Macroeconomic Forecasting: Statistically Adequate, Temporal Principal Components

Brian Arthur D'Orazio

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Economics

Aris Spanos, Chair

Richard Ashley

Niloy Bose

Ali Habibnia

April 17, 2023

Blacksburg, Virginia

Keywords: Principal Component Analysis, Forecasting, Vector Autoregressive Models, Statistical Adequacy

Macroeconomic Forecasting: Statistically Adequate, Temporal Principal Components

Brian Arthur D'Orazio

(ABSTRACT)

The main goal of this dissertation is to expand upon the use of Principal Component Analysis (PCA) in macroeconomic forecasting, particularly in cases where traditional principal components fail to account for all of the systematic information making up common macroeconomic and financial indicators. At the outset, PCA is viewed as a statistical model derived from the reparameterization of the Multivariate Normal model in Spanos (1986). To motivate a PCA forecasting framework prioritizing sound model assumptions, it is demonstrated, through simulation experiments, that model mis-specification erodes reliability of inferences. The Vector Autoregressive (VAR) model at the center of these simulations allows for the Markov (temporal) dependence inherent in macroeconomic data and serves as the basis for extending conventional PCA. Stemming from the relationship between PCA and the VAR model, an operational out-of-sample forecasting methodology is prescribed incorporating statistically adequate, temporal principal components, i.e. principal components which capture not only Markov dependence, but all of the other, relevant information in the original series. The macroeconomic forecasts produced from applying this framework to several, common macroeconomic indicators are shown to outperform standard benchmarks in terms of predictive accuracy over longer forecasting horizons.

Macroeconomic Forecasting: Statistically Adequate, Temporal Principal Components

Brian Arthur D'Orazio

(GENERAL AUDIENCE ABSTRACT)

The landscape of macroeconomic forecasting and nowcasting has shifted drastically in the advent of big data. Armed with significant growth in computational power and data collection resources, economists have augmented their arsenal of statistical tools to include those which can produce reliable results in big data environments. At the forefront of such tools is Principal Component Analysis (PCA), a method which reduces the number of predictors into a few factors containing the majority of the variation making up the original data series. This dissertation expands upon the use of PCA in the forecasting of key, macroeconomic indicators, particularly in instances where traditional principal components fail to account for all of the systematic information comprising the data. Ultimately, a forecasting methodology which incorporates temporal principal components, ones capable of capturing both time dependence as well as the other, relevant information in the original series, is established. In the final analysis, the methodology is applied to several, common macroeconomic and financial indicators. The forecasts produced using this framework are shown to outperform standard benchmarks in terms of predictive accuracy over longer forecasting horizons.

Dedication

To all of the people who have encouraged and supported me along this journey.

Acknowledgments

First, I would like to extend a special thank you to my advisor, Dr. Aris Spanos. Throughout my graduate education, Dr. Spanos not only renewed my appreciation for econometrics through his comprehensive knowledge of the field, but he also fostered an environment where I could build self-confidence. He never discouraged my inquisitive nature, and I learned much from our discussions. I would also like to thank my committee members, Dr. Richard Ashley, Dr. Niloy Bose and Dr. Ali Habibnia, for their comments and questions. Each conversation with Dr. Ashley and Dr. Habibnia provided me with new research directions, as well as a desire to understand more about econometrics. Outside of research, Dr. Bose's belief in his students inspired me to achieve more, and I appreciate all of the advice and encouragement he offered during my time at Virginia Tech.

Finishing this dissertation would have been next to impossible without the constant love and support of my parents, Cosmo and Diane, and my brother, Anthony. They have always pushed me to persevere and overcome challenges, and they have made me a better person for it. I will forever be indebted to their selflessness and generosity.

Last but not least, I would like to thank all of my friends for being there during this endeavor. At Virginia Tech, I especially appreciated the time I spent with Jonathan Gendron, Haoyu Wang, Hoang Phuong Do, Chang Geun Song and Zichao Yang. Their emotional support and wisdom were invaluable for completing this dissertation. Outside of Virginia Tech, I would like to thank my long-time friend Michael Rand for encouraging me to pursue a doctorate in economics.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Principal component analysis (PCA) is one of the most commonly utilized techniques for reducing the number of dimensions in a dataset, with earliest known applications dating back to Pearson (1901) and Hotelling (1933). Initially, the use of PCA was quite limited, as principal components are constructed from the decomposition of the variance-covariance matrix, which requires the computation of matrix determinants, characteristic polynomials and eigenvalues/eigenvectors (Anderson, 2003). However, despite initial drawbacks, the demand for tools which allow for the analysis of many predictors has grown as an increasing amount of data has become more readily available. Studies which utilize high dimensional datasets, where the number of predictors exceeds the number of observations, often face what is referred to as the "curse of dimensionality," a situation where standard statistical techniques are rendered ineffective (Bellman, 1961). PCA is especially relevant in these environments, as reducing the dimensionality of large datasets to a few, key factors is its main purpose.

In macroeconomics, while the idea of using the common variation among different data series to create a general index which captures business cycles has existed since Burns and Mitchell (1946), it was not until the seminal papers Sargent and Sims (1977) and Geweke (1977) that a formal framework for dynamic factor models (DFM) was created. Since then, the advent of "big data" and an increased ease of access to thousands of macroeconomic and financial data series has spurred a subsection of econometric literature solely dedicated to these models and their relation to macroeconomic forecasting; see Stock and Watson (2016). However, principal components remain at the core of the many such models as one of the primary methods for consistent estimation of the underlying latent factors (Bernanke et al., 2005, Stock and Watson, 2002b). See Barigozzi et al. (2019) for a recent example of a generalized dynamic factor model which utilizes dynamic principal components in its estimation.

The focus of this research is to create a formal forecasting framework where principal components can capture the characteristics inherent in many macroeconomic time series. Ultimately, the fundamental goal of this dissertation is the establishment of a statistically adequate principal components model, one which takes it account all of the systematic information in the dataset. To accomplish this, the traditional definition of principal components in this dissertation is extended to account for temporal dependence using a link between PCA and the Normal Vector Autoregressive (VAR) model, following Solat (2018) and Solat and Tsang (2021). These "temporal" principal components differ from existing Dynamic PCA found in Barigozzi et al. (2019), Brillinger (1981), which form principal components based on spectral analysis in the frequency domain. Such dynamic principal components, however, assume that the series are stationary from the outset, making them largely inapplicable for macroeconomic data.

1.2 Overview of Chapters

This dissertation is organized as follows. In order to understand the logic behind the temporal principal components at the end of the dissertation, the remainder of this chapter is dedicated to shedding light on the traditional definition of principal components and drawing out the connection between PCA and the simple, Multivariate Normal model found in Spanos (1986).

Chapter 2 reviews the model-based frequentist approach to inference discussed in Spanos (2021) and explores the effects of statistical mis-specification in the context of Vector Autoregressive (VAR) models. This is particularly relevant, as the temporal principal components are derived via a connection to the VAR model. Through simulation, it is shown that statistical mis-specification of the VAR model can lead to misleading inferences, where the nominal and actual error probabilities substantially differ.

Chapter 3 formally introduces the Normal, Temporal Principal Components model, where the estimated principal components capture information on time dependence. A forecasting methodology emphasizing statistical adequacy is then designed around these principal components, and forecasts are estimated for commonly used macroeconomic data series. The out-of-sample forecasts generated from the statistically adequate model are shown to outperform models which neglect statistical adequacy, such as the FAVAR (Bernanke et al., 2005).

Chapter 4 provides a summary of the dissertation contributions alongside a discussion on future extensions.

1.3 The Traditional Perspective on Principal Components

1.3.1 Multivariate Normal Model Specification

In order to understand the traditional perspective of principal components, one must first understand and fully describe the Multivariate Normal model; the link between the two will become clear when defining principal components. Let $\{\mathbf{Z}_t, t \in \mathbb{N}\}$ be a vector stochastic process, where each random vector $\mathbf{Z}_t$ is of length $(k \times 1)$. Specifically,

$$\mathbf{Z}_t = \begin{bmatrix} X_{1,t} \\ X_{2,t} \\ \vdots \\ X_{k,t} \end{bmatrix},$$

where the random variable $X_{i,t}$, in this context, will denote the macroeconomic predictor of interest i at time t. Moreover, assume that

$$E(\mathbf{Z}_t) = \begin{bmatrix} E(X_{1,t}) \\ E(X_{2,t}) \\ \vdots \\ E(X_{k,t}) \end{bmatrix} \equiv \boldsymbol{\mu},$$

which is also a vector of length $(k \times 1)$. The $(k \times k)$ contemporaneous variance-covariance matrix, $Cov(\mathbf{Z}_t) = E((\mathbf{Z}_t - \boldsymbol{\mu})(\mathbf{Z}_t - \boldsymbol{\mu})^\top)$, is defined as:

$$Cov(\mathbf{Z}_t) = \begin{bmatrix} Var(X_{1,t}) & Cov(X_{1,t}, X_{2,t}) & \dots & Cov(X_{1,t}, X_{k,t}) \\ Cov(X_{2,t}, X_{1,t}) & Var(X_{2,t}) & \dots & Cov(X_{2,t}, X_{k,t}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_{k,t}, X_{1,t}) & Cov(X_{k,t}, X_{2,t}) & \dots & Var(X_{k,t}) \end{bmatrix} \equiv \boldsymbol{\Sigma}.$$

Note that here it is assumed that the variance-covariance matrix is positive definite. To illustrate the full model specification of the Multivariate Normal model, complete with a set of testable model assumptions, this section will utilize the probabilistic reduction (PR) approach in Spanos (2019a). This approach first requires the partitioning of the space of possible statistical models using three, broad categories of assumptions, namely (i) Distribution, (ii) Dependence, (iii) Heterogeneity. More information on this reduction can be found in Chapter 2, which delves deeper into the model-based frequentist approach to inference, as well as the definition and role of the probabilistic reduction approach. For the Multivariate Normal model, the stochastic process $\{\mathbf{Z}_t, t \in \mathbb{N}\}$ is assumed to be (i) multivariate normal, (ii) Independent(I), (iii) Identically distributed(ID).

These assumptions ultimately simplify the joint distribution of the sample, $D(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n)$, as shown below:

$$D(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n; \boldsymbol{\psi}) \stackrel{\text{I}}{=} \prod_{t=1}^{n} D_t(\mathbf{Z}_t; \boldsymbol{\phi}_t) \stackrel{\text{IID}}{=} \prod_{t=1}^{n} D(\mathbf{Z}_t; \boldsymbol{\phi}).$$

For notation clarification, the distribution 'D' is a generalization meant to represent either the probability density function or cumulative distribution function. When deriving the log-likelihood function in the subsequent section, the notation for the probability density function will be specified using the conventional notation. Hence, the joint distribution is simply the product of the distributions of each $\mathbf{Z}_t$.[1] Under normality, the reduction implies that:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_n \end{bmatrix} \sim N \left(\boldsymbol{\mu}_y = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{bmatrix}, \boldsymbol{\Sigma}_y = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \mathbf{0} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{\Sigma} \end{bmatrix} \right)$$

,

[1] In this particular setup, $\boldsymbol{\phi} := (\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

where $\boldsymbol{\mu}_y$ is a vector of length $(nk \times 1)$ and $\boldsymbol{\Sigma}_y$ is a block diagonal matrix of size $(nk \times nk)$. Given the assumptions above, the reduction simplifies to the statistical model in Table 1.1 (Spanos, 1986). In other words, the reduction shows that for $t \in \mathbb{N}$, $\mathbf{Z}_t \sim \text{NIID}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Table 1.1: Multivariate Normal Model

	Statistical GM:	$\mathbf{Z}_t = \boldsymbol{\mu} + \mathbf{u}_t, t \in \mathbb{N}$,
[1]	Normality:	$\mathbf{Z}_t \sim N(.,.)$
[2]	Constant Mean:	$E(\mathbf{Z}_t) = \boldsymbol{\mu}$
[3]	Constant Covariance:	$Cov(\mathbf{Z}_t) = \boldsymbol{\Sigma}$
[4]	Independence:	$\mathbf{Z}_t$ is an independent process

Note that in Table 1.1, $\mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma})$.

1.3.2 Maximum Likelihood Estimation

To derive the Maximum Likelihood Estimators of $\boldsymbol{\phi} := (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, first observe that the probability density function of $\mathbf{Z}_t$ from the reduction is assumed to be multivariate normal, i.e.:

$$f(\mathbf{Z}_t; \boldsymbol{\phi}) = 2\pi^{-\frac{k}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{Z}_t - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Z}_t - \boldsymbol{\mu})\right).$$

Thus, the the likelihood function for the sample takes the form:

$$\begin{aligned} L(\boldsymbol{\psi}; \mathbf{Y}) \propto \prod_{t=1}^{T} f(\mathbf{Z}_t; \boldsymbol{\phi}) &= \prod_{t=1}^{T} 2\pi^{-\frac{k}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{Z}_t - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Z}_t - \boldsymbol{\mu})\right) \\ &= 2\pi^{-\frac{kT}{2}} |\boldsymbol{\Sigma}|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}(\mathbf{Z}_t - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Z}_t - \boldsymbol{\mu})\right), \end{aligned}$$

where $|\boldsymbol{\Sigma}|$ denotes the determinant of $\boldsymbol{\Sigma}$. Hence, the log-likelihood function is:

$$\begin{aligned}\log L(\boldsymbol{\psi};\mathbf{Y}) &= -\frac{kT}{2}\log(2\pi) - \frac{T}{2}\log(|\boldsymbol{\Sigma}|) - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{Z}_t-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{Z}_t-\boldsymbol{\mu}) \\ &= c - \frac{T}{2}\log(|\boldsymbol{\Sigma}|) - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{Z}_t-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{Z}_t-\boldsymbol{\mu}).\end{aligned}$$

Since it is assumed that $\boldsymbol{\Sigma}$ is invertible, the log likelihood function can be rewritten as:

$$\log L(\boldsymbol{\psi};\mathbf{Y}) = c + \frac{T}{2}\log(|\boldsymbol{\Sigma}^{-1}|) - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{Z}_t-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{Z}_t-\boldsymbol{\mu}).$$

Taking partial derivatives and setting them equal to zero yields the maximum likelihood estimators:

$$\begin{aligned}\frac{\partial\log L(\boldsymbol{\psi};\mathbf{Y})}{\partial\boldsymbol{\mu}} &= \sum_{t=1}^{T}\boldsymbol{\Sigma}^{-1}(\mathbf{Z}_t-\boldsymbol{\mu}) = \mathbf{0} \\ &\Rightarrow \hat{\boldsymbol{\mu}}_{MLE} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t.\end{aligned}$$

$$\begin{aligned}\frac{\partial\log L(\boldsymbol{\psi};\mathbf{Y})}{\partial\boldsymbol{\Sigma}^{-1}} &= \frac{T}{2}\boldsymbol{\Sigma} - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{Z}_t-\boldsymbol{\mu})(\mathbf{Z}_t-\boldsymbol{\mu})^\top = \mathbf{0} \\ &\Rightarrow \hat{\boldsymbol{\Sigma}}_{MLE} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{Z}_t-\hat{\boldsymbol{\mu}}_{MLE})(\mathbf{Z}_t-\hat{\boldsymbol{\mu}}_{MLE})^\top\end{aligned}$$

Thus, the Maximum likelihood estimation reveals that the MLE for $\boldsymbol{\phi} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ take the following forms:

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t, \;\; \hat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{Z}_t-\hat{\boldsymbol{\mu}})(\mathbf{Z}_t-\hat{\boldsymbol{\mu}})^\top,$$

i.e. the sample mean and sample variance, respectively; see Spanos (1986) for further explanation on the derivations.[2]

1.3.3 The Normal Principal Component Model

The Normal Principal Components model is specified as a re-parameterization of the Multivariate Normal model above. Hence, consider the $(k \times 1)$ random vector $\mathbf{Z}_t$ defined during the discussion of the simple Multivariate Normal model. Reflecting the earlier setup, let $E(\mathbf{Z}_t) = \boldsymbol{\mu}$ and $Cov(\mathbf{Z}_t) = E((\mathbf{Z}_t - \boldsymbol{\mu})(\mathbf{Z}_t - \boldsymbol{\mu})^\top) = \boldsymbol{\Sigma}$. The primary aim of principal components is to reduce the number of dimensions in the dataset while also retaining as much of the variability in the original dataset as possible (Jolliffe, 2002).

Assuming that the contemporary covariance matrix $\boldsymbol{\Sigma}$ is positive definite and symmetric, it follows that the matrix admits a spectral decomposition of the form:

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top, \tag{1.1}$$

where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0)$ denotes the ranked eigenvalues of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Gamma} := (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_k)$ is a $(k \times k)$ matrix whose columns consist of orthonormal eigenvectors corresponding to the ranked eigenvalues. One then selects the first $p, p \leq k$ eigenvectors to form linear combinations with the demeaned, random vector, i.e., formulates the first p

[2]The simplification of the MLE for the variance-covariance matrix uses properties of invertible and symmetric matrices. As shown in Greene (2018), since it is assumed that $\boldsymbol{\Sigma}$ is invertible and symmetric,

$$\frac{\partial \log|\boldsymbol{\Sigma}^{-1}|}{\partial \boldsymbol{\Sigma}^{-1}} = \left((\boldsymbol{\Sigma}^{-1})^{-1}\right)^\top = \boldsymbol{\Sigma}$$

Moreover, it can be shown that:

$$\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{A}} = \mathbf{x}\mathbf{x}^\top,$$

where $\mathbf{A}$ is a symmetric matrix and $\mathbf{x}$ is a column vector.

principal components:

$$\mathbf{C}_t = \mathbf{\Gamma}_p^\top(\mathbf{Z}_t - \boldsymbol{\mu}) \tag{1.2}$$

When normality is imposed:

$$\mathbf{C}_t \sim N(\mathbf{0}, \mathbf{\Lambda}_p).$$

Similar to the specification in Solat (2018), the Normal Principal Components model is described in Table 1.2 below, along with the corresponding probabilistic assumptions. Note that the error term, denoted as $\mathbf{u}_t$, includes the approximation error, which stems from the exclusion of certain principal components.

