}{4\pi i}\oint_{|z|=1} z^{-1}\left(I_{np}\otimes e_p^{\top}\otimes I_n\right)\left(A^{-1}\left(z^{-1}\right)\otimes\Lambda^{-1}\right)\left(e_p^{\top}\otimes I_n\right)\frac{dz}{z}\\
=&\left(I_{np}\otimes e_p^{\top}\otimes I_n\right)\left[\left(\frac{1}{4\pi i}\oint_{|z|=1} z^{-1}A^{-1}\left(z^{-1}\right)\frac{dz}{z}\right)\otimes\Lambda^{-1}\right]\left(e_p^{\top}\otimes I_n\right).
\end{aligned}
$$

Within the equation, we substitute z with $\frac{1}{w}$. As a result,

$$
\frac{1}{4\pi i}\oint_{|z|=1} z^{-1}A^{-1}\left(z^{-1}\right)\frac{dz}{z}=\frac{-1}{4\pi i}\oint_{|w|=1} wA^{-1}(w)w\frac{-dw}{w^2}=\frac{1}{4\pi i}\oint_{|w|=1} A^{-1}(w)dw=0.
$$

doi:10.6342/NTU202301697

The other parts of the third and fourth lines follow similar computations. Combining the result of the first and second lines, the final result is

$$F_{\Lambda,C}(\theta) = \begin{bmatrix} \frac{1}{2}\Lambda^{-1} \otimes \Lambda^{-1} & O_{n^2 \times n^2 p} \\ O_{n^2 p \times n^2} & F_C(\theta) \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\Lambda^{-1} \otimes \Lambda^{-1} & O_{n^2 \times n^2 p} \\ O_{n^2 p \times n^2} & f(\theta) \otimes \Lambda^{-1} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{2}\Lambda^{-1} & O_{n \times np} \\ O_{np \times n} & f(\theta) \end{bmatrix} \otimes \Lambda^{-1}$$

Here, $F_{\Lambda,C}(\theta)$ denotes the Fisher information matrix with unknown Λ, and $F_C(\theta)$ denotes the Fisher information matrix with known Λ in Section 3.2.

By Komaki (1999), the Fisher information matrix $g(\theta)$ of an autoregressive model (AR) with unknown variance σ^2 is

$$g_{IJ} = \left(\begin{array}{c|ccc} g_{00} & \cdots & g_{0i} & \cdots \\ \hline \vdots & \cdots & \cdots & \cdots \\ g_{i0} & \vdots & g_{ij} & \vdots \\ \vdots & \vdots & \cdots & \vdots \end{array} \right), \text{ with } \begin{cases} g_{00} = \frac{1}{2\sigma^4} \\ g_{0i} = g_{i0} = 0 \\ g_{ij} = \frac{1}{1 - z_i z_j} \end{cases}.$$

Here, g_{00} denotes the Fisher information within the parameter σ^2, and g_{ij} is the Fisher information matrix within the parameters z_i, which is a change of coordinate from the parameters a_i. Since the AR model is the one-dimensional case of the VAR(p) model, the similarity of the Information matrix is apparent. In both the AR and VAR cases, the Fisher information within the covariance Λ are both of the form $\frac{1}{2}\Lambda^{-1} \otimes \Lambda^{-1}$. In addition, the Fisher information between the white noise covariance Λ and the regressive parameters C_i are all zero. This suggests that the two sets of parameters are information orthogonal parameters, which suggests that their maximal likelihood estimators are asymptotically uncorrelated.

However, the Fisher information matrix within the lag parameters is fundamentally different. In the AR(p) model, the information matrix is independent of σ^2, so when considering the superharmonic functions, we only need to consider a function with the lag parameters as the input. In contrast, the information matrix in the VAR(p) model consists of parameters from both C_i and Λ. So the corresponding superharmonic function must take

doi:10.6342/NTU202301697

both Λ and C_i as inputs. In Section 3.5, we will provide some examples to demonstrate the difference between the AR(p) and VAR(p) models.

3.5 Example: VAR(1) Model with Dimension Two

To demonstrate our results, we consider the simplest case: VAR(1) with $n = 2$

$$\begin{bmatrix} y_{t,1} \\ y_{t,2} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} y_{t-1,1} \\ y_{t-1,2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t,1} \\ \varepsilon_{t,2} \end{bmatrix}, \quad \varepsilon \sim \mathcal{N}(0, \Lambda).$$

Since the Fisher information matrix of Λ remains the same in all cases, it will be considered a known parameter for the following examples.