Table 1.2: Normal Principal Components Model

	Statistical GM:	$\mathbf{C}_t = \mathbf{\Gamma}_p^\top(\mathbf{Z}_t - \boldsymbol{\mu}) + \mathbf{u}_t,$
[1]	Normality:	$\mathbf{C}_t \sim N(.,.)$
[2]	Linearity:	$E(\mathbf{C}_t) = \mathbf{\Gamma}_p^\top(\mathbf{Z}_t - \boldsymbol{\mu}),$
[3]	Constant Covariance:	$Cov(\mathbf{C}_t) = \mathbf{\Lambda}_p,$
[4]	Independence:	$\{\mathbf{C}_t, t \in \mathbb{N}\}$ is an independent process
[5]	t-invariance:	$\boldsymbol{\theta} := \{\boldsymbol{\mu}, \mathbf{\Lambda}_p, \mathbf{\Gamma}_p\}$ does not change with t

For more information regarding the conventional approach to specifying principal components, see Jolliffe (2002) or Anderson (2003). The authors of the respective books thoroughly discuss the process of defining principal components through the lens of a constrained optimization problem, where the objective function being maximized is the variance of $\boldsymbol{\gamma}_i^\top(\mathbf{Z}_t - \boldsymbol{\mu})$ subject to $\boldsymbol{\gamma}_i^\top\boldsymbol{\gamma}_i = 1$, for $i = 1, \ldots, k$.

1.3.4 Estimation of Principal Components

Under the normality assumption, the derivation of the Maximum Likelihood Estimators (MLE) for the parameters in $\boldsymbol{\theta} := \{\boldsymbol{\mu}, \boldsymbol{\Lambda}_p, \boldsymbol{\Gamma}_p\}$ is possible because of the invariance of the MLE to different parameterizations (Spanos, 1986). The definition of parameterization invariance is given in Spanos (1986, 2019a):

Definition 1.1 (Parameterization Invariance of MLE)**.** Let $\hat{\boldsymbol{\theta}}$ be the MLE of $\boldsymbol{\theta}$. If $\boldsymbol{\phi} = h(\boldsymbol{\theta})$ is a Borel function of $\boldsymbol{\theta}$, then the MLE of $\boldsymbol{\phi}$ exists and is given by: $\hat{\boldsymbol{\phi}} = h(\hat{\boldsymbol{\theta}})$.

Given this property, the MLE for $\boldsymbol{\theta}$ are derived by simply applying the matrix algebra above on the MLE of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ obtained from the simple Multivariate Normal model:

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{t=1}^{n}\mathbf{Z}_t, \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{t=1}^{n}(\mathbf{Z}_t - \hat{\boldsymbol{\mu}})(\mathbf{Z}_t - \hat{\boldsymbol{\mu}})^\top. \tag{1.3}$$

In the context of principal components, the MLE for the eigenvectors and eigenvalues are obtained from applying spectral decomposition, or alternatively singular value decomposition, on the sample covariance matrix $\hat{\boldsymbol{\Sigma}}$. In the case of spectral decomposition, this amounts to factoring $\hat{\boldsymbol{\Sigma}}$ into:

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Gamma}}^\top,$$

The estimated principal components are then constructed using the first p ordered eigenvalues and corresponding eigenvectors:

$$\hat{\mathbf{C}}_t = \hat{\boldsymbol{\Gamma}}_p^\top(\mathbf{Z}_t - \hat{\boldsymbol{\mu}}).$$

Discussion of the sampling distributions for the principal components estimators can be found in Jolliffe (2002) or Anderson (2003). However, these sampling distributions require the use of asymptotic results and are not explicitly derived here. For more information regarding the exact form of these asymptotic sampling distributions, see Anderson (2003).

Chapter 2

Statistical Adequacy and the Reliability of Inference: Vector Autoregressive Models

2.1 Introduction

The primary aim of this chapter is to motivate the unreliability of inference that stems from model mis-specification, specifically within the context of Vector Autoregressive (VAR) models. In model-based, frequentist inference, reliability emanates directly from the use of statistically adequate models, i.e. models which capture all of the systematic, statistical information in the data (Spanos, 2009, 2018, 2019a). Failure to account for all of the relevant information in the data generates discrepancies between the actual and nominal error probabilities, thereby undermining the trustworthiness of any inferences. This was found to be the case in earlier works, such as Spanos and McGuirk (2001) and Spanos (2009). The simulations within this chapter also show that neglecting systematic information has a devastating effect on the sampling distributions of the VAR model estimators and creates gaps between the actual and nominal error probabilities, extending the results of the previous studies to include different forms of mean heterogeneity.

2.1.1 Statistical Models in Model-based Frequentist Inference

The foundation for model-based frequentist inference is the pre-specified statistical model (Spanos, 2021). Following Spanos (2009, 2010a), this takes the form:

$$\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x}) = \{f(\mathbf{x}; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}, \mathbf{x} \in \mathbb{R}^n_X, \Theta \subset \mathbb{R}^p, n > p, \tag{2.1}$$

where $f(\mathbf{x}; \boldsymbol{\theta})$ denotes the joint distribution of the sample $\mathbf{X} := (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$, $\mathbf{x} \in \mathbb{R}^n_X$ denotes the sample space, and $\Theta \subset \mathbb{R}^p$ signifies the parameter space. The joint distribution in the definition above encapsulates the probabilistic information characterizing the stochastic process $\{\mathbf{X}_k, k \in \mathbb{N}\}$, which serves as the basis for the data $\mathbf{x}_0 := (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$. Thus, the statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$ is viewed as a particular parameterization of the stochastic process, whose probabilistic assumptions are chosen to capture all of the systematic information in the data (Spanos, 2010a). This view is directly in line with the goal of frequentist inference: to learn about the true generating mechanism from the data, such that the data $\mathbf{x}_0$ make up a typical realization of the stochastic process (2010a).

2.1.2 Reliability of Inference

In Spanos (2010a), Spanos questions where the untrustworthiness in evidence in the field of economics comes from. As argued in the paper, one of the instigators of untrustworthy evidence is statistical inadequacy, where statistical inadequacy implies that the modeling assumptions imposed on the data $\mathbf{x}_0 := (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ are invalid for the data. The joint distribution is the conduit through which invalid assumptions affect inferences. Namely, in order to derive the sampling distributions of estimators, statistics, etc., and hence the error probabilities, one needs to go through the joint distribution of the sample $f(\mathbf{x}; \boldsymbol{\theta})$

(Spanos, 2019a). If the joint distribution is mis-specified, then any derivation which uses this distribution will also be mis-specified.

Thus, a statistically adequate model must first be established, such that the selected model captures all of the chance regularity patterns in the data. To accomplish this requires rigorous model mis-specification (M-S) testing and respecification. Only once statistical adequacy is achieved can one move on to inference procedures, such as estimation, testing, and prediction. Table 2.1 below outlines the stages of the empirical modeling and inference process as found in Spanos (2019a). After the respecification stage, it is assumed that a statistically adequate model has been attained.

Table 2.1: Stages in Empirical Modeling and Inference

Modeling	1. Specification 2. Estimation 3. Mis-specification Testing 4. Respecification
Inference	5. Estimation, testing, prediction, simulation

2.2 Normal VAR(p) Model Specification

2.2.1 The Probabilistic Reduction Approach

To illustrate the probabilistic reduction approach in Spanos and McGuirk (2001), consider the following example. Let $\{\mathbf{Z}_t, t \in \mathbb{N}\}$ be a vector stochastic process, where $\mathbf{Z}_t$ is of length $(k \times 1)$. The first step in utilizing the probabilistic reduction approach is to restrict, or partition, the space of statistical models by imposing probabilistic assumptions based on the three broad categories of (i) Distribution, (ii) Dependence, (iii) Heterogeneity.[1] In this case, the stochastic process $\{\mathbf{Z}_t, t \in \mathbb{N}\}$ is assumed to be a (i) multivariate normal, (ii) Markov dependent of order p, and (iii) Covariance Stationary process. A stochastic process is defined to be covariance, or second-order, stationary if:

$$E(\mathbf{Z}_t) = \boldsymbol{\mu}, \;\; Cov(\mathbf{Z}_t, \mathbf{Z}_{t-p}) = E((\mathbf{Z}_t - \boldsymbol{\mu})(\mathbf{Z}_{t-p} - \boldsymbol{\mu})^\top) = \boldsymbol{\Sigma}_{ij}, \;\; i, j = 0, 1, \ldots, p.$$

Given the assumptions above, the joint distribution of the sample, $D(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$, can be reduced as follows:

$$\begin{aligned} D(\mathbf{Z}_1, \ldots, \mathbf{Z}_n; \boldsymbol{\psi}) &= D(\mathbf{Z}_1^p; \boldsymbol{\phi}_0) \prod_{t=p+1}^{n} D_t(\mathbf{Z}_t | \mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}_t) \\ &\overset{\text{M}}{=} D(\mathbf{Z}_1^p; \boldsymbol{\phi}_0) \prod_{t=p+1}^{n} D_t(\mathbf{Z}_t | \mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}_t) \\ &\overset{\text{M/S}}{=} D(\mathbf{Z}_1^p; \boldsymbol{\phi}_0) \prod_{t=p+1}^{n} D(\mathbf{Z}_t | \mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}), \end{aligned} \tag{2.2}$$

where $\mathbf{Z}_1^p := (\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_p)$ and $\mathbf{Z}_{t-1}^{t-p} := (\mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \ldots, \mathbf{Z}_{t-p})$. Note that the second line of the reduction comes from assumption (ii) Markov Dependence, and the third line comes from assumption (iii) Covariance Stationary. Following Hamilton (1994), the first p observations will be ignored, resulting in the approximate form of the log-likelihood function. See Appendix G for more insight into the exact form of the log-likelihood function. Also, as was the case in Chapter 1, the distribution 'D' above is a generalization and is in this case meant to represent the probability density function. Under the assumption that $D(\mathbf{Z}_t | \mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi})$ is

[1] Restricting the stochastic process based solely on a distributional assumption, such as normality for example, will not result in a workable statistical model, since the number of dimensions and unknown parameters will increase with the sample size. To render the model operational, restrictions need to also be placed on both the dependence and heterogeneity of $\mathbf{Z}_t$ (Spanos, 2019a) for further examples.

normal, this implies that distribution of $D(\mathbf{Z}_t|\mathbf{Z}_{t-1}^{t-p};\boldsymbol{\phi})$ takes the form:

$$\begin{bmatrix} \mathbf{Z}_t \\ \mathbf{Z}_{t-1} \\ \vdots \\ \mathbf{Z}_{t-p} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01} & \boldsymbol{\Sigma}_{02} & \dots & \boldsymbol{\Sigma}_{0p} \\ \boldsymbol{\Sigma}_{10}^{\top} & \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \dots & \boldsymbol{\Sigma}_{1p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{p0}^{\top} & \boldsymbol{\Sigma}_{p1}^{\top} & \boldsymbol{\Sigma}_{p2}^{\top} & \dots & \boldsymbol{\Sigma}_{pp} \end{bmatrix}, \right) \tag{2.3}$$

where $\boldsymbol{\mu}$ is the mean vector of $\mathbf{Z}_t$ and is of vector of length $((p+1)k \times 1)$, $\boldsymbol{\Sigma}_{ij}$ is a $k \times k$ covariance matrix for $i, j = 0, 1, \dots, p$. The VAR(p) model specification stems directly from this joint distribution. To simplify, first partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as follows (Poudyal, 2012):

$$\begin{bmatrix} \mathbf{Z}_t, (k \times 1) \\ \mathbf{Z}_{t-1}^{t-p}, (pk \times 1) \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}, (k \times 1) \\ \boldsymbol{\mu}_{pk}, (pk \times 1) \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{00}, (k \times k) & \boldsymbol{\Sigma}_{01}, (k \times pk) \\ \boldsymbol{\Sigma}_{10}^{\top}, (pk \times k) & \boldsymbol{\Sigma}_{11}, (pk \times pk) \end{bmatrix};$$

Given this partition, the conditional distribution is derived to be:

$$D(\mathbf{Z}_t|\mathbf{Z}_{t-1}^{t-p};\boldsymbol{\Theta}) \sim N(\mathbf{a}_0 + \mathbf{A}^{\top}\mathbf{Z}_{t-1}^{t-p}, \boldsymbol{\Omega}), \tag{2.4}$$

where the $\boldsymbol{\Theta} := (\mathbf{a}_0, \mathbf{A}, \boldsymbol{\Omega})$ are:

$$\mathbf{A}^{\top} = \boldsymbol{\Sigma}_{01}\boldsymbol{\Sigma}_{11}^{-1} \tag{2.5}$$

$$\mathbf{a}_0 = \boldsymbol{\mu} - \mathbf{A}^{\top}\boldsymbol{\mu}_{pk} \tag{2.6}$$

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma}_{00} - \boldsymbol{\Sigma}_{01}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{01}^{\top} \tag{2.7}$$

In this case, the complete specification, including the model assumptions, takes the form found in Spanos (1986), found in Table 2.2 below.

Table 2.2: Normal Vector Autoregressive (VAR(p)) Model

		Statistical GM: $\mathbf{Z}_t = \mathbf{a}_0 + \mathbf{A}^\top \mathbf{Z}_{t-1}^{t-p} + \mathbf{u}_t,\ \ t \in \mathbb{N}$,
[1]	Normal:	$D(\mathbf{Z}_t \vert \mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\theta})$, for $\mathbf{Z}_{t-1}^{t-p} := (\mathbf{Z}_{t-1}, ... \mathbf{Z}_{t-p})$ is Normally distributed,
[2]	Linearity:	$E(\mathbf{Z}_t \vert \sigma(\mathbf{Z}_{t-1}^{t-p})) = \mathbf{a}_0 + \mathbf{A}^\top \mathbf{Z}_{t-1}^0$,
[3]	Homoskedasticity:	$Var(\mathbf{Z}_t \vert \sigma(\mathbf{Z}_{t-1}^{t-p})) = \boldsymbol{\Omega}$,
[4]	Markov(p):	$\{\mathbf{Z}_t,\ t \in \mathbb{N}\}$ is a Markov(p) process,
[5]	t-invariance:	$\boldsymbol{\theta} := (\mathbf{a}_0, \mathbf{A}, \boldsymbol{\Omega})$ are t-invariant.
		$\mathbf{a}_0 = \boldsymbol{\mu} - \mathbf{A}^\top \boldsymbol{\mu}_{pk}$, $\mathbf{A}^\top = \boldsymbol{\Sigma}_{01} \boldsymbol{\Sigma}_{11}^{-1}$, $\boldsymbol{\Omega} = \boldsymbol{\Sigma}_{00} - \boldsymbol{\Sigma}_{01} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{01}^\top$

2.2.2 Model Specification: VAR(1) With Mean-Heterogeneity

Suppose that assumption (iii), that the stochastic process $\{\mathbf{Z}_t, t \in \mathbb{N}\}$ is Covariance Stationary, is relaxed to allow for mean heterogeneity. As a simple example, consider the case where the mean of the process takes the following form:

$$E(\mathbf{Z}_t) = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t.$$

To further simplify the problem, suppose that assumption (ii) is Markov(1) dependent. Therefore, the distribution $D(\mathbf{Z}_t | \mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi})$ is:

$$\begin{bmatrix} \mathbf{Z}_t \\ \mathbf{Z}_{t-1} \end{bmatrix} \sim N \left(\begin{bmatrix} \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t \\ \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 (t-1) \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01} \\ \boldsymbol{\Sigma}_{10}^\top & \boldsymbol{\Sigma}_{11} \end{bmatrix}, \right) \tag{2.8}$$

This setup leads to VAR(1) model specification found in Table 2.3. The key difference between Table 2.2 and Table 2.3 lies in the parameterization of the constant term $\mathbf{a}_0$ and

Table 2.3: Normal Vector Autoregressive (VAR(1)) Model with Mean Heterogeneity

	Statistical GM:	$\mathbf{Z}_t = \mathbf{a}_0 + \boldsymbol{\delta}t + \mathbf{A}_1^\top \mathbf{Z}_{t-1} + \mathbf{u}_t,\ t\in\mathbb{N}$,
[1]	Normal:	$D(\mathbf{Z}_t \vert \mathbf{Z}_{t-1}; \boldsymbol{\theta})$, is Normally distributed,
[2]	Linearity:	$E(\mathbf{Z}_t \vert \sigma(\mathbf{Z}_{t-1})) = \mathbf{a}_0 + \boldsymbol{\delta}t + \mathbf{A}_1^\top \mathbf{Z}_{t-1}$,
[3]	Homoskedasticity:	$Var(\mathbf{Z}_t \vert \sigma(\mathbf{Z}_{t-1})) = \boldsymbol{\Omega}$,
[4]	Markov(p):	$\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$ is a Markov(1) process,
[5]	t-invariance:	$\boldsymbol{\theta}:=(\mathbf{a}_0, \boldsymbol{\delta}, \mathbf{A}_1, \boldsymbol{\Omega})$ are t-invariant.
		$\mathbf{a}_0 = (\mathbf{I}_2 - \mathbf{A}_1^\top)\boldsymbol{\mu}_0 + \mathbf{A}_1^\top \boldsymbol{\mu}_1,\ \boldsymbol{\delta} = (\mathbf{I}_2 - \mathbf{A}_1^\top)\boldsymbol{\mu}_1,\ \mathbf{A}_1^\top = \boldsymbol{\Sigma}_{01}\boldsymbol{\Sigma}_{11}^{-1}$,
		$\boldsymbol{\Omega} = \boldsymbol{\Sigma}_{00} - \boldsymbol{\Sigma}_{01}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{01}^\top$

the inclusion of $\boldsymbol{\delta}$. The parameterization for these terms is derived as follows:

$$\begin{aligned}
\boldsymbol{\mu}(t) - \mathbf{A}_1^\top \boldsymbol{\mu}(t-1) &= \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t - \mathbf{A}_1^\top(\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1(t-1)) \\
&= (\mathbf{I}_2 - \mathbf{A}_1^\top)\boldsymbol{\mu}_0 + \mathbf{A}_1^\top \boldsymbol{\mu}_1 + (\mathbf{I}_2 - \mathbf{A}_1^\top)\boldsymbol{\mu}_1 t \\
&= \mathbf{a}_0 + \boldsymbol{\delta}t.
\end{aligned} \tag{2.9}$$

Keep in mind that this particular example only considers the instance where there is a linear trend in the mean. In reality, however, it is not uncommon to encounter other forms of mean heterogeneity, all of which can be included in the model specification without too much difficulty. For example, another instance of mean heterogeneity that is ever-present in macroeconomic data is seasonality. One way to model seasonality is through the use of trigonometric polynomials (Ghysels and Osborn, 2001).

Following Spanos (2019a) and Ghysels and Osborn (2001), suppose that the distribution

$D(\mathbf{Z}_t|\mathbf{Z}_{t-1}^{t-p};\boldsymbol{\phi})$ is:

$$\begin{bmatrix}\mathbf{Z}_t\\ \mathbf{Z}_{t-1}\end{bmatrix}\sim N\left(\begin{bmatrix}\boldsymbol{\mu}_0+\boldsymbol{\mu}_1(t)\\ \boldsymbol{\mu}_0+\boldsymbol{\mu}_1(t-1)\end{bmatrix},\begin{bmatrix}\boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01}\\ \boldsymbol{\Sigma}_{10}^{\top} & \boldsymbol{\Sigma}_{11}\end{bmatrix},\right) \tag{2.10}$$

where

$$\mu_{1,i}(t)=\sum_{k=1}^{S/2}\gamma_{i,k}cos(\frac{2\pi kt}{S})+\psi_{i,k}sin(\frac{2\pi kt}{S}),i=1,2,\ldots,n. \tag{2.11}$$

Assume for simplicity that $S=4$, which is representative of quarterly data. Much like the linear trend case above, the main difference between this specification and that found in Table 2.2 above is in the parameterization of $\mathbf{a}_0$ and the inclusion of new terms capturing the seasonality. The derivation of these model parameters is:

$$\begin{aligned}\boldsymbol{\mu}(t)-\mathbf{A}_1^{\top}\boldsymbol{\mu}(t-1)&=\boldsymbol{\mu}_0+\boldsymbol{\gamma}_1cos\left(\frac{2\pi t}{4}\right)+\boldsymbol{\psi}_1sin\left(\frac{2\pi t}{4}\right)+\boldsymbol{\gamma}_2cos\left(\frac{2(2)\pi t}{4}\right)\\ &-\mathbf{A}_1^{\top}\left[\boldsymbol{\mu}_0+\boldsymbol{\gamma}_1cos\left(\frac{2\pi(t-1)}{4}\right)+\boldsymbol{\psi}_1sin\left(\frac{2\pi(t-1)}{4}\right)+\boldsymbol{\gamma}_2cos\left(\frac{2(2)\pi(t-1)}{4}\right)\right]\\ &=\left(\mathbf{I}_2-\mathbf{A}_1^{\top}\right)\boldsymbol{\mu}_0+\left(\boldsymbol{\gamma}_1+\mathbf{A}_1^{\top}\boldsymbol{\psi}_1\right)cos\left(\frac{\pi t}{2}\right)+\left(\boldsymbol{\psi}_1-\mathbf{A}_1^{\top}\boldsymbol{\gamma}_1\right)sin\left(\frac{\pi t}{2}\right)\\ &+\left(\boldsymbol{\gamma}_2+\mathbf{A}_1^{\top}\boldsymbol{\gamma}_2\right)cos\left(\pi t\right)\\ &=\mathbf{a}_0+\boldsymbol{\delta}_1cos\left(\frac{\pi t}{2}\right)+\boldsymbol{\delta}_2sin\left(\frac{\pi t}{2}\right)+\boldsymbol{\delta}_3cos\left(\pi t\right).\end{aligned}$$

Note that the fourth seasonal term is omitted since it is equal to zero. Hence, the full VAR(1) model specification with seasonal mean heterogeneity is provided in Table 2.4.

Table 2.4: Normal VAR(1) Model with Seasonal Mean Heterogeneity (quarterly)

	Statistical GM:	$\mathbf{Z}_t = \mathbf{a}_0 + \boldsymbol{\delta}_1 cos\left(\frac{\pi t}{2}\right) + \boldsymbol{\delta}_2 sin\left(\frac{\pi t}{2}\right) + \boldsymbol{\delta}_3 cos\left(\pi t\right) + \mathbf{A}_1^\top \mathbf{Z}_{t-1} + \mathbf{u}_t,\ \ t\in\mathbb{N},$
[1]	Normal:	$D(\mathbf{Z}_t \vert \mathbf{Z}_{t-1};\boldsymbol{\theta})$, is Normally distributed,
[2]	Linearity:	$E(\mathbf{Z}_t \vert \sigma(\mathbf{Z}_{t-1})) = \mathbf{a}_0 + \boldsymbol{\delta}_1 cos\left(\frac{\pi t}{2}\right) + \boldsymbol{\delta}_2 sin\left(\frac{\pi t}{2}\right) + \boldsymbol{\delta}_3 cos\left(\pi t\right) + \mathbf{A}_1^\top \mathbf{Z}_{t-1},$
[3]	Homoskedasticity:	$Var(\mathbf{Z}_t \vert \sigma(\mathbf{Z}_{t-1})) = \boldsymbol{\Omega},$
[4]	Markov(p):	$\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$ is a Markov(1) process,
[5]	t-invariance:	$\boldsymbol{\theta}$:=$(\mathbf{a}_0, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3, \mathbf{A}_1, \boldsymbol{\Omega})$ are t-invariant.
		$\mathbf{a}_0 = (\mathbf{I}_2 - \mathbf{A}_1^\top)\boldsymbol{\mu}_0,\ \boldsymbol{\delta}_1 = (\boldsymbol{\gamma}_1 + \mathbf{A}_1^\top \boldsymbol{\psi}_1),\ \boldsymbol{\delta}_2 = (\boldsymbol{\psi}_1 + \mathbf{A}_1^\top \boldsymbol{\gamma}_1),$
		$\boldsymbol{\delta}_3 = (\boldsymbol{\gamma}_2 + \mathbf{A}_1^\top \boldsymbol{\gamma}_2),\ \mathbf{A}_1^\top = \boldsymbol{\Sigma}_{01}\boldsymbol{\Sigma}_{11}^{-1},\ \boldsymbol{\Omega} = \boldsymbol{\Sigma}_{00} - \boldsymbol{\Sigma}_{01}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{01}^\top$

2.3 Model Mis-specification in the VAR(1)

Although the model specifications in Tables 2.2, 2.3, and 2.4 are quite similar, neglecting the specific trends inherent in the data can be catastrophic. As shown in Spanos and McGuirk (2001), in the context of the Linear Regression model, the omission of key, statistical information can lead to erroneous inferences. For example, they found that model mis-specification can lead to major differences between the actual and nominal error probabilities. These findings were supported in McGuirk and Spanos (2009), where it was shown that the method of 'error-fixing' can produce not only biased and inconsistent estimates, but also can cause misleading test results. To highlight the impact of model mis-specification on the reliability of inference in the context of vector autoregressive models, even the instance where only one of the model assumptions is invalid, consider the following VAR(1) simulation.

2.4 Simulation Design

The simulation design in this chapter is similar to that found in Spanos and McGuirk (2001), Spanos (2009), and Spanos (2019a). For this simulation, assume that the joint distribution given in equation (2.3) takes the form:

$$\begin{bmatrix} \mathbf{Y}_t \\ \mathbf{Y}_{t-1} \end{bmatrix} \sim N\left(\begin{bmatrix} 3.0 \\ 6.0 \\ 3.0 \\ 6.0 \end{bmatrix}, \begin{bmatrix} 1.3 & 0.7 & 0.7 & 0.5 \\ 0.7 & 1.8 & 0.4 & 0.6 \\ 0.7 & 0.4 & 1.3 & 0.7 \\ 0.5 & 0.6 & 0.7 & 1.8 \end{bmatrix} \right)$$

Then the parameterization for the VAR(1) is given as:

$$\mathbf{A}_1^\top = \begin{pmatrix} 0.7 & 0.5 \\ 0.4 & 0.6 \end{pmatrix} \begin{pmatrix} 1.3 & 0.7 \\ 0.7 & 1.8 \end{pmatrix}^{-1} = \begin{pmatrix} 0.4919 & 0.0865 \\ 0.1622 & 0.2703 \end{pmatrix}$$

$$\mathbf{a}_0 = (\mathbf{I}_2 - \mathbf{A}_1^\top)\boldsymbol{\mu} = \begin{pmatrix} 1.01 \\ 3.89 \end{pmatrix}$$

$$E(\mathbf{Y}_t|\mathbf{Y}_{t-1}) = \mathbf{a}_0 + \mathbf{A}_1^\top \mathbf{Y}_{t-1} = \begin{pmatrix} 1.01 \\ 3.89 \end{pmatrix} + \begin{pmatrix} 0.4919 & 0.0865 \\ 0.1622 & 0.2703 \end{pmatrix} \mathbf{Y}_{t-1}.$$

$$\boldsymbol{\Omega} = \begin{pmatrix} 1.3 & 0.7 \\ 0.7 & 1.8 \end{pmatrix} - \begin{pmatrix} 0.7 & 0.5 \\ 0.4 & 0.6 \end{pmatrix} \begin{pmatrix} 1.3 & 0.7 \\ 0.7 & 1.8 \end{pmatrix}^{-1} \begin{pmatrix} 0.7 & 0.4 \\ 0.5 & 0.6 \end{pmatrix} = \begin{pmatrix} 0.912 & 0.4514 \\ 0.4514 & 1.573 \end{pmatrix}$$

The statistical GM is therefore:

$$\mathbf{Y}_t = \mathbf{a}_0 + \mathbf{A}_1^\top \mathbf{Y}_{t-1} + \mathbf{u}_t, \mathbf{u}_t \sim NMd(\mathbf{0}, \boldsymbol{\Omega}).$$

Hence, the steps to simulate from $\mathbf{Y}_t$ are:

1. Specify the true parameter values (done above).
2. Generate the observation $\mathbf{Y}_1$, where $\mathbf{Y}_1 \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{00})$.
3. Draw one observation from $D(\mathbf{Y}_2|\mathbf{Y}_1) \sim N(\mathbf{a}_0 + \mathbf{A}_1^\top \mathbf{Y}_1, \boldsymbol{\Omega})$.
4. Repeat the third step n times, recursively substituting in each new value of $\mathbf{Y}_t$ as $\mathbf{Y}_{t-1}$ in the next step, for $t = 3, \ldots, n$.

Steps 1-4 are then repeated R times, thus creating R replications. For each replication, the VAR(1) model is estimated, and the estimated parameters and corresponding standard errors are recorded to calculate the corresponding sampling distributions. Error probabilities and power are then calculated for the following set of hypotheses:

$$\mathrm{H}_0 : \beta_i = \beta_i^* \quad \text{vs.} \quad \mathrm{H}_1 : \beta_i \neq \beta_i^*$$

In this case, the test statistics:

$$\boldsymbol{\tau}_0(\mathbf{y}_i) = \frac{\hat{\beta}_i - \beta_i^*}{\sqrt{Var(\hat{\beta}_i)}} \overset{\beta_i^*}{\sim} St(n-k-1), \;\; \boldsymbol{\tau}_1(\mathbf{y}_i) = \frac{\hat{\beta}_i - \beta_i^*}{\sqrt{Var(\hat{\beta}_i)}} \overset{\beta_\mathrm{i}^\mathrm{a}}{\sim} St(\delta; n-k-1),$$

where $\delta = \frac{\beta_i^* - \beta_i^a}{\sqrt{Var(\hat{\beta}_i)}}$ is the non-centrality parameter and k is the number of predictors, are evaluated for each data realization $\mathbf{y}_i, i = 1\ldots, R$.[2] Note that the only reason that it is possible to calculate the error probabilities is because the actual values for the model parameters are known; they were assumed from the outset.

2.5 Experiment 1: VAR(1) With Trending (Linear) Mean

The first mis-specification to be explored is that of a neglected mean heterogeneity, i.e. assumption [5] in Table 2.2 is invalid, and the correct model specification is that found in Table 2.3. It is assumed in this case that the true model contains mean heterogeneity in the form of a linear trend. Let

$$\begin{bmatrix} \mathbf{Y}_t \\ \mathbf{Y}_{t-1} \end{bmatrix} \sim N\left(\begin{bmatrix} 3.0 + 0.5t \\ 6.0 + 0.25t \\ 3.0 + 0.5(t-1) \\ 6.0 + 0.25(t-1) \end{bmatrix}, \begin{bmatrix} 1.3 & 0.7 & 0.7 & 0.5 \\ 0.7 & 1.8 & 0.4 & 0.6 \\ 0.7 & 0.4 & 1.3 & 0.7 \\ 0.5 & 0.6 & 0.7 & 1.8 \end{bmatrix} \right)$$

Then the parameterization for the VAR(1) is given as:

$$\mathbf{A}_1^\top = \begin{pmatrix} 0.7 & 0.5 \\ 0.4 & 0.6 \end{pmatrix} \begin{pmatrix} 1.3 & 0.7 \\ 0.7 & 1.8 \end{pmatrix}^{-1} = \begin{pmatrix} 0.4919 & 0.0865 \\ 0.1622 & 0.2703 \end{pmatrix}$$

[2] The estimated standard error is used in place of $\sqrt{Var(\hat{\beta}_i)}$.

$$\mathbf{a}_0 = (\mathbf{I}_2 - \mathbf{A}_1^\top)\boldsymbol{\mu}_0 + \mathbf{A}_1\boldsymbol{\mu}_1,$$

$$\boldsymbol{\delta} = (\mathbf{I}_2 - \mathbf{A}_1^\top)\boldsymbol{\mu}_1$$

$$\boldsymbol{\Omega} = \begin{pmatrix} 1.3 & 0.7 \\ 0.7 & 1.8 \end{pmatrix} - \begin{pmatrix} 0.7 & 0.5 \\ 0.4 & 0.6 \end{pmatrix} \begin{pmatrix} 1.3 & 0.7 \\ 0.7 & 1.8 \end{pmatrix}^{-1} \begin{pmatrix} 0.7 & 0.4 \\ 0.5 & 0.6 \end{pmatrix} = \begin{pmatrix} 0.912 & 0.4514 \\ 0.4514 & 1.573 \end{pmatrix}$$

where $\boldsymbol{\mu}_0 = [3, 6]$ and $\boldsymbol{\mu}_1 = [0.5, 0.25]$. Therefore:

$$E(\mathbf{Y}_t|\mathbf{Y}_{t-1}) = \boldsymbol{\alpha}_0 + \boldsymbol{\delta}t + \mathbf{A}_1^\top\mathbf{Y}_{t-1} = \begin{pmatrix} 1.27 \\ 4.04 \end{pmatrix} + \begin{pmatrix} 0.232 \\ 0.101 \end{pmatrix} t + \begin{pmatrix} 0.4919 & 0.0865 \\ 0.1622 & 0.2703 \end{pmatrix} \mathbf{Y}_{t-1} \quad (2.12)$$

The statistical GM is thus:

$$\mathbf{Y}_t = \mathbf{a}_0 + \boldsymbol{\delta}t + \mathbf{A}_1^\top\mathbf{Y}_{t-1} + \mathbf{u}_t, \quad (2.13)$$

where $\mathbf{u}_t \sim NMd(\mathbf{0}, \boldsymbol{\Omega})$, $t \in \mathbb{N}$. The simulation procedure is identical to that in the case with no trend, with the exception that steps 2 and 3 now have an additional term:

1. Specify the true parameter values (done above).
2. Generate the observation $\mathbf{Y}_1$, where $\mathbf{Y}_1 \sim N(\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1(1), \boldsymbol{\Sigma}_{00})$.
3. Draw one observation from $D(\mathbf{Y}_2|\mathbf{Y}_1) \sim N(\mathbf{a}_0 + \boldsymbol{\delta}(2) + \mathbf{A}_1^\top\mathbf{Y}_1, \boldsymbol{\Omega})$.
4. Repeat the third step n times, recursively substituting in each new value of $\mathbf{Y}_t$ as $\mathbf{Y}_{t-1}$ in the next step, for $t = 3, \ldots, n$.