The simplest case to consider is when C and Λ are diagonal matrices. Then the model reduces to two independent AR(1) models, i.e.

$$\begin{aligned} y_{t,1} &= C_{11} y_{t-1,1} + \varepsilon_{t,1} \\ y_{t,2} &= C_{22} y_{t-1,2} + \varepsilon_{t,2}. \end{aligned}$$

The Fisher information matrix is simply $F(\theta) = \begin{bmatrix} \frac{1}{1-C_{11}{}^2} & 0 \\ 0 & \frac{1}{1-C_{22}{}^2} \end{bmatrix}$. The superharmonic function, in this case, is

$$\phi = (1 - C_{11})^{\frac{1}{2}} (1 - C_{22})^{\frac{1}{2}}, \text{ with } \Delta\phi = -2\phi.$$

This case shows that for a VAR model that reduces to an AR model, the Fisher information matrix is independent of Λ, and therefore the superharmonic function is independent of Λ.

However, in a similar case when C is diagonal, but with $\Lambda = \begin{bmatrix} p, & r \\ r, & q \end{bmatrix}$. The FIM becomes

$$F(\theta) = \begin{bmatrix} \frac{pq}{1-C_{11}^2} & \frac{-r^2}{1-C_{11}C_{22}} \\ \frac{-r^2}{1-C_{11}C_{22}} & \frac{pq}{1-C_{22}^2} \end{bmatrix} \frac{1}{\det(\Lambda)} \tag{3.4}$$

doi:10.6342/NTU202301697

Although the model is very similar to the previous case, we can see that the Fisher information matrix now contains the white noise parameters p, r, q. The dependence between $\epsilon_{t,1}$ and $\epsilon_{t,2}$ reflexes on the information matrix, resulting in the non-diagonal entries being non-zero. From (Tanaka, 2018), we know that for a metric of the form

$$F(\theta) = \begin{bmatrix} \frac{1}{1-C_{11}^2} & \frac{1}{1-C_{11}C_{22}} \\ \frac{1}{1-C_{11}C_{22}} & \frac{1}{1-C_{22}^2} \end{bmatrix},$$

the superharmonic function could be $1 - C_{11}C_{22}$. However, in (3.4), the superharmonic function must also include p, q, r as inputs, so its form is still unknown.

For the general case of the VAR(1) model with dimension two, the Fisher information matrix is as follows

$$\begin{aligned} F(\theta) = &\begin{bmatrix} pD_{11} + qD_{14} + r(D_{12} + D_{13}), & pD_{31} + qD_{34} + r(D_{32} + D_{33}) \\ pD_{21} + qD_{24} + r(D_{22} + D_{23}), & pD_{41} + qD_{44} + r(D_{42} + D_{43}) \end{bmatrix} \\ &\otimes \begin{bmatrix} p, & r \\ r, & q \end{bmatrix}^{-1} \end{aligned}$$

where $D = (I_4 - \mathrm{C} \otimes \mathrm{C})^{-1}$. This case shows that to compute and obtain the superharmonic function of the general model, we must further simplify the metric or change to another appropriate coordinate. For our current explicit form of the matrix, the possible superharmonic function remains too complicated to predict.

The last case we like to highlight is the special case when $\Lambda = \sigma^2 I_2$, the Fisher information matrix is

$$\begin{aligned} F(\theta) &= f(\theta) \otimes \Lambda^{-1} \\ &= \left\{ \left[(\mathrm{vec}\, I_2)^{\mathrm{T}} \otimes I_2 \right] \left[I_2 \otimes ((I_4 - \mathrm{C} \otimes \mathrm{C})^{-1} \mathrm{vec}(\Lambda)) \right] \right\} \otimes \Lambda^{-1} \\ &= \left\{ \left[(\mathrm{vec}\, I_2)^{\mathrm{T}} \otimes I_2 \right] \left[I_2 \otimes ((I_4 - \mathrm{C} \otimes \mathrm{C})^{-1} \mathrm{vec}(I_2)) \right] \right\} \otimes I_2 \\ &= \begin{bmatrix} D_{11} + D_{14}, & D_{31} + D_{34} \\ D_{21} + D_{24}, & D_{41} + D_{44} \end{bmatrix} \otimes I_2, \end{aligned}$$