2.6 Experiment 2: VAR(1) with Seasonality

The second case explored in this simulation is when there is seasonality that is unaccounted for in the model. Hence, the true model is that depicted in Table 2.4. Let

$$\begin{bmatrix} \mathbf{Y}_t \\ \mathbf{Y}_{t-1} \end{bmatrix} \sim N\left(\begin{bmatrix} 3.0 + \mu_1(t) \\ 6.0 + \mu_2(t) \\ 3.0 + \mu_1(t-1) \\ 6.0 + \mu_2(t-1) \end{bmatrix}, \begin{bmatrix} 1.3 & 0.7 & 0.7 & 0.5 \\ 0.7 & 1.8 & 0.4 & 0.6 \\ 0.7 & 0.4 & 1.3 & 0.7 \\ 0.5 & 0.6 & 0.7 & 1.8 \end{bmatrix} \right),$$

where:

$$\mu_{1,i}(t) = \sum_{k=1}^{S/2} \gamma_{i,k} cos(\frac{2\pi kt}{S}) + \psi_{i,k} sin(\frac{2\pi kt}{S}), i = 1, 2$$

In this case, let $S = 4$. Given this, let the seasonal parameters be:

$\boldsymbol{\gamma}_1 = [\gamma_{1,1}, \gamma_{1,2}] = [0.90, 0.25]$

$\boldsymbol{\psi}_1 = [\psi_{1,1}, \psi_{1,2}] = [0.20, 0.85]$

$\boldsymbol{\gamma}_2 = [\gamma_{2,1}, \gamma_{2,2}] = [0.40, 0.65]$

$\boldsymbol{\psi}_2 = [\psi_{1,1}, \psi_{1,2}] = [0.75, 0.45]$

and let $\boldsymbol{\mu}_0 = [3, 6]$ (given above). Then the parameterization for the VAR(1) is given as:

$$\mathbf{A}_1^\top = \begin{pmatrix} 0.7 & 0.5 \\ 0.4 & 0.6 \end{pmatrix} \begin{pmatrix} 1.3 & 0.7 \\ 0.7 & 1.8 \end{pmatrix}^{-1} = \begin{pmatrix} 0.4919 & 0.0865 \\ 0.1622 & 0.2703 \end{pmatrix}$$

$$\mathbf{\Omega} = \begin{pmatrix} 1.3 & 0.7 \\ 0.7 & 1.8 \end{pmatrix} - \begin{pmatrix} 0.7 & 0.5 \\ 0.4 & 0.6 \end{pmatrix} \begin{pmatrix} 1.3 & 0.7 \\ 0.7 & 1.8 \end{pmatrix}^{-1} \begin{pmatrix} 0.7 & 0.4 \\ 0.5 & 0.6 \end{pmatrix} = \begin{pmatrix} 0.912 & 0.4514 \\ 0.4514 & 1.573 \end{pmatrix}$$

$$E(\mathbf{Y}_t|\mathbf{Y}_{t-1}) = \mathbf{a}_0 + \boldsymbol{\delta}_1 cos\left(\frac{\pi t}{2}\right) + \boldsymbol{\delta}_2 sin\left(\frac{\pi t}{2}\right) + \boldsymbol{\delta}_3 cos\left(\pi t\right) + \mathbf{A}_1^{\top}\mathbf{Y}_{t-1},$$

where the derivation of the constant and trigonometric terms comes from Section 2.2.2 above. The statistical GM is therefore:

$$\begin{aligned}\mathbf{Y}_t &= \mathbf{a}_0 + \boldsymbol{\delta}_1 cos\left(\frac{\pi t}{2}\right) + \boldsymbol{\delta}_2 sin\left(\frac{\pi t}{2}\right) + \boldsymbol{\delta}_3 cos\left(\pi t\right) + \mathbf{A}_1^{\top}\mathbf{Y}_{t-1} + \mathbf{u}_t \\ &= \begin{bmatrix} 1.01 \\ 3.89 \end{bmatrix} + \begin{bmatrix} 1.07 \\ 0.512 \end{bmatrix} cos\left(\frac{\pi t}{2}\right) + \begin{bmatrix} -0.264 \\ 0.636 \end{bmatrix} sin\left(\frac{\pi t}{2}\right) + \begin{bmatrix} 0.653 \\ 0.891 \end{bmatrix} cos\left(\pi t\right) + \mathbf{A}_1^{\top}\mathbf{Y}_{t-1} + \mathbf{u}_t,\end{aligned}$$

where $\mathbf{u}_t \sim NMd(\mathbf{0}, \mathbf{\Omega})$, $t \in \mathbb{N}$. As in the case of the linear trend, the simulation procedure is identical to the original simulation procedure, but now contains the new model parameters:

1. Specify the true parameter values (done above).

2. Generate the observation $\mathbf{Y}_1$, where $\mathbf{Y}_1 \sim N(\boldsymbol{\mu}_0 + \boldsymbol{\gamma}_1 cos\left(\frac{\pi(1)}{2}\right) + \boldsymbol{\psi}_1 sin\left(\frac{\pi(1)}{2}\right) + \boldsymbol{\gamma}_2 cos\left(\pi(1)\right), \mathbf{\Sigma}_{00})$.

3. Draw one observation from

$$D(\mathbf{Y}_2|\mathbf{Y}_1) \sim N(\mathbf{a}_0 + \boldsymbol{\delta}_1 cos\left(\frac{\pi(2)}{2}\right) + \boldsymbol{\delta}_2 sin\left(\frac{\pi(2)}{2}\right) + \boldsymbol{\delta}_3 cos\left(\pi(2)\right) + \mathbf{A}_1^{\top}\mathbf{Y}_1, \mathbf{\Omega}).$$

4. Repeat step three n times, recursively substituting in each new value of $\mathbf{Y}_t$ as $\mathbf{Y}_{t-1}$ in the next step, for $t = 3, \ldots, n$.

2.7 Simulation Results

2.7.1 Experiment 1 Results

The estimated sampling distributions and Type I error probabilities for Experiment 1 are given in Table 2.5 below. Note that the empirical sampling distributions of the $R = 20000$ replication estimates given in Table 2.5 are computed using simple averages.

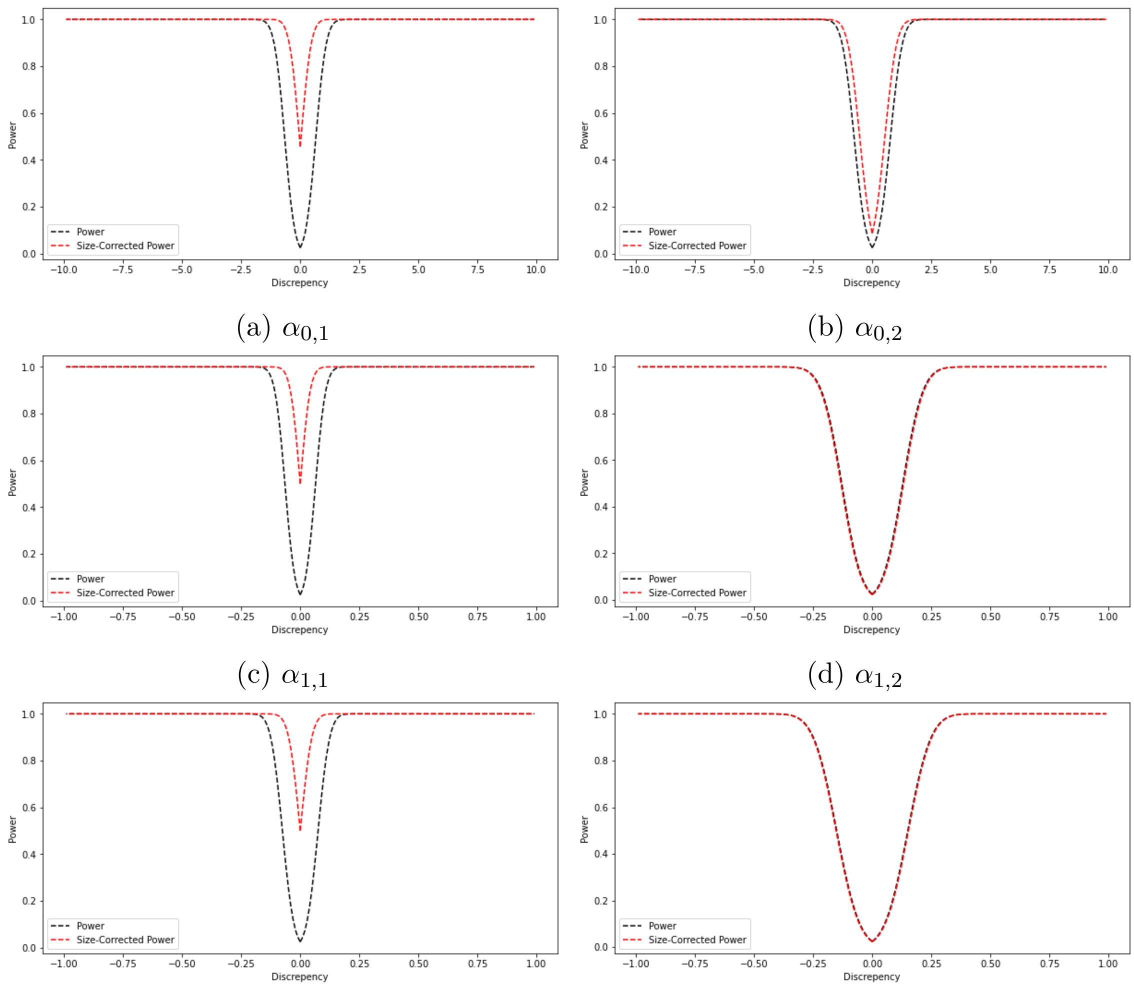
Table 2.5: Simulation Results Comparison VAR(1) Model with Linear Trend

R = 20,000	Adequate Model:				Mis-specified Model:			
	True: $\mathbf{Z}_t = \boldsymbol{\alpha}_0 + \mathbf{A}_1^\intercal \mathbf{Z}_{t-1} + \mathbf{u}_t$				True: $\mathbf{Z}_t = \boldsymbol{\alpha}_0 + \boldsymbol{\delta} t + \mathbf{A}_1^\intercal \mathbf{Z}_{t-1} + \mathbf{u}_t$			
	Estim: $\mathbf{Z}_t = \boldsymbol{\alpha}_0 + \mathbf{A}_1^\intercal \mathbf{Z}_{t-1} + \mathbf{u}_t$				Estim: $\mathbf{Z}_t = \boldsymbol{\alpha}_0 + \mathbf{A}_1^\intercal \mathbf{Z}_{t-1} + \mathbf{u}_t$			
	n = 100		n = 200		n = 100		n = 200	
Parameters	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$\hat{\alpha}_{0,1}[\alpha_{0,1} = 1.01]$	1.092	0.4622	1.046	0.3186	0.2569	0.4085	0.2170	0.2806
$\hat{\alpha}_{0,2}[\alpha_{0,2} = 3.89]$	4.037	0.5934	3.963	0.4183	3.726	0.5347	3.643	0.3723
$\hat{\alpha}_{1,1}[\alpha_{1,1} = 0.492]$	0.4636	0.0979	0.4784	0.0677	0.9608	0.0450	0.9635	0.0311
$\hat{\alpha}_{1,2}[\alpha_{1,2} = 0.0865]$	0.0864	0.0831	0.0865	0.0575	0.0725	0.0899	0.0716	0.0621
$\hat{\alpha}_{2,1}[\alpha_{2,1} = 0.1622]$	0.1568	0.1304	0.1595	0.0891	0.3806	0.0528	0.3748	0.0370
$\hat{\alpha}_{2,2}[\alpha_{2,2} = 0.2703]$	0.2486	0.1058	0.2596	0.0751	0.2359	0.1042	0.2498	0.0739
t-statistics	Mean	$(\alpha = 0.05)$	Mean	$(\alpha = 0.05)$	Mean	$(\alpha = 0.05)$	Mean	$(\alpha = 0.05)$
$\tau = \frac{\lvert\hat{\alpha}_{0,1} - \alpha_{0,1}\rvert}{\sqrt{Var(\hat{\alpha}_{0,1})}}$	0.8162	0.0525	0.8047	0.0493	2.095	0.5487	3.138	0.9140
$\tau = \frac{\lvert\hat{\alpha}_{0,2} - \alpha_{0,2}\rvert}{\sqrt{Var(\hat{\alpha}_{0,2})}}$	0.8056	0.0515	0.8100	0.0511	0.9158	0.0812	1.184	0.1737
$\tau = \frac{\lvert\hat{\alpha}_{1,1} - \alpha_{1,1}\rvert}{\sqrt{Var(\hat{\alpha}_{1,1})}}$	0.8221	0.0547	0.8075	0.0515	9.872	1.000	14.35	1.000
$\tau = \frac{\lvert\hat{\alpha}_{1,2} - \alpha_{1,2}\rvert}{\sqrt{Var(\hat{\alpha}_{1,2})}}$	0.8042	0.0521	0.7982	0.0514	0.7651	0.0401	0.7722	0.0442
$\tau = \frac{\lvert\hat{\alpha}_{2,1} - \alpha_{2,1}\rvert}{\sqrt{Var(\hat{\alpha}_{2,1})}}$	0.8127	0.0546	0.7975	0.0506	3.926	0.9880	5.532	1.000
$\tau = \frac{\lvert\hat{\alpha}_{2,2} - \alpha_{2,2}\rvert}{\sqrt{Var(\hat{\alpha}_{2,2})}}$	0.8053	0.0501	0.8083	0.0500	0.7951	0.0460	0.7988	0.0460

A quick glance at the results in Table 2.5 reveals how devastating mis-specification can be for vector autoregressive models. In the case of the statistically adequate model, the estimated means of the model parameters are excellent approximations for the true values found in brackets. Increasing the number of observations in the simulated data set further improves the accuracy of the estimated means. Moreover, the actual Type I error probabilities for the t-tests, when the estimated model is statistically adequate, approximate the the nominal error probability ($\alpha = 0.05$). Once again, increasing the number of observations to $n = 200$ enhances the accuracy.

Failure to account for mean heterogeneity, as shown under the "Mis-specified Model" in Table 2.5, causes the estimated means of the model parameter to be highly inaccurate, and the inaccuracy is further exacerbated when the number of observations increases. This is indicative of inconsistent estimators. Unlike the statistically adequate model, the actual Type I error probabilities no longer provide an accurate approximation for the nominal error probability. The discrepancies between the actual and nominal error probability become larger as the sample size increases to $n = 200$.

Given the relationship between the Type I and Type II error probabilities, it is clear that if the actual and nominal Type I error probabilities differ, so too will the actual and nominal Type II error probabilities. Thus, a statistically inadequate model will lead to not only incorrect error probabilities, but also will affect the power of the test to detect discrepancies from the null hypothesis. To emphasize the effect that model mis-specifcation has on the power of the t-test, see Figure 1. In the case of the statistically inadequate model, one needs to use the *size adjusted* power (McGuirk and Spanos, 2009), evaluated using the actual Type I error probability. Figure 1 shows the difference between the power and size-corrected power of the t-test in the case where a trend has been neglected, when $n = 200$. It is immediately apparent that disparities exist between the nominal power, which is evaluated at ($\alpha = 0.05$) and the size-corrected power. Thus, the power of the test is distorted, and any results stemming from the t-test are untrustworthy.



(a) $\alpha_{0,1}$ (b) $\alpha_{0,2}$

(c) $\alpha_{1,1}$ (d) $\alpha_{1,2}$

(e) $\alpha_{2,1}$ (f) $\alpha_{2,2}$

Figure 2.1: Power and Size-Corrected Power Curves: Linear Trend $(n = 200, \nu = 196)$

2.7.2 Experiment 2 Results

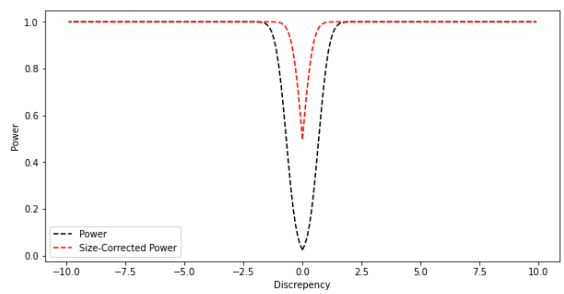
The estimated sampling distributions and actual Type I error probabilities for the t-test, in the case where seasonality is present in the mean, is given in Table 2.6. Note that the statistically adequate model is identical to the one described in Table 2.5.

Table 2.6: Simulation Results Comparison VAR(1) Model with Seasonality

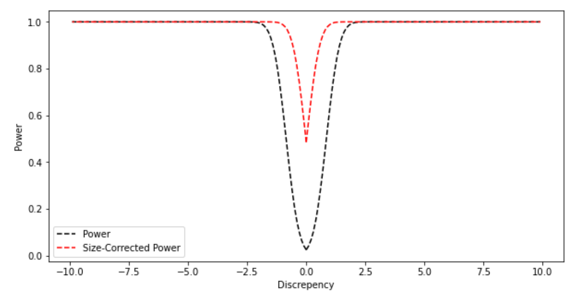
R = 20,000	Adequate Model:				Mis-specified Model:			
	True: $\mathbf{Z}_t = \boldsymbol{\alpha}_0 + \mathbf{A}_1^{\intercal}\mathbf{Z}_{t-1} + \mathbf{u}_t$				True: $\mathbf{Z}_t = \boldsymbol{\alpha}_0 + \boldsymbol{\delta}_1 cos(\frac{\pi t}{2}) + \boldsymbol{\delta}_2 sin(\frac{\pi t}{2}) + \boldsymbol{\delta}_3 cos(\pi t) + \mathbf{A}_1^{\intercal}\mathbf{Z}_{t-1} + \mathbf{u}_t$			
	Estim: $\mathbf{Z}_t = \boldsymbol{\alpha}_0 + \mathbf{A}_1^{\intercal}\mathbf{Z}_{t-1} + \mathbf{u}_t$				Estim: $\mathbf{Z}_t = \boldsymbol{\alpha}_0 + \mathbf{A}_1^{\intercal}\mathbf{Z}_{t-1} + \mathbf{u}_t$			
	n = 100		n = 200		n = 100		n = 200	
Parameters	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$\hat{\alpha}_{0,1}[\alpha_{0,1} = 1.01]$	1.092	0.4622	1.046	0.3186	3.235	0.5122	3.176	0.3616
$\hat{\alpha}_{0,2}[\alpha_{0,2} = 3.89]$	4.037	0.5934	3.963	0.4183	5.607	0.6065	5.546	0.4320
$\hat{\alpha}_{1,1}[\alpha_{1,1} = 0.492]$	0.4636	0.0979	0.4784	0.0677	0.4136	0.1025	0.423	0.0724
$\hat{\alpha}_{1,2}[\alpha_{1,2} = 0.0865]$	0.0864	0.0831	0.0865	0.0575	-0.2451	0.0862	-0.2405	0.0606
$\hat{\alpha}_{2,1}[\alpha_{2,1} = 0.1622]$	0.1568	0.1304	0.1595	0.0891	0.3090	0.1326	0.3073	0.0915
$\hat{\alpha}_{2,2}[\alpha_{2,2} = 0.2703]$	0.2486	0.1058	0.2596	0.0751	-0.0892	0.1136	-0.0781	0.0808
t-statistics	Mean	$(\alpha = 0.05)$	Mean	$(\alpha = 0.05)$	Mean	$(\alpha = 0.05)$	Mean	$(\alpha = 0.05)$
$\tau = \frac{\lvert\hat{\alpha}_{0,1} - \alpha_{0,1}\rvert}{\sqrt{Var(\hat{\alpha}_{0,1})}}$	0.8162	0.0525	0.8047	0.0493	4.480	0.9945	6.209	1.000
$\tau = \frac{\lvert\hat{\alpha}_{0,2} - \alpha_{0,2}\rvert}{\sqrt{Var(\hat{\alpha}_{0,2})}}$	0.8056	0.0515	0.8100	0.0511	2.773	0.7914	3.820	0.9718
$\tau = \frac{\lvert\hat{\alpha}_{1,1} - \alpha_{1,1}\rvert}{\sqrt{Var(\hat{\alpha}_{1,1})}}$	0.8221	0.0547	0.8075	0.0515	0.9364	0.0779	1.052	0.1190
$\tau = \frac{\lvert\hat{\alpha}_{1,2} - \alpha_{1,2}\rvert}{\sqrt{Var(\hat{\alpha}_{1,2})}}$	0.8042	0.0521	0.7982	0.0514	3.563	0.9488	4.989	0.9990
$\tau = \frac{\lvert\hat{\alpha}_{2,1} - \alpha_{2,1}\rvert}{\sqrt{Var(\hat{\alpha}_{2,1})}}$	0.8127	0.0546	0.7975	0.0506	1.196	0.1689	1.557	0.3120
$\tau = \frac{\lvert\hat{\alpha}_{2,2} - \alpha_{2,2}\rvert}{\sqrt{Var(\hat{\alpha}_{2,2})}}$	0.8053	0.0501	0.8083	0.0500	3.109	0.8729	4.292	0.9893

In the same spirit as Experiment 1, Experiment 2 shows that neglecting seasonality, and thus violating assumption [5] in Table 2.4, will cause the mean of the estimated parameters to be extremely inaccurate. Moreover, as was the case in Experiment 1, the estimated means become even more inaccurate when the number of observations increases to $n = 200$, which is indicative that the estimators of the model parameters are inconsistent.