where $D = (I_4 - \mathrm{C} \otimes \mathrm{C})^{-1}$. Note that the FIM is independent of Λ, which differs from

doi:10.6342/NTU202301697

all the above cases. Only when the white noise in each dimension is independent and identically distributed following $\mathcal{N}(0, \sigma^2)$, the matrix will be independent of Λ which is more similar to the AR(p) case. This phenomenon stays true for all dimensions of the VAR(p) model, so it is indeed an important case. With this special case, the superharmonic function could be independent of Λ and can be derived following a similar method according to Tanaka (2018).

doi:10.6342/NTU202301697

doi:10.6342/NTU202301697

Chapter 4 Conclusion and Outlooks

4.1 Conclusion

We started this research to calculate the superharmonic prior of a VAR(p) model, and we have computed the explicit Fisher information matrices to describe the model manifold. However, the calculation of the explicit form of Jeffreys prior and the Laplacian of superharmonic functions remains an obstacle. The main difference between the AR(p) models and the VAR(p) models is that for the general VAR(p) model, the Fisher metric of the model manifold contains the parameter Λ. As a result, the corresponding superharmonic function is unlikely to be independent of Λ, and it is difficult to predict its possible forms. However, we study a special case when $\Lambda = \sigma^2 I_n$. In this case, the Fisher metric is independent of Λ, so it will likely be a good first step for predicting and calculating the superharmonic function. Also, according to Tanaka (2018), a change of coordinate on the model manifold is essential to simplify the computation. Hence, we have several future directions to take.

4.2 Future Directions

A viable change of variable of the AR(p) model is to consider the eigenvalue decomposition of the parameter matrix C and use the eigenvalues to construct a one-to-one relation with the original parameters. However, the parameter matrix, in this case, is a

doi:10.6342/NTU202301697

companion matrix of the form

$$C = \begin{bmatrix} a_1, & a_2, & \cdots, & a_{p-1}, & a_p \\ -1 & 0 & & 0 & 0 \\ 0 & -1 & & \vdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 0 \end{bmatrix} = QDQ^{-1}, \quad D = \mathrm{diag}\left(z_1, z_2, \cdots, z_p\right)$$

Because there are p parameters to be estimated in the companion matrix C, and there are also p parameters on the diagonal entries of D, the relation between the transformation is clear. However, for a general parameter matrix of a VAR(1) model

$$C = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{n1} & \cdots & c_{nn} \end{bmatrix}$$

the number of parameters is n^2, which exceeds the number of its eigenvalues, so the above method can not apply.

A possible choice is to consider the singular value decomposition of $C = U\Sigma V^\top$. Here, U and V are unitary matrices, and Σ is a diagonal matrix. We need to define n^2 parameters from these three matrices. Since we assume stationarity for our model, the eigenvalues $\{z_i\}$ of matrix C are inside the unit circle. Therefore, when it comes to determining the superharmonic function, the possible candidate could be $\phi(\theta) = \prod_{i<j}(1 - z_i z_j)$ just like the one from an AR(p) model. This function not only ensures that it will be positive but also suggests that $\Delta\phi = -\frac{p(p-1)}{2}\phi$ in the AR(p) setting (Tanaka, 2018). So we hope this function's property will be preserved in the VAR(p) setting.

Another future direction is to establish the existence of the superharmonic functions on the VAR(p) model manifold. Komaki (2006) pointed out that when the sectional curvature of the model manifold is negative everywhere, the superharmonic function must exist. So with the Fisher metric of the VAR(p) model we derived, we can use it to calculate the sectional curvature of the model manifold and verify the existence of the superharmonic function,

doi:10.6342/NTU202301697

Last but not least, the Fisher information matrix of the VAR(p) model is of the form $F(\theta) = f(\theta) \otimes \Lambda^{-1}$, which suggests that although the matrix has size $n^2p \times n^2p$, its property is mainly determined by $f(\theta)$, a $np \times np$ matrix. Note that the VAR model is a sub-model of the vector autoregressive moving-average (VARMA) model, so if we can extend our work to compute the Fisher information matrix of the VARMA model, the intrinsic relations between the parameters could be further examined and discussed.

doi:10.6342/NTU202301697

doi:10.6342/NTU202301697

References

Amari, S. i. (2016), *Information geometry and its applications*, Vol. 194, Springer.