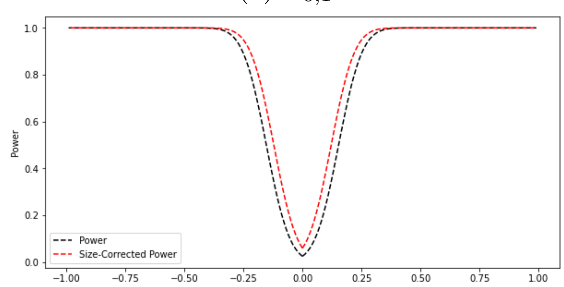
Similarly, the actual Type I error probabilities are not remotely close to the nominal ($\alpha =$ 0.05), and in most cases, the actual error probability approaches one as the number of observations increases. Thus, the actual and nominal Type I error probabilities are different in the presence of neglected seasonality, and this problem remains as the number of observations increases. Since the actual Type I error probabilities are erroneous, the actual Type II error probabilities will also be poor approximations of the nominal Type II error probabilities. Moreover, power of the test will be highly distorted, as shown in Figure 2.2. Note that the actual power of the test is the size-corrected power, which uses the actual Type I error probability. Given the extreme differences between the the actual, size-corrected power, and the nominal power, Figure 2.2 once again demonstrates the unreliability of inference in the presence of model mis-specification.



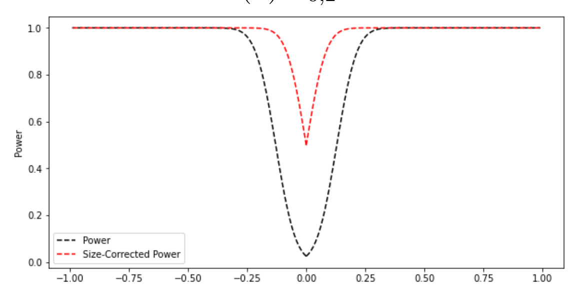
(a) $\alpha_{0,1}$



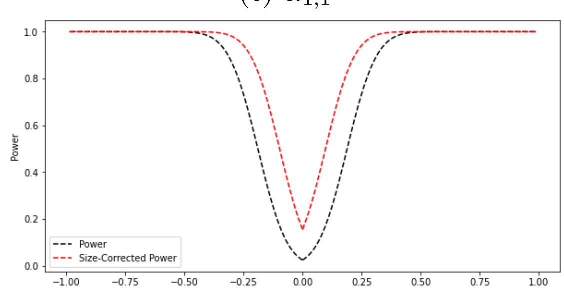
(b) $\alpha_{0,2}$



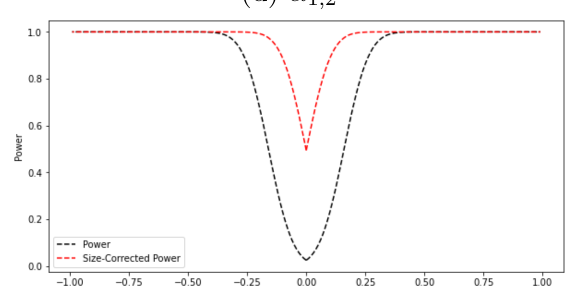
(c) $\alpha_{1,1}$



(d) $\alpha_{1,2}$



(e) $\alpha_{2,1}$



(f) $\alpha_{2,2}$

Figure 2.2: Power and Size-Corrected Power: Seasonality ($n = 200, \nu = 196$)

2.8 Discussion

In conclusion, the main purpose of this chapter is to motivate what disasters can occur when inference is performed on a statistically mis-specified model. The cornerstone of frequentist statistics is the underlying statistical model, $\mathcal{M}_\theta(\mathbf{x})$, which contains a set of internally consistent and testable assumptions (Spanos, 2019a). When assumptions are erroneously imposed on the data, the consequences can be severe. As noted in Spanos (2019a), the only way to avert the problem in its entirety is to first verify that the model assumptions are valid for the data. Hence, one must first establish that the statistical model is statistically adequate, i.e., $\mathcal{M}_\theta(\mathbf{x})$ must take into account all of the chance regularity patterns in the data. This can be accomplished through rigorous model mis-specification testing and respecification (Spanos, 2010a).

Ultimately, the results of the experiments performed above are consistent with those in the earlier work of McGuirk and Spanos (2009), Spanos (2009) and Spanos and McGuirk (2001). Violating the underlying model assumptions, in this case assumption [5] in Table 2.2, will destroy any chance at achieving reliable and trustworthy inferences. This was alarmingly transparent in the experiments, where not only the estimated sampling distributions were highly inaccurate for the model parameters, but the actual error probabilities were extremely poor approximations of the nominal error probabilities. Furthermore, the problems demonstrated above do not disappear as the number of observations increases; if anything, the estimates deviate farther from the true values. It is important to note that the only reason that such an analysis could be performed is because the true parameter values were known from the outset. In practice, the true parameter values and error probabilities are unknown, implying that in reality, it is impossible to capture the differences between the actual and nominal error probabilities. Hence, the only way to avoid such scenarios is to ensure that the all of the statistical model adequately captures all of the features in the data.

Another point of consideration is that these simulations only explore the violation of assumption [5] in Table 2.2, namely the assumption that the model parameters are invariant to changes in time (the index). However, this is not the only potential source of model mis-specification. Any deviation from assumptions [1]-[5] will result in undependable inferences; see Spanos (2009) for the example of Markov dependence in the context of the Linear Regression model and McGuirk and Spanos (2009), Spanos and McGuirk (2001) for more examples.

Chapter 3

Macroeconomic Forecasting using Temporal Principal Components

3.1 Introduction

Over the last few decades, the number of papers exploring PCA has blossomed into a variety of directions. One such direction is the estimation of factor models to nowcast and forecast key macroeconomic data series, which stems from Sargent and Sims (1977) and Geweke (1977). Moreover, the idea of using indices to capture common behavior between correlated macroeconomic variables was suggested as early as Burns and Mitchell (1946). The factor models used today typically utilize principal components as consistent estimators of the latent factors; see Stock and Watson (2002a,b), Giannone et al. (2008), Stock and Watson (2016).

The aim of this chapter is to extend the use of PCA to instances where traditional principal components fail to account for all of the systematic information making up common macroeconomic and financial indicators. It was shown in Chapter 2 that statistical mis-specification, or neglecting important systematic information, can lead to erroneous inferences, even in the trivial case of omitted trends and seasonality. This became incredibly apparent through the differences in actual and nominal error probabilities. In the case of traditional principal components, as defined in Chapter 1, the stochastic process was assumed to independent and identically distributed. When these probabilistic assumptions are incorrect, however, the estimated principal components will be mis-specified, and any forecasts produced using them will be misleading. Thus, temporal principal components are defined to capture the dependence inherent in macroeconomic time series. These temporal principal components are then used to augment forecasts from statistically adequate autoregressive models. Ultimately, a baseline framework is established for not only achieving superior forecasts in terms of accuracy, but also in terms of capturing all of the probabilistic assumptions inherent in each macroeconomic data series.

3.2 Markov Dependence and the Normal Vector Autoregressive Model

Having established the framework for the traditional perspective on principal components in Chapter 1, as well as the relationship between principal components and the Multivariate Normal model, the next step is to replace the independence assumption in the Multivariate Normal model with Markov dependence and the identically distributed assumption with the assumption that the stochastic process is second-order stationary. To begin, let $\{\mathbf{Z}_t, t \in \mathbb{N}\}$ be a vector stochastic process, where $\mathbf{Z}_t$ is of length $(k \times 1)$. The notation in this setup is similar to that used in Chapter 1. Here, Markov dependence of order p and second-order stationary are defined as follows:

Definition 3.1 (Markov Dependence(p))**.** Let $\{\mathbf{Z}_t, t \in \mathbb{N}\}$ be a stochastic process. The process $\{\mathbf{Z}_t, t \in \mathbb{N}\}$ is Markov dependent of order p, $p \geq 1$, if $D_t(\mathbf{Z}_t|\mathbf{Z}_{t-1}, \ldots, \mathbf{Z}_1; \boldsymbol{\psi}_t) = D_t(\mathbf{Z}_t|\mathbf{Z}_{t-1}, \ldots, \mathbf{Z}_{t-p}, \boldsymbol{\theta}_t)$. In other words, given $\mathbf{Z}_{t-1}, \ldots, \mathbf{Z}_{t-p}$, $\mathbf{Z}_t$ is *conditionally independent* of $\mathbf{Z}_{t-p+1}, \ldots, \mathbf{Z}_1$ (Spanos, 2019a).

Definition 3.2 (Second-Order Stationary)**.** Let $\{\mathbf{Z}_t, t \in \mathbb{N}\}$ be a vector stochastic process, where $\mathbf{Z}_t$ is of length $(k \times 1)$. The process $\{\mathbf{Z}_t, t \in \mathbb{N}\}$ is second-order, or covariance, stationary if $E(\mathbf{Z}_t) = \boldsymbol{\mu}$ and $Cov(\mathbf{Z}_t, \mathbf{Z}_{t-p}) = E((\mathbf{Z}_t - \boldsymbol{\mu})(\mathbf{Z}_{t-p} - \boldsymbol{\mu})^\top) = \boldsymbol{\Sigma}_{ij}$, for $i, j = 0, 1, \ldots, p$.

This alteration in the probabilistic assumptions results in the Normal Vector Autoregressive (VAR) model, which is crucial in the establishment and specification of temporal principal components. As done for the Multivariate Normal model, this section will utilize the probabilistic reduction approach in Spanos and McGuirk (2001) and Spanos (2019a) to provide a complete specification for the VAR(p) model, including the statistical generating mechanism as well as the model assumptions and parameterizations. Here, the stochastic process $\{\mathbf{Z}_t, t \in \mathbb{N}\}$ is assumed to be a (i) multivariate normal, (ii) Markov dependent of order p, and (iii) Covariance Stationary process.

Given the assumptions above, the joint distribution of the sample, $D(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$, can be reduced as follows:

$$\begin{aligned} D(\mathbf{Z}_1, \ldots, \mathbf{Z}_n; \boldsymbol{\psi}) &\overset{\text{M}}{=} D(\mathbf{Z}_1^p; \boldsymbol{\phi}_0) \prod_{t=p+1}^{n} D_t(\mathbf{Z}_t | \mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}_t) \\ &\overset{\text{M/S}}{=} D(\mathbf{Z}_1^p; \boldsymbol{\phi}_0) \prod_{t=p+1}^{n} D(\mathbf{Z}_t | \mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}), \end{aligned} \tag{3.1}$$

where $\mathbf{Z}_1^p := (\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_p)$ and $\mathbf{Z}_{t-1}^{t-p} := (\mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \ldots, \mathbf{Z}_{t-p})$. Note that the second line of the reduction comes from assumption (ii) Markov Dependence, and the third line comes from assumption (iii) Covariance Stationary. As done in previously, the first p observations are ignored, which will result in the approximate form of the log-likelihood function; see Appendix G for details on the exact form of the log-likelihood function. Under the assumption that $D(\mathbf{Z}_t|\mathbf{Z}_{t-1}^{t-p};\boldsymbol{\phi})$ is normal, this implies that distribution of $D(\mathbf{Z}_t|\mathbf{Z}_{t-1}^{t-p};\boldsymbol{\phi})$ takes the form:

$$\mathbf{Y}_t = \begin{bmatrix} \mathbf{Z}_t \\ \mathbf{Z}_{t-1} \\ \vdots \\ \mathbf{Z}_{t-p} \end{bmatrix} \sim N\left(\boldsymbol{\mu}_y = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{bmatrix}, \boldsymbol{\Sigma}_y = \begin{bmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01} & \boldsymbol{\Sigma}_{02} & \dots & \boldsymbol{\Sigma}_{0p} \\ \boldsymbol{\Sigma}_{10}^\top & \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \dots & \boldsymbol{\Sigma}_{1p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{p0}^\top & \boldsymbol{\Sigma}_{p1}^\top & \boldsymbol{\Sigma}_{p2}^\top & \dots & \boldsymbol{\Sigma}_{pp} \end{bmatrix} \right), \tag{3.2}$$

where $\boldsymbol{\mu}$ is the mean vector of $\mathbf{Z}_t$ and is of vector of length $((p+1)k \times 1)$, $\boldsymbol{\Sigma}_{ij}$ is a $(k \times k)$ covariance matrix for $i, j = 0, 1, \dots, p$. To specify the VAR(p) model directly from this joint distribution, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are first partitioned as follows (Poudyal and Spanos, 2022):

$$\begin{bmatrix} \mathbf{Z}_t, (k \times 1) \\ \mathbf{Z}_{t-1}^{t-p}, (pk \times 1) \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}, (k \times 1) \\ \boldsymbol{\mu}_{pk}, (pk \times 1) \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{00}, (k \times k) & \boldsymbol{\Sigma}_{01}, (k \times pk) \\ \boldsymbol{\Sigma}_{10}^\top, (pk \times k) & \boldsymbol{\Sigma}_{11}, (pk \times pk) \end{bmatrix};$$

Hence, the conditional distribution is derived to be:

$$D(\mathbf{Z}_t|\mathbf{Z}_{t-1}^{t-p};\boldsymbol{\Theta}) \sim N(\mathbf{a}_0 + \mathbf{A}^\top \mathbf{Z}_{t-1}^{t-p}, \boldsymbol{\Omega}), \tag{3.3}$$

where the $\boldsymbol{\Theta} := (\mathbf{a}_0, \mathbf{A}, \boldsymbol{\Omega})$ are:

$$\mathbf{A}^\top = \boldsymbol{\Sigma}_{01}\boldsymbol{\Sigma}_{11}^{-1} \tag{3.4}$$

$$\mathbf{a}_0 = \boldsymbol{\mu} - \mathbf{A}^\top \boldsymbol{\mu}_{pk} \tag{3.5}$$

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma}_{00} - \boldsymbol{\Sigma}_{01}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{01}^\top \tag{3.6}$$

In this case, the complete specification, including the model assumptions, takes the form found in the table below (Spanos, 1986).

Table 3.1: Normal Vector Autoregressive (VAR(p)) Model

	Statistical GM:	$\mathbf{Z}_t = \mathbf{a}_0 + \mathbf{A}^\top \mathbf{Z}_{t-1}^{t-p} + \mathbf{u}_t,\ t\in\mathbb{N},$
[1]	Normal:	$D(\mathbf{Z}_t \vert \mathbf{Z}_{t-1}^{t-p};\boldsymbol{\Theta})$, for $\mathbf{Z}_{t-1}^{t-p}{:=}(\mathbf{Z}_{t-1},...\mathbf{Z}_{t-p})$ is Normally distributed,
[2]	Linearity:	$E(\mathbf{Z}_t \vert \sigma(\mathbf{Z}_{t-1}^{t-p})) = \mathbf{a}_0 + \mathbf{A}^\top \mathbf{Z}_{t-1}^{t-p},$
[3]	Homoskedasticity:	$Var(\mathbf{Z}_t \vert \sigma(\mathbf{Z}_{t-1}^{t-p})) = \boldsymbol{\Omega},$
[4]	Markov(p):	$\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$ is a Markov(p) process,
[5]	t-invariance:	$\boldsymbol{\Theta}{:=}(\mathbf{a}_0, \mathbf{A}, \boldsymbol{\Omega})$ are t-invariant.
		$\mathbf{a}_0 = \boldsymbol{\mu} - \mathbf{A}^\top \boldsymbol{\mu}_{pk}$, $\mathbf{A}^\top = \boldsymbol{\Sigma}_{01}\boldsymbol{\Sigma}_{11}^{-1}$, $\boldsymbol{\Omega} = \boldsymbol{\Sigma}_{00} - \boldsymbol{\Sigma}_{01}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{01}^\top$

3.2.1 Estimation of the VAR Model

One way to estimate the VAR(p) model using Maximum Likelihood estimation is to find the MLE for the parameters of the joint distribution, namely $\boldsymbol{\theta} := (\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$. Following the definitions presented in the first chapter, the distribution 'D' in the reduction above is meant as a generalization either representing the probability density function or the cumulative distribution function. To make it clear that the density function is being described, the 'D' has been replaced with the conventional notation. In this case, $\mathbf{Y}_t$ follows a multivariate normal distribution:

$$f(\mathbf{Y}_t;\boldsymbol{\theta}) = 2\pi^{-\frac{(p+1)k}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{y}}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{Y}_t - \boldsymbol{\mu}_y)^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_t - \boldsymbol{\mu}_y)\right).$$

The likelihood for the sample of size T, i.e. $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_T$ is therefore:

$$\begin{aligned} L(\boldsymbol{\theta}; \mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_T) \propto \prod_{t=1}^{T} f(\mathbf{Y}_t;\boldsymbol{\theta}) &= \prod_{t=1}^{T} 2\pi^{-\frac{(p+1)k}{2}} |\boldsymbol{\Sigma}_y|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{Y}_t - \boldsymbol{\mu}_y)^\top \boldsymbol{\Sigma}_y^{-1}(\mathbf{Y}_t - \boldsymbol{\mu}_y)\right) \\ &= 2\pi^{-\frac{(p+1)kT}{2}} |\boldsymbol{\Sigma}_y|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}(\mathbf{Y}_t - \boldsymbol{\mu}_y)^\top \boldsymbol{\Sigma}_y^{-1}(\mathbf{Y}_t - \boldsymbol{\mu}_y)\right). \end{aligned}$$

Hence, the log-likelihood is:

$$\log L(\boldsymbol{\theta}; \mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_T) = c + \frac{T}{2}\log(|\boldsymbol{\Sigma}_y^{-1}|) - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{Y}_t - \boldsymbol{\mu}_y)^\top \boldsymbol{\Sigma}_y^{-1}(\mathbf{Y}_t - \boldsymbol{\mu}_y).$$

Taking derivatives with respect to the parameters in $\boldsymbol{\theta}$ yields the same maximum likelihood estimators as those found in the Multivariate Normal specification, only now there is a temporal component which must be taken into consideration. Given the MLE for $\boldsymbol{\mu}_y$ and $\boldsymbol{\Sigma}_y$, one can then apply the parameterization in Table 3.1 to derive the MLE for the VAR(p) model.

As an alternative MLE formulation, following similar notation to that found in Hamilton (1994), suppose that the objective is to estimate the VAR(p) found in Table 3.1 above. The statistical generating mechanism can be expanded to:

$$\mathbf{Z}_t = \mathbf{a}_0 + \mathbf{A}_1^\top \mathbf{Z}_{t-1} + \cdots + \mathbf{A}_p^\top \mathbf{Z}_{t-p} + \mathbf{u}_t,$$

where $\mathbf{u}_t \sim \text{NIID}(\mathbf{0}, \boldsymbol{\Omega})$. As done in Hamilton (1994), the equation above can again be rewritten into a compact form. To do so, let

$$\mathbf{X}_t = \begin{bmatrix} 1 \\ \mathbf{Z}_{t-1} \\ \mathbf{Z}_{t-2} \\ \vdots \\ \mathbf{Z}_{t-p} \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_p \end{bmatrix}$$

Here, $\mathbf{X}_t$ is a vector of size $((kp+1) \times 1)$ and $\mathbf{B}$ is a $((kp+1) \times k)$ matrix. Thus, the conditional, $D(\mathbf{Z}_t | \mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\Theta})$, obtained from the probabilistic reduction approach above, can be written as

$$(\mathbf{Z}_t|\mathbf{Z}_{t-1}^{t-p};\boldsymbol{\Theta}) \sim \mathrm{N}(\mathbf{B}^\top\mathbf{X}_t,\boldsymbol{\Omega}).$$

Under a multivariate normal distribution, this implies that

$$f(\mathbf{Z}_t|\mathbf{Z}_{t-1}^{t-p};\boldsymbol{\Theta}) = 2\pi^{-\frac{k}{2}}|\boldsymbol{\Omega}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\mathbf{Z}_t-\mathbf{B}^\top\mathbf{X}_t)^\top\boldsymbol{\Omega}^{-1}(\mathbf{Z}_t-\mathbf{B}^\top\mathbf{X}_t)\right).$$

As in the case of the Multivariate Normal, the likelihood function for the sample is proportional to the product of the individual conditional probability density functions, i.e.

$$\begin{aligned} L(\boldsymbol{\Theta};\mathbf{Y}_1,\mathbf{Y}_2,\ldots,\mathbf{Y}_T) &= \prod_{t=1}^{T} 2\pi^{-\frac{k}{2}}|\boldsymbol{\Omega}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\mathbf{Z}_t-\mathbf{B}^\top\mathbf{X}_t)^\top\boldsymbol{\Omega}^{-1}(\mathbf{Z}_t-\mathbf{B}^\top\mathbf{X}_t)\right) \\ &= 2\pi^{-\frac{kT}{2}}|\boldsymbol{\Omega}|^{-\frac{T}{2}}\exp\left(-\frac{1}{2}\sum_{t=1}^{T}(\mathbf{Z}_t-\mathbf{B}^\top\mathbf{X}_t)^\top\boldsymbol{\Omega}^{-1}(\mathbf{Z}_t-\mathbf{B}^\top\mathbf{X}_t)\right) \end{aligned}$$

Therefore, the log-likelihood function is given as:

$$\log L(\boldsymbol{\Theta};\mathbf{Y}_1,\mathbf{Y}_2,\ldots,\mathbf{Y}_T) = c + \frac{T}{2}\log(|\boldsymbol{\Omega}^{-1}|) - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{Z}_t-\mathbf{B}^\top\mathbf{X}_t)^\top\boldsymbol{\Omega}^{-1}(\mathbf{Z}_t-\mathbf{B}^\top\mathbf{X}_t).$$

Taking partial derivatives of the log-likelihood function with respect to $\mathbf{B}$:[1]

$$\frac{\partial \log L(\boldsymbol{\Theta};\mathbf{Y}_1,\mathbf{Y}_2,\ldots,\mathbf{Y}_T)}{\partial\mathbf{B}} = -\frac{1}{2}2\sum_{t=1}^{T}\boldsymbol{\Omega}^{-1}(\mathbf{Z}_t-\mathbf{B}^\top\mathbf{X}_t)\mathbf{X}_t^\top$$

Setting this equal to zero yields the MLE for the matrix of coefficients with the constant

[1] Note that for a vector $\mathbf{x}$ and a vector $\mathbf{a}$:

$$\frac{\partial\mathbf{x}^\top\mathbf{a}}{\partial\mathbf{x}} = \mathbf{a}^\top.$$

This property is particularly useful in computing the partial derivative of the log-likelihood with respect to $\mathbf{B}$.

included:

$$\frac{\partial \log L(\boldsymbol{\Theta}; \mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_T)}{\partial \mathbf{B}} = -\frac{1}{2} 2 \sum_{t=1}^{T} \boldsymbol{\Omega}^{-1} (\mathbf{Z}_t - \mathbf{B}^\top \mathbf{X}_t) \mathbf{X}_t^\top = \mathbf{0}$$

$$\Rightarrow \mathbf{B}^\top = \left[\sum_{t=1}^{T} \mathbf{Z}_t \mathbf{X}_t^\top \right] \left[\sum_{t=1}^{T} \mathbf{X}_t \mathbf{X}_t^\top \right]^{-1}$$

Hence,

$$\hat{\mathbf{B}}_{MLE} = \left[\left[\sum_{t=1}^{T} \mathbf{Z}_t \mathbf{X}_t^\top \right] \left[\sum_{t=1}^{T} \mathbf{X}_t \mathbf{X}_t^\top \right]^{-1} \right]^\top$$

As shown in both Hamilton (1994) and Lüketpohl (2005), the maximum likelihood estimates for the VAR(p) model coefficients, when normality is imposed, will be identical to those found using multivariate least squares. Essentially, one can estimate the VAR(p) model via OLS equation-by-equation, regressing the elements of $\mathbf{Z}_t$ on the constant and the p lags; see Hamilton (1994) for verification. Similarly,

$$\frac{\partial \log L(\boldsymbol{\Theta}; \mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_T)}{\partial \boldsymbol{\Omega}^{-1}} = \frac{T}{2} \boldsymbol{\Omega} - \frac{1}{2} \sum_{t=1}^{T} (\mathbf{Z}_t - \mathbf{B}^\top \mathbf{X}_t)(\mathbf{Z}_t - \mathbf{B}^\top \mathbf{X}_t)^\top = \mathbf{0}$$

$$\Rightarrow \hat{\boldsymbol{\Omega}}_{MLE} = \frac{1}{T} \sum_{t=1}^{T} (\mathbf{Z}_t - \hat{\mathbf{B}}_{MLE}^\top \mathbf{X}_t)(\mathbf{Z}_t - \hat{\mathbf{B}}_{MLE}^\top \mathbf{X}_t)^\top.$$

Defining $\boldsymbol{\varepsilon}_t = \mathbf{Z}_t - \mathbf{B}^\top \mathbf{X}_t$, the expression above can be reduced to:

$$\hat{\boldsymbol{\Omega}}_{MLE} = \frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t^\top,$$

which is again analogous to the multivariate OLS results.

3.2.2 Time Dependent 'Temporal' Principal Components

To establish a statistically adequate forecasting framework centered around PCA, it is first necessary to establish principal components which capture the time dependence inherent in most macroeconomic data series. In other words, the goal is to create 'temporal' principal components which reflect the underlying probabilistic structure of the original data. Much like the traditional principal components and their relationship to the Multivariate Normal model, the temporal principal components in this chapter hold a relationship with the VAR(p) model. Here, principal components are constructed from the temporal variance-covariance matrix. Recall the joint distribution resulting from the probabilistic reduction of the VAR(p):

$$\mathbf{Y}_t = \begin{bmatrix} \mathbf{Z}_t \\ \mathbf{Z}_{t-1} \\ \vdots \\ \mathbf{Z}_{t-p} \end{bmatrix} \sim N \left(\boldsymbol{\mu}_y = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{bmatrix}, \boldsymbol{\Sigma}_y = \begin{bmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01} & \boldsymbol{\Sigma}_{02} & \dots & \boldsymbol{\Sigma}_{0p} \\ \boldsymbol{\Sigma}_{10}^\top & \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \dots & \boldsymbol{\Sigma}_{1p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{p0}^\top & \boldsymbol{\Sigma}_{p1}^\top & \boldsymbol{\Sigma}_{p2}^\top & \dots & \boldsymbol{\Sigma}_{pp} \end{bmatrix} \right),$$

From this point, the objective is to construct principal components from the temporal covariance matrix $\boldsymbol{\Sigma}_y$. First, note that this covariance matrix is a positive definite, symmetric matrix. Therefore, the previous discussion on the derivation of principal components applies here. Hence, the temporal covariance matrix can be decomposed using spectral decomposition:

$$\boldsymbol{\Sigma}_y = \mathbf{V}\mathbf{D}\mathbf{V}^\top,$$

where $\mathbf{V} := (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{(p+1)k})$ is the $(p+1)k \times (p+1)k$ matrix whose columns are eigenvectors, and the diagonal matrix $\mathbf{D} = \text{diag}(\lambda_1 \geq \lambda_2 \cdots \geq \lambda_{(p+1)k} \geq 0)$ denotes the ranked eigenvalues of the temporal covariance matrix.

As done with the traditional principal components, the first r, $r \leq (p+1)k$, eigenvectors are chosen to form linear combinations with the demeaned data vector, now augmented with the first p lags:

$$\mathbf{F}_t = \mathbf{V}_r^\top (\mathbf{Y}_t - \boldsymbol{\mu}_y), \ \ t \in \mathbb{N}.$$

Note that $\mathbf{V}_r$ denotes the r columns, or eigenvectors, of the matrix $\mathbf{V}$ that correspond to the r largest eigenvalues. This equation can be expanded to:

$$\mathbf{F}_t = \mathbf{V}_r^\top \left(\begin{bmatrix} \mathbf{Z}_t \\ \mathbf{Z}_{t-1} \\ \vdots \\ \mathbf{Z}_{t-p} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{bmatrix} \right), \ \ t \in \mathbb{N}.$$

Since it has been assumed that $\mathbf{Y}_t$ is assumed to follow a multivariate normal distribution, i.e. $\mathbf{Y}_t \sim N(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$, additional meaning can be given to the principal components. Namely, it follows that:

$$\mathbf{F}_t \sim N(\mathbf{0}, \mathbf{D}_r),$$

where $\mathbf{D}_r$ is the $(r \times r)$ diagonal matrix containing the largest r eigenvalues, ordered by magnitude. This leads to the Normal, Temporal Principal Components Model found in Table 3.2. As was the case in the traditional specification, the error term, denoted as $\mathbf{u}_t$, includes the approximation error, which stems from the exclusion of certain principal components. Discussion on methods for choosing which principal components to retain can be found in the following sections.

Table 3.2: Normal, Temporal Principal Components Model

Statistical GM:		$\mathbf{F}_t = \mathbf{V}_r^\top(\mathbf{Y}_t - \boldsymbol{\mu}_y) + \mathbf{u}_t,\ \mathbf{Y}_t = [\mathbf{Z}_t, \mathbf{Z}_{t-1}, \ldots, \mathbf{Z}_{t-p}]^\top$
[1]	Normality:	$\mathbf{F}_t \sim N(.,.)$
[2]	Linearity:	$E(\mathbf{F}_t) = \mathbf{V}_r^\top(\mathbf{Y}_t - \boldsymbol{\mu}_y)$,
[3]	Constant Covariance:	$Cov(\mathbf{F}_t) = \mathbf{D}_r$,
[4]	Markov:	$\{\mathbf{Z}_t, t \in \mathbb{N}\}$ is an Markov(p) process
[5]	t-invariance:	$\boldsymbol{\theta} := \{\boldsymbol{\mu}_y, \mathbf{D}_r, \mathbf{V}_r\}$ does not change with t

As stated above, the aim of of the methodology presented in this chapter is to expand upon the traditional definition of principal components to account for all of the systematic information in the data. Specifically, the goal is to create a framework in which principal components can capture the dependence information inherent in many macroeconomic time series via the probabilistic reduction approach. To this end, it would be a mistake not to mention how the 'temporal' principal components defined here differ from those used in the present-day dynamic factor models. In summary, the key difference between the 'temporal' principal components and the dynamic principal components is that the former are derived directly from their relationship with a vector autoregressive model in the time domain, whereas the latter is formulated using spectral analysis in the frequency domain (Brillinger, 1981). Furthermore, dynamic principal components are not applicable for the macroeconomic data, since they initially assume stationarity.

3.2.3 Estimation of Temporal Principal Components

Given the normality assumption, the density function for $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_T$ is:

$$f(\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_T; \boldsymbol{\theta}) = 2\pi^{-\frac{(p+1)kT}{2}} |\boldsymbol{\Sigma}_y|^{-\frac{T}{2}} exp\left(-\frac{1}{2}\sum_{t=1}^{T}(\mathbf{Y}_t - \boldsymbol{\mu}_y)^\top \boldsymbol{\Sigma}_y^{-1}(\mathbf{Y}_t - \boldsymbol{\mu})\right)$$

From the discussion on the simple Multivariate Normal model above, the MLE for $\boldsymbol{\Theta} := (\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$ equal:

$$\hat{\boldsymbol{\mu}}_y = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Y}_t, \quad \hat{\boldsymbol{\Sigma}}_y = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{Y}_t - \hat{\boldsymbol{\mu}}_y)(\mathbf{Y}_t - \hat{\boldsymbol{\mu}}_y)^\top,$$

Thus, the MLE for the temporal covariance matrix is merely the sample covariance matrix of the original variables plus the lags of those variables. Using the property that the MLE are parameterization invariant, as discussed in Chapter 1, the MLE for $\mathbf{D}, \mathbf{V}$ come from the spectral decomposition of the extended covariance matrix:

$$\hat{\boldsymbol{\Sigma}}_y = \hat{\mathbf{V}}\hat{\mathbf{D}}\hat{\mathbf{V}}^\top.$$

The estimated principal components are then constructed using the r largest eigenvalues and the corresponding eigenvectors.

3.3 Forecasting Framework

3.3.1 Step 1: Determining the Number Trends and Lags

It is standard practice to first center the data columns before using principal components (Jolliffe, 2002). While demeaning the data often is accomplished by subtracting the sample mean, this can lead to issues when the data is trending. Moreover, one can take first differences of the natural logarithms of the macroeconomic variables to not only eliminate any lingering trends, but to also avoid lingering effects from unit roots (Stock and Watson, 2002b). In this case, however, the primary interest is the forecasting of the original predictors, not the changes in those predictors. Hence, mean heterogeneity is modeled directly through estimated trend polynomials and trigonometric polynomials for seasonality. The number of lags necessary to capture the dependence are subsequently determined through the estimation of an autoregressive model on each of the predictors and a vector autoregressive model.

Modeling Seasonality through Trigonometric Polynomials

Seasonal variables are constructed using trigonometric functions. Let S denote the season of interest. As shown in Ghysels and Osborn (2001), estimating seasonal variables amounts to estimating the following equations:

$$\sum_{k=1}^{S/2} \gamma_k cos(\frac{2\pi kt}{S}) + \psi_k sin(\frac{2\pi kt}{S}), \tag{3.7}$$

where t denotes the time variable. In the case where the data is quarterly, i.e. $S = 4$, this amounts to estimating:

$$\sum_{k=1}^{2} \gamma_k cos(\frac{2\pi kt}{4}) + \psi_k sin(\frac{2\pi kt}{4}). \tag{3.8}$$

Modeling Mean-heterogeneity through Trend Polynomials

Time variables used in the autoregressive model are constructed using the method discussed in Michaelides and Spanos (2020). This is done to avoid near-multicollinearity caused from the inclusion of high degrees of time polynomials and the misleading statistical results that ensue. Namely, as mentioned in Michaelides and Spanos (2020), the inclusion of trend polynomials to capture mean-heterogeneity can lead to unstable estimates due to an ill-conditioned matrix in the OLS estimation; see Spanos and McGuirk (2002) and Spanos (2019b) for more discussion on the near-multicollinearity. To combat this, the authors propose first re-scaling the time variable to numbers within the range [-1,1] using the equation:

$$t_* = \frac{2t - n - 1}{n - 1}, \tag{3.9}$$

where t is the original time index, and n is the total number of observations. When the number of trend terms is large, say greater than the sixth or seventh power (Michaelides and Spanos, 2020), scaling proves to be insufficient to eliminate the near-multicollinearity problem. However, employing orthogonal polynomials, in addition to the re-scaling, was found to have reduced the evidence of ill-conditioning. Thus, the second step is to compute the Chebyshev polynomial of the first kind, via:

$$t^k(t_*) = \frac{(-1)^k 2^k (k)!}{(2k)!}(1 - t_*^2)^{\frac{1}{2}}(\frac{d}{dt_*})^k(1 - t_*^2)^{k-\frac{1}{2}}, k \geq 0 \tag{3.10}$$

Note that other orthogonal polynomials can be considered, as shown in Michaelides and Spanos (2020), but the choice will depend on the range of the time variable. Here, in order to use the Chebyshev polynomials, the time variable must take values within the range [-1,1]. From this point onward, the notation for the scaled Chebyshev polynomial is simplified to t_*^k, where k is the degree of the trend.

Choosing the Number of Trends and Lags

Once the trend polynomials and seasonality terms have been constructed, the Normal, AR(p) model with mean heterogeneity (Spanos, 2019a), given in Table 3.3 below, is estimated for each data series.

Note that, in Table 3.3, $\mathbf{X}_{t-1}^{t-p} := \left(X_{t-1}, \ldots, X_{t-p}\right)$. Additionally, the constant term in Table 3.3, $a_0(t)$ contains all of the added seasonal and trend polynomials as well as potential

Table 3.3: Normal, Autoregressive (AR(p)) Model with Mean Heterogeneity

	Statistical GM:	$X_t = \alpha_0(t) + \sum_{i=1}^{p} \alpha_i X_{t-i} + u_t, t \in \mathbb{N},$
[1]	Normality:	$(X_t, X_{t-1}, \ldots, X_{t-p}) \sim N(.,.)$
[2]	Linearity:	$\mathrm{E}\left(\mathrm{X}_t \middle| \sigma\left(\mathbf{X}_{t-1}^{t-p}\right)\right) = \alpha_0(t) + \sum_{i=1}^{p} \alpha_i X_{t-i},$
[3]	Homoskedasticity:	$\mathrm{Var}\left(\mathrm{X}_t \middle| \sigma\left(\mathbf{X}_{t-1}^{t-p}\right)\right) = \sigma_0^2,$
[4]	Markov:	$\{\mathrm{X}_t, t \in \mathbb{N}\}$ is Markov of order p
[5]	t-invariance:	$\left(\alpha_0(t), \alpha_1, \ldots, \alpha_p, \sigma_0^2\right)$ does not change with t

dummy variables for outliers.[2] The number of lags and trends are selected based on statistical adequacy of the models. Here, the notion of statistical adequacy follows from the discussion in Section 2.1.2. Simply put, a statistically adequate model is one which encapsulates all of the chance regularity patterns in the data (Spanos, 2019a). To test for departures from the model assumptions imposed on the data, auxiliary regressions are conducted on the residuals of each data series, see Spanos (2018) for details on such auxiliary regressions. An example of these auxiliary regressions where the dependent variable is the natural log of GDP can be found in Appendix D. After thoroughly testing for potential mis-specification, the models are then re-specified to account for the information not captured in the current version.

Once a statistically adequate autoregressive model has been estimated for each predictor, a vector autoregressive model VAR(L) is estimated to jointly determine the number of lags and trends. It is important to clarify that the number of trends and lags determined in this model may differ from that found in the AR(p) model estimation, hence the distinction between p and L. The specification of the VAR(L) model with trending and seasonal means can be respectively found in Tables 2.3 and 2.4. Moreover, the estimation of the VAR(L) is carried

[2]Observations are only considered to be outliers if they significantly affect the other model estimates during their inclusion or omission.

out using OLS for each equation separately. This way, the model mis-specification testing can be conducted on the residuals of each equation separately. The minimum number of trends and lags that establish statistical adequacy for all of the VAR(L) equations is recorded and then used for the construction of temporal principal components.

How/Why this Method Differs

The aforementioned method for determining the number of trends and lags stands in stark contrast to other, popular methods utilized in papers such as Stock and Watson (2002a) and Bok et al. (2018). As seen in Bok et al. (2018), data downloaded from the FRED database is typically seasonally-adjusted prior to analysis. This seasonal adjustment often uses the X-13 ARIMA-SEATS Seasonal Adjustment program, which is available on the U.S. Census Bureau website.[3] As the name suggests, X-13 ARIMA-SEATS utilizes ARIMA models to subtract out the seasonal component of a dataset. Modeling the seasonality directly is therefore seen as unnecessary and therefore omitted from the analysis, as the downloaded data is assumed to have taken seasonal cycles into account. Implicitly, however, this procedure assumes that the ARIMA model is statistically adequate for the data. If there is model mis-specification, then the subtracted seasonal component will be erroneous, as will any inferences/forecasts which use the adjusted data.

Another point worth mentioning is that information criterion, such as the commonly used Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) (Akaike, 1973, Schwarz, 1978) are not considered when determining the number of trends and lags. As shown in Spanos (2010b), these types of model selection procedures can be unreliable, especially when in the presence of model mis-specification. If, for example, the statistically adequate model is not in the set of possible models being considered, the AIC and BIC will

[3] https://www.census.gov/data/software/x13as.html

select a mis-specified model. Moreover, Spanos (2010b) finds that even if a statistically adequate model is in the possible set, the information criterion may select a different model. To avoid the possibility that these model selection procedures will select statistically inadequate autoregressive and vector autoregressive models, this forecasting framework instead employs the alternative proposed in Spanos (2010b). That is, achieving statistical adequacy is the lone criterion for selecting the number of trends and lags in the autoregressive and vector autoregressive models.

3.3.2 Step 2: Detrending and De-seasonalizing the Data

Having determined a statistically adequate autoregressive model for each predictor, where the predictor i at time t is denoted as $X_{i,t}$ $(i = 1, \ldots, k), (t = 1, \ldots, T)$, the data is then detrended and de-seasoned using the estimated trends and lags. Hence, detrending and de-seasonalizing amounts to estimating[4]:

$$X_{i,t} = \overbrace{\alpha_0 + \sum_{w=1}^{W} \delta_i t_*^w + \sum_{j=1}^{S/2} \gamma_j cos(\frac{2\pi jt}{S}) + \psi_j sin(\frac{2\pi jt}{S})}^{\mu_i(t)} + u_t, \tag{3.11}$$

where $\mu_i(t)$ is the mean of predictor i. Subtracting out the estimated mean yields the detrended predictor, $\hat{X}_{i,t}(d)$:

$$\hat{X}_{i,t}(d) = X_{i,t} - \hat{\mu}_i(t). \tag{3.12}$$

3.3.3 Step 3: Sample Variance-Covariance Matrix

A new dataset containing both the detrended data series and their lags is formulated based on the number of trends and lags for the system determined via the statistically adequate vector autoregressive model. Let $\hat{\mathbf{Z}}_t(d) = [\hat{X}_{1,t}(d), \hat{X}_{2,t}(d), \ldots, \hat{X}_{k,t}(d)]^\top$ denote the $(k \times 1)$ detrended random vector. Following notation similar to that when defining the Vector

[4]While the estimation of each autoregressive model was conducted separately for all of the data series, and this will be an integral part of the methodology prescribed below, the choice of the number of lags and included trend variables is the same across all of the predictors. These results come from the estimated vector autoregressive model. To reiterate, the number of lags and trends may differ between the two models.

Autoregressive model, let

$$\hat{\mathbf{Y}}_t(d) := \left[\hat{\mathbf{Z}}_t(d), \hat{\mathbf{Z}}_{t-1}(d), \hat{\mathbf{Z}}_{t-2}(d), \ldots, \hat{\mathbf{Z}}_{t-L}(d)\right]^\top,$$

where L is the number of lags. The newly constructed data matrix contains the detrended and therefore demeaned data columns, i.e. it is assumed that $E(\hat{\mathbf{Z}}_t(d)) = \mathbf{0}$ and therefore $E(\hat{\mathbf{Y}}_t(d)) = \mathbf{0}$. Thus, after constructing the detrended data matrix with lags included, the sample covariance matrix is computed via:

$$\hat{\mathbf{\Sigma}}_y = \frac{1}{T}\sum_{t=1}^{T}[\hat{\mathbf{Y}}_t(d)][\hat{\mathbf{Y}}_t(d)]^\top$$

Note that $\hat{\mathbf{\Sigma}}_y$ is a $(L+1)k \times (L+1)k$ matrix. There are two important considerations regarding the sample temporal covariance matrix to keep in mind at this point. First, the inclusion of the lagged variables in the original dataset poses the problem of missing values when computing the covariance matrix. One solution to this problem is to simply omit the first few rows of the data matrix where the missing values occur. However, in doing so, the assumption that the means of each column is zero is no longer valid.[5] The second method, which is used here, involves adding the estimated constant term in each trend equation back to the respective data column. This will not change the covariances, since the variance of a constant term has no bearing.

Moreover, units of measurement will have a major impact on the contribution of each predictor to the principal components. In the case where units of measurement differ among the predictors, it is common practice to first standardize the covariance matrix, i.e. to generate

[5]Note that, in Python, the libraries numpy (Harris et al., 2020) and pandas (Wes McKinney, 2010) will produce slightly different covariance matrices for this exact reason. Pandas will attempt to fill the missing values using the sample mean of the data column, whereas numpy will drop the rows of the matrix containing missing values.

the correlation matrix; see Jolliffe (2002), Jolliffe and Cadima (2016) for more information on standardization of principal components. Here, this is accomplished by pre and post multiplying the covariance matrix by the inverse of the square-root of the diagonal entries of the covariance matrix, or the inverse of the standard deviation of each data column. To summarize, let $\hat{\mathbf{S}}$ denote a diagonal matrix containing the square root of the diagonal elements of $\hat{\boldsymbol{\Sigma}}_y$. It follows that the correlation matrix, $Corr(\hat{\mathbf{Y}}_t(d))$, equals:

$$Corr(\hat{\mathbf{Y}}_t(d)) = \hat{\mathbf{S}}^{-1}\hat{\boldsymbol{\Sigma}}_y\hat{\mathbf{S}}^{-1}.$$

3.3.4 Step 4: Estimating the Temporal Principal Components

Assuming that the units of measurement differ among the data series necessitates the use of the standardized correlation matrix, $Corr(\hat{\mathbf{Y}}_t(d))$, to estimate principal components. Spectral decomposition is first applied to correlation matrix to obtain its eigenvalues and corresponding eigenvectors:

$$Corr(\hat{\mathbf{Y}}_t(d)) = \hat{\mathbf{V}}\hat{\mathbf{D}}\hat{\mathbf{V}}^{\top}$$

Having completed this step, the temporal principal components defined in Table 3.2 are then estimated:

$$\hat{\mathbf{F}}_t = \hat{\mathbf{V}}^{\top}(\mathbf{Y}_t - \hat{\boldsymbol{\mu}}_y) = \hat{\mathbf{V}}^{\top}(\hat{\mathbf{Y}}_t(d))$$

As mentioned above, the second equals sign follows from the assumption that $E(\hat{\mathbf{Y}}_t(d)) = \mathbf{0}$, which implies that the data has been demeaned correctly.

3.3.5 Step 5: Selecting Principal Components

There exists a degree of subjectivity when selecting which principal components to keep and which to drop; this stems from the diversity of methods for principal component selection. Methods used to choose the number of principal components include ad-hoc rules such as the scree plot, tests based on distributional assumptions, and other methods such as bootstrapping, permutation and cross validation (Jolliffe, 2002, Josse and Husson, 2012). In Bai and Ng (2002), the authors determine the number of factors to use in approximate factor models based on a developed information criterion. Newer methods, such as sparse principal component analysis, utilize variable selection algorithms like the ridge and lasso regressions to not only reduce dimensionality, but to also shrink the number of components utilized (Zou et al., 2006).

This framework uses follows a combination of distributional assumptions, scree plot, and statistical significance to select both the number of principal components and which of these principal components to keep in the final forecasting model. The first method to select the number of principal components to keep is the traditional method for determining the cutoff via the proportion of total variance, π_r, explained by the included r principal components:

$$\pi_r = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_r}{\lambda_1 + \lambda_2 + \cdots + \lambda_{(L+1)k}} = \frac{\sum_{i=1}^{r} \lambda_i}{\text{trace}(\mathbf{\Sigma}_y)}. \tag{3.13}$$

Here, one specifies a threshold for the percentage of variance explained, such as 95%, and selects the r eigenvectors corresponding to the r eigenvalues whose summation reaches this threshold. This method of selecting principal components can be fully integrated into a hypothesis testing framework (Anderson, 2003). Namely, suppose the question of interest is whether the last $(L+1)k-r$ principal components can be ignored based on their contribution to the total variance. As stated in Anderson (2003), in this case, the null hypothesis can be formulated as:

$$H_0 : f(\boldsymbol{\lambda}) = \frac{\sum_{i=r+1}^{(L+1)k} \lambda_i}{\text{trace}(\boldsymbol{\Sigma}_y)} \geq c, \tag{3.14}$$

where c is the pre-specified threshold for the percentage of variance explained. Under equality in 3.14, Anderson shows that the asymptotic variance of $f(\hat{\boldsymbol{\lambda}})$, where $\hat{\boldsymbol{\lambda}}$ is the vector of estimated eigenvalues, equals (2003):[6]

$$Var(f(\hat{\boldsymbol{\lambda}})) \xrightarrow[T\to\infty]{} 2\left(\frac{c}{\text{trace}(\boldsymbol{\Sigma}_y)}\right)^2 (\lambda_1^2 + \cdots + \lambda_r^2) + 2\left(\frac{1-c}{\text{trace}(\boldsymbol{\Sigma}_y)}\right)^2 (\lambda_{r+1}^2 + \cdots + \lambda_{(L+1)k}^2). \tag{3.15}$$

Therefore, the null hypothesis in 3.14 is rejected if

$$\sqrt{T}\left[\left(\frac{\sum_{i=r+1}^{(L+1)k} \hat{\lambda}_i}{\text{trace}(\hat{\boldsymbol{\Sigma}}_y)}\right) - c)\right] < z_\alpha \sqrt{Var(f(\hat{\boldsymbol{\lambda}}))},$$

where z_α denotes the α-significance level critical point of the standard normal distribution, and the MLE have been substituted in the asymptotic variance in 3.15. For brevity, the theorems and intermediate steps have not been fully illustrated here; see Anderson (2003) for a more complete picture on the procedure.[7] Other than the proportion of total variance explained, this chapter uses a scree plot is also used to indicate which components to keep. Finally, after estimating the subsequently defined FAR(p,L) model, only principal components which are significantly (statistically) related to the variable of interest are kept for use in the forecasts.

[6] This follows from Theorem 4.2.3 in Anderson (2003).

[7] Alternatively, Anderson shows that one can also test whether the sum of the smallest roots, i.e. the sum of the smallest variances of the last principal components is less than a pre-specified threshold.

3.3.6 Step 6: Estimating the Dynamic Factor-Augmented Autoregressive Model (FAR(p,L))

Given that the aim of this line of research is to incorporate the dependence information depicted in the VAR(L) model into the the autoregressive models (AR(p)) via estimated principal components, a new model specification that captures all of the assumptions needs to be transcribed. Thus, the specification needs to include conditioning information from both the original autoregressive model and the temporal principal components, which were derived from the VAR(L) model. The culmination is the Dynamic Factor-Augmented Autoregressive model (FAR(p,L)), which is contained in Table 3.4 below.

Table 3.4: Dynamic Factor Augmented Autoregressive Model (FAR(p,L))

	Statistical GM:	$X_t = a_0(t) + \sum_{i=1}^{p} \beta_i X_{t-i} + \sum_{j=1}^{r} \delta_j f_{j,t} + u_t,\ \ t\in\mathbb{N},\ r \leq (L+1)k$
[1]	Normal:	$(X_t, X_{t-1}, \ldots, X_{t-p}, f_{1,t}, \ldots f_{r,t})$ is Normally distributed,
[2]	Linearity:	$E(X_t \vert \sigma(\mathbf{W}^0_{t-1})) = a_0(t) + \sum_{i=1}^{p} \beta_i X_{t-i} + \sum_{j=1}^{r} \delta_j f_{j,t},$
[3]	Homoskedasticity:	$Var(X_t \vert \sigma(\mathbf{W}^0_{t-1})) = \omega^2,$
[4]	Markov:	$\{X_t, t\in\mathbb{N}\}$ is a Markov(p) process, $\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$ is a Markov(L) process
[5]	t-invariance:	$\boldsymbol{\theta}\mathbf{:=}(a_0(t), \beta_i, \ldots, \beta_p, \delta_1, \ldots, \delta_r, \omega^2)$ are t-invariant.
		$\mathbf{Z}_t = [X_{1,t}, X_{2,t}, \ldots, X_{k,t}]^\top,\ X_t \in \{X_{1,t}, X_{2,t}, \ldots, X_{k,t}\},$ $\mathbf{Y}_t = [\mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \ldots, \mathbf{Z}_{t-L}]^\top,\ \mathbf{F}_t = \mathbf{V}_r^\top(\mathbf{Y}_t - \boldsymbol{\mu}_y),$ $\mathbf{F}_t = (f_{1,t}, f_{2,t}, \ldots, f_{r,t}),\ \mathbf{W}^0_{t-1} := (X_{t-1}, \ldots, X_{t-p}, \mathbf{Z}_{t-1}, \ldots, \mathbf{Z}_{t-L})$

Note that the information contained in $\mathbf{W}^0_{t-1}$ depends on the data availability and frequency of release. In practice, it may very well be the case that lower frequency data, such as monthly and quarterly data, will not be available at the time of forecasting. It is for this reason that the temporal covariance matrix used to construct the temporal principal components may not contain the contemporaneous matrix with those variables included. Ultimately, in terms of forecasting, the omission or inclusion of such variables should not impact the results significantly, due to the nature of principal components.

Moreover, this formulation of the FAR(p,L) model should not be confused with the FAVAR model proposed in Bernanke et al. (2005), which is based on the dynamic factor model. The main difference is that the FAR(p,L) model is based on the reduction of the joint distribution of the sample based on the imposed probabilistic assumptions inherent in the AR(p) and VAR(L) models (Spanos, 1986). For more information on FAVAR models as well as their relation to structural dynamic factor models, see Bernanke et al. (2005) and Stock and Watson (2016). Moreover, the FAVAR model will be used as a benchmark for the FAR(p,L). Details regarding the model and benchmark are found in Appendix I.

3.3.7 Step 7: Model Mis-specification Testing of the FAR(p,L) Model

Model mis-specification testing is conducted on the estimated residuals from the FAR(p,L) in order to ensure that adding the estimated principal components does not affect the statistical adequacy of the original autoregressive model. Auxiliary regressions on the estimated residuals, as shown in Spanos (2018, 2019a), detect departures from the probabilistic assumptions invoked in the model. For example, in the context of the normal, FAR(p,L) model with cubic trends, auxiliary regressions could take the following form:

$$\widehat{u}_t = \overbrace{a_0 + \sum_{i=1}^{3} \delta_i t_*^i + \sum_{j=1}^{r} \gamma_j f_{j,t} + \sum_{m=1}^{p} \gamma_m y_{t-m}}^{\text{original specification}} + \overbrace{a_1 t_*^4 + a_2 t_*^5}^{[5]} + \overbrace{\beta_1 \hat{y}_t^2}^{[2]} + \overbrace{\beta_2 \hat{u}_{t-1} + \beta_3 \hat{u}_{t-2}}^{[4]} + v_t$$

$$\widehat{u_t}^2 = \delta_0 + \overbrace{\delta_1 t_*^1 + \delta_2 t_*^2}^{[5]} + \overbrace{\psi_1 \hat{y}_t^2}^{[3]} + \overbrace{\gamma_1 \hat{u}_{t-1}^2 + \gamma_2 \hat{u}_{t-2}^2}^{[4]} + v_t$$

Table 3.4 shows the null hypotheses for testing each assumption explicitly. Note that these mis-specification tests detect departures from the probabilistic assumptions by testing the additional terms, not the terms from the original model specification.

Table 3.4: M-S Tests: Factor Augmented Autoregression

Assumptions	**Null Hypothesis**
Joint (Regression Function)	H_0: $a_1 = a_2 = \beta_1 = \beta_2 = \beta_3 = 0$
Linearity	H_0: $\beta_1 = 0$
Dependence (Mean)	H_0: $\beta_2 = \beta_3 = 0$
1st t-invariance	H_0: $a_1 = a_2 = 0$
Joint (Skedastic Function)	H_0: $\delta_1 = \delta_2 = \psi_1 = \gamma_1 = \gamma_2 = 0$
Heteroskedasticity	H_0: $\psi_1 = 0$
Dependence (Variance)	H_0: $\gamma_1 = \gamma_2 = 0$
2nd t-invariance	H_0: $\delta_1 = \delta_2 = 0$

3.3.8 Step 8: Constructing Out-of-Sample Forecasts

The main focus of this chapter is to find out how the modelling methodology described in Steps 1-7 can be used to forecast macroeconomic variables. Hence, in the final step, out-of-sample forecasts are conducted for each of the predictors to find out whether the addition of principal components increases forecasting accuracy. First, a training dataset and testing dataset are established; here, the last twelve quarters of each predictor make up testing dataset, i.e. Q1 2017 through Q4 2019.

Forecasts are then constructed recursively from the estimated equation. Namely, forecasts are iteratively computed one quarter ahead over the forecasting horizon using the statistical generating mechanism. Consider, without loss of generality, the simple yet relevant benchmark AR(1) model. The estimated equation takes the following form:

$$\hat{y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 y_{t-1}$$

To forecast the next time period, i.e. time $t+1$, one can use the estimated value from the previous time period:

$$\hat{y}_{t+1} = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{y}_t$$

This process is repeated to then forecast h periods ahead, where h is defined to be the forecasting horizon. Ultimately, the forecast at time $t+h$ would then be:

$$\hat{y}_{t+h} = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{y}_{t+h-1}.$$

In the context of the FAR(p,L) model, the forecasting follows a similar recursive procedure. Using the statistical generating mechanism defined in Table 3.4, for a FAR(1,3), the estimated equation from the training set is given as:

$$\hat{y}_t = \hat{\alpha}_0(t) + \hat{\alpha}_1 \hat{y}_{t-1} + \sum_{j=1}^{r} \hat{\delta}_j \hat{f}_{j,t}$$

Like the benchmark model above, forecasting h steps ahead amounts to plugging in the estimated values into the statistical generating mechanism:

$$\hat{y}_{t+h} = \hat{\alpha}_0(t+h) + \sum_{i=1}^{p} \hat{\alpha}_i \hat{y}_{t+h-i} + \sum_{j=1}^{r} \hat{\delta}_j \hat{f}_{j,t+h}$$

A few things to note about this setup. Firstly, in this framework, the FAR(p,L) model utilizes principal components which have been constructed from the full sample. When forecasting in practice, however, observations whose dates are beyond the training set, in other words the out-of-sample observations, will generally be unknown to the modeller at the time the forecasts are created.

3.3.9 Summary of Forecasting Steps

To summarize, the forecasting framework is given in the following steps:

Step 1. Determine the number of trends and lags via statistically adequate AR(p) and VAR(L) models.

Step 2. Detrend and de-seasonalize the data using the estimated trends from the statistically adequate VAR(L) model.

Step 3. Estimate the sample, temporal (with lags) covariance matrix from the VAR(L) model.

Step 4. Estimate the temporal principal components using spectral decomposition or SVD.

Step 5. Choose which components to keep based on statistical significance, scree plot, etc.

Step 6. Estimate the Dynamic Factor-Augmented Autoregressive Model (FAR(p,L)).

Step 7. Check the statistical adequacy of the FAR(p,L) using M-S testing.

Step 8. Construct out-of-sample forecasts iteratively over the forecast horizon.

3.4 Data

Predictors

This chapter utilizes twelve widely used macroeconomic and financial data series, including both leading and lagging indicators, to demonstrate the amenability of this principal components method to forecasting economic data, see Bok et al. (2018) for an example of macroeconomic nowcasting using these predictors. Moreover, while at the outset twelve data series will not likely constitute a "big data" problem, or rather an issue with dimensionality (Bellman, 1961), macroeconomic data series often display various forms of dependence which require the inclusion of lagged variables. Stacking these lags with the original data set will ultimately increase the number of dimensions in the data series. Furthermore, the aim of this chapter is to present a methodology for the general forecasting of macroeconomic variables. As such, it is simple to include more data series into the set of predictors.

The macroeconomic data in this analysis was primarily sourced from the FRED database, which is maintained by the Federal Reserve Bank of St. Louis [8]. Since the data frequency varies between quarterly and monthly observations, the monthly data series are aggregated to a quarterly frequency. Although current macroeconomic nowcasting techniques use smoothing methods, such as the application of the Kalman Filter, smoothing the data imposes restrictions that may lead to model mis-specification.[9] Aggregation can also be accomplished through a variety of methods including averages and interpolation, but these methods risk running into similar restrictions on the data. Thus, quarterly variables are constructed using the end-of-quarter value for each predictor. Seasonal data on real GDP only reaches as far back as Q1 2002. Therefore, the more readily available nominal GDP with seasonality is

[8] https://fred.stlouisfed.org/

[9] Aggregation does lead to the loss of some information, but since the goal here is to ensure that the forecasting model is statistically adequate, this loss is less relevant.

transformed to real GDP using the GDP deflator.[10]

After variable transformations and adjustments, quarterly data from Q1 1992 to Q4 2019 remain for the analysis. Figure 3.1 below shows the time series plots of the original data. Note that the version of the data without the seasonal cycles subtracted out were chosen when available, as seasonality is directly modeled using trigonometric polynomials. This is further expanded upon when establishing the forecasting methodology. See Appendix C for more details on each data series.

Note that descriptive statistics, such as the sample mean and sample variance, are highly misleading when the data series are not independent and identically distributed (IID). It is clear from the Figure 3.1 that the predictors considered are not IID, and exhibit trends, dependence, and seasonality. Hence, these descriptive statistics are excluded from this analysis.

3.5 Results

3.5.1 Detrending and De-seasonalizing Results

The first step in the forecasting procedure amounts to the detrending and de-seasonalizing of each predictor. As mentioned in Section 3.3.2. above, this is accomplished through the estimation of autoregressive models for each predictor. See Appendix D for the estimated autoregressive equations and mis-specification testing results for the natural log of GDP.

After estimating the autoregressive equations for each series, the data series is modeled

[10] Both the GDP deflator and CPI were considered for this conversion. However, since the CPI is already included as a predictor, the GDP deflator was selected. Selecting the CPI to use for the conversion did not significantly change the preliminary results below. Moreover, it is important to note that the GDP deflator was seasonally adjusted. This adjustment had very little effect on the results.

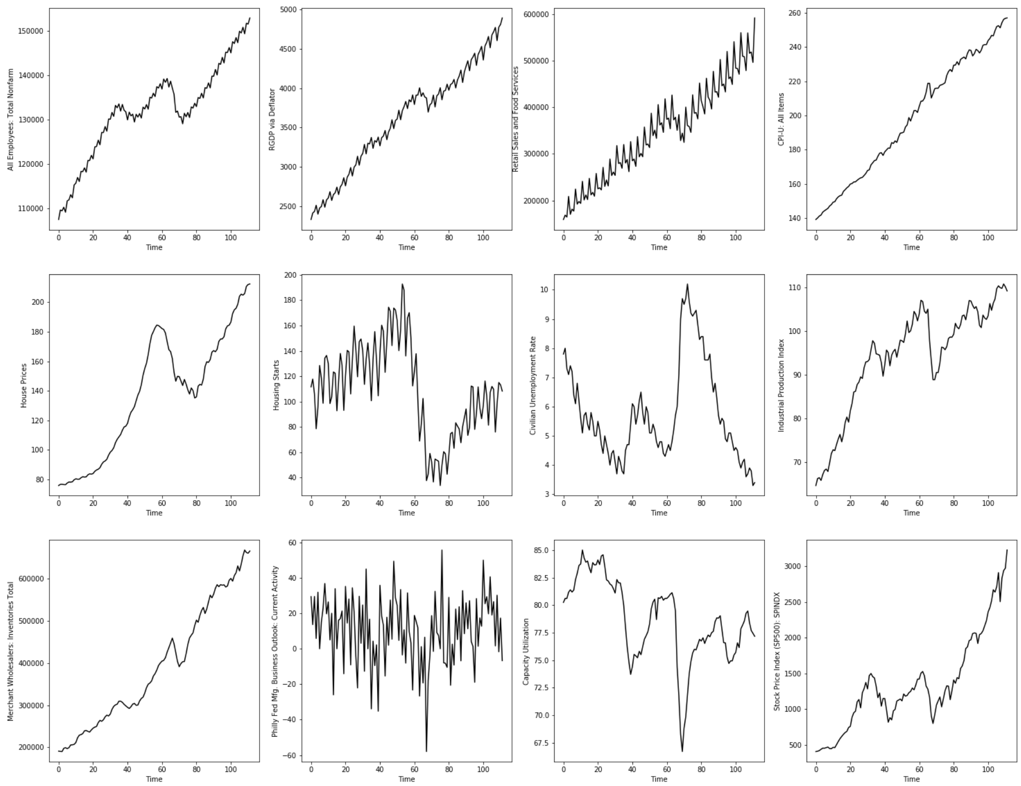


Figure 3.1: Time Series Plots: Original Predictors

together via a VAR model. As described in the forecasting methodology section, the number of trends and lags in the VAR(L) model are determined via statistical adequacy. Based on the testing methodology outlined in Poudyal and Spanos (2022), three lags and third degree trends were selected for the estimation of the 'temporal' principal components.[11] At the heart of the prescribed joint M-S tests for the VAR(L) model lies the following auxiliary regressions estimated for each equation:

[11] Performing the M-S testing on each equation separately might reveal that different equations require a distinct combination of trends and number of lags to achieve statistical adequacy. Therefore, in order to achieve statistical adequacy for the system of equations, the minimum trend degree and number of lags which render all of the equations statistically adequate is selected.

$$\widehat{u}_t = \alpha_0 + \beta_1\hat{y}_t + \overbrace{\beta_2\hat{y}_t^2}^{[2]} + \overbrace{\beta_3\hat{u}_{t-1} + \beta_4\hat{u}_{t-2}}^{[4]} + \overbrace{\gamma_1 t_*^4 + \gamma_2 t_*^5}^{[5]} + v_{1,t}$$

$$\widehat{u}_t^{\,2} = \delta_0 + \overbrace{\delta_1 t_*^1 + \delta_2 t_*^2}^{[5]} + \overbrace{\psi_1\hat{y}_t^2}^{[3]} + \overbrace{\omega_1\hat{u}_{t-1}^2 + \omega_2\hat{u}_{t-2}^2}^{[4]} + v_{2,t}.$$

Note that in this case, the fitted values are used in lieu of the original specification in order to save on degrees of freedom. Because the data is aggregated to the quarterly frequency, and the number of observations is reduced, the total number of degrees of freedom are limited. Hence, including the original specification will rapidly consume degrees of freedom and produce questionable results. For robustness, various forms of the auxiliary regressions were explored for each of the auxiliary regressions, as done in Poudyal and Spanos (2022).

Table 3.5 summarizes the joint M-S tests in terms of the assumptions being tested and corresponding null hypotheses. Also, observe that regardless of the form of the auxiliary regression being selected, all five of the model assumptions for the VAR(L) model, which can be found in Table 3.1, are being tested; see Spanos (2018) for more details.

Importantly, the assumptions of normality and dynamic heteroskedasticity are not examined here, but can be analyzed using the Student's t VAR(L) model; see Poudyal and Spanos (2022) for details. F-statistics and corresponding p-values resulting from the joint mis-specification tests are given in Table 3.6. Here, each column represents an individual equation in the VAR(3) system. Statistically significant test results, evaluated at the $\alpha = 0.01$ level, are given in bold font.[12]

The results in Table 3.6 indicate that the VAR(3) model with cubic trends and trigonometric

[12]Following Spanos (2014) and Poudyal and Spanos (2022), the significance level is chosen to be $\alpha = 0.01$ due to the number of observations available. Ultimately, there is an inverse relationship between the number of observations and the p-value; as the number of observations increases, the p-value shrinks. Thus, the power of the test to detect tiny discrepancies from the null hypothesis increases with the number of observations.

Table 3.5: M-S Tests: Normal VAR(3) Model

Assumptions	**Null Hypothesis**
Joint (Regression Function)	H_0: $\beta_2 = \beta_3 = \beta_4 = \gamma_1 = \gamma_2 = 0$
Linearity	H_0: $\beta_2 = 0$
Markov(3)	H_0: $\beta_3 = \beta_4 = 0$
1st t-invariance	H_0: $\gamma_1 = \gamma_2 = 0$
Joint (Skedastic Function)	H_0: $\delta_1 = \delta_2 = \psi_1 = \omega_1 = \omega_2 = 0$
Homoskedasticity	H_0: $\psi_1 = 0$
Dynamic Heteroskedasticity	H_0: $\omega_1 = \omega_2 = 0$
2nd t-invariance	H_0: $\delta_1 = \delta_2 = 0$

polynomials for seasonality is statistically adequate. Hence, the extended sample covariance matrix, including three lags for each predictor, will be used to formulate the temporal principal components.

Table 3.6: M-S Test Results for VAR(3) Model

Equation	$lngdp_t$	$lnnonfarm_t$	$lnretail_t$	$lncpi_t$	$lnhp_t$	$lnhs_t$
Joint (regression)	0.42[.8318]	0.21[.9562]	0.14[.9827]	0.12[.9870]	0.38[8607]	0.91[.4779]
Linearity	0.00[.9969]	0.37[.5446]	0.56[.4561]	0.17[.6790]	0.00[.9667]	3.76[.0553]
Dependence (Mean)	1.01[.3695]	0.16[.8544]	0.02[.9836]	0.12[.8838]	0.80[.4525]	0.47[.6250]
1st t-invariance	0.05[.9482]	0.12[.8829]	0.04[.9577]	0.18[.8382]	0.15[.8645]	0.50[.6071]
Joint (skedastic)	2.56[.0318]	2.68[.0255]	1.33[.2565]	0.69[.6343]	1.67[.1489]	0.53[.7527]
Homoskedasticity	1.88[.1729]	0.80[.3725]	0.42[.5175]	0.17[.6818]	0.09[.7674]	0.12[.7272]
Dependence (Variance)	2.31[.1045]	4.80[.0102]	2.48[.0888]	0.53[.5892]	2.90[.0598]	1.08[.3422]
2nd t-invariance	1.85[.1620]	3.26[.0424]	1.11[.3341]	1.38[.2574]	0.55[.5814]	0.31[.7335]

Equation	$lnip_t$	$lninv_t$	$lnsp500_t$	$outlook_t$	$cap.util_t$	ur_t
Joint (regression)	0.07[.9963]	0.09[.9944]	0.25[.9406]	0.06[.9972]	0.12[.9888]	0.53[.7552]
Linearity	0.00[.9681]	0.00[.9966]	0.06[.8147]	0.11[.7386]	0.09[.7706]	0.36[.5513]
Dependence (Mean)	0.11[.8993]	0.18[.8396]	0.59[.5589]	0.10[.9079]	0.01[.9911]	1.15[.3199]
1st t-invariance	0.09[.9240]	0.03[.9671]	0.02[.9808]	0.01[.9894]	0.12[.8839]	0.02[.9822]
Joint (skedastic)	1.14[.3465]	1.40[.2291]	0.73[.6022]	1.07[.3838]	0.64[.6731]	0.77[.5743]
Homoskedasticity	2.56[.1124]	0.19[.6635]	1.72[.1926]	0.39[.5319]	0.27[.6031]	0.82[.3687]
Dependence (Variance)	1.29[.2797]	3.35[.0389]	0.82[.4430]	0.29[.7481]	0.43[.6488]	0.84[.4357]
2nd t-invariance	2.13[.1246]	0.14[.8723]	0.82[.4446]	1.89[.1560]	1.41[.2494]	1.21[.3011]

In summary, the estimated autoregressive equations and VAR model equations suggest the use of three lags, the first three scaled, Chebyshev polynomials, the seasonal trigonometric polynomial, as well as dummies for the outliers at Q3 2008 and Q4 2008 (D67 and D68, respectively) to capture the trends and dependence.[13] Using this information, each series is detrended according to the methodology prescribed above; plots of the detrended series can found in Figure 3.2. After subtracting out the trends, the lags of the detrended data series are added back into the dataset.

3.5.2 Principal Components via the Covariance Matrix

Principal components are formed for both the estimated covariance matrix, $\hat{\mathbf{\Sigma}}_y$, and the estimated standardized correlation matrix, $Corr(\hat{\mathbf{Y}}_t(d))$ following the steps outlined in the methodology section. Here, the threshold for total explained variance is set to 95%. However, this level of explained variance is subject to change, as higher thresholds increase the number of components needed.[14] Scree plots and plots for percentage of variance explained at this threshold are given in Figure 3.3.

The proportion of total variance explained and the scree plots above indicate that the first 9 eigenvalues and corresponding eigenvectors should be retained to form principal components when using the covariance matrix, whereas the standardized correlation matrix utilizes the first 14 eigenvalues and corresponding eigenvectors. Note that in the case of the standardized correlation matrix, the number of principal components needed to achieve the same

[13] Outliers were only considered if they were statistically significant and affected the number of trends and lags for the system. For example, although the autoregressive model for the natural log of GDP indicated that the observation dummy during the 2008 financial crisis was unnecessary, the dummy was kept because there was an effect in other equations. Moreover, in the forecasting stage, the dummy variables were removed entirely, since omitting the effects from observations unnecessarily will unduly affect forecast performance; Lenza and Primiceri (2022) serves as a recent example on how removing the observations from the COVID