Berger, J. O. and Bernardo, J. M. (1992), 'On the development of the reference prior method', *Bayesian statistics* **4**(4), 35–60.

Brown, L. D. (1971), 'Admissible estimators, recurrent diffusions, and insoluble boundary value problems', *The Annals of Mathematical Statistics* **42**(3), 855–903.

Chen, W. Y. and Louck, J. D. (1996), 'The combinatorial power of the companion matrix', *Linear Algebra and its Applications* **232**, 261–278.

Choi, J. and Mullhaupt, A. P. (2015), 'Geometric shrinkage priors for kählerian signal filters', *Entropy* **17**(3), 1347–1357.

Cross, J. L., Hou, C. and Poon, A. (2020), 'Macroeconomic forecasting with large bayesian vars: Global-local priors and the illusion of sparsity', *International Journal of Forecasting* **36**(3), 899–915.

Den Haan, W. J. and Levin, A. (1998), 'Vector autoregressive covariance matrix estimation', *University of California, San Diego, manuscript* .

Ghosh, M. (2011), 'Objective Priors: An Introduction for Frequentists', *Statistical Science* **26**(2), 187–202.
URL: *https://doi.org/10.1214/10-STS338*

Hammarling, S. J. (1982), 'Numerical solution of the stable, non-negative definite lyapunov equation lyapunov equation', *IMA Journal of Numerical Analysis* **2**(3), 303–323.

Higham, N. J. (2008), *Functions of matrices: theory and computation*, SIAM.

doi:10.6342/NTU202301697

Kadiyala, K. R. and Karlsson, S. (1997), 'Numerical methods for estimation and inference in bayesian var-models', *Journal of Applied Econometrics* **12**(2), 99–132.

Klein, A. (2000), 'A generalization of whittle's formula for the information matrix of vector-mixed time series', *Linear algebra and its applications* **321**(1-3), 197–208.

Komaki, F. (1999), 'An estimating method for parametric spectral densities of gaussian time series', *Journal of Time Series Analysis* **20**(1), 31–50.

Komaki, F. (2006), 'Shrinkage priors for bayesian prediction'.

Langley, T. E., McNeill, A., Lewis, S., Szatkowski, L. and Quinn, C. (2012), 'The impact of media campaigns on smoking cessation activity: a structural vector autoregression analysis', *Addiction* **107**(11), 2043–2050.

Opgen-Rhein, R. and Strimmer, K. (2007), 'Learning causal networks from systems biology time course data: an effective model selection procedure for the vector autoregressive process', *BMC bioinformatics* **8**(2), 1–8.

Prüser, J. (2021), 'The horseshoe prior for time-varying parameter vars and monetary policy', *Journal of Economic Dynamics and Control* **129**, 104188.

Sims, C. A. (1980), 'Macroeconomics and reality', *Econometrica: journal of the Econometric Society* pp. 1–48.

Stock, J. H. and Watson, M. W. (2001), 'Vector autoregressions', *Journal of Economic perspectives* **15**(4), 101–115.

Tanaka, F. (2003), 'The sectional curvature of ar model manfolds', *Tensor* .

Tanaka, F. (2018), 'Superharmonic priors for autoregressive models', *Information Geometry* **1**(2), 215–235.

Tanaka, F. and Komaki, F. (2008), 'A superharmonic prior for the autoregressive process of the second-order', *Journal of Time Series Analysis* **29**(3), 444–452.

Welch, B. L. and Peers, H. (1963), 'On formulae for confidence points based on integrals of weighted likelihoods', *Journal of the Royal Statistical Society: Series B (Methodological)* **25**(2), 318–329.

doi:10.6342/NTU202301697

Whittle, P. (1963), 'On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix', *Biometrika* **50**(1-2), 129–134.

Wild, B., Eichler, M., Friederich, H. C., Hartmann, M., Zipfel, S. and Herzog, W. (2010), 'A graphical vector autoregressive modelling approach to the analysis of electronic diary data', *BMC medical research methodology* **10**, 1–13.

doi:10.6342/NTU202301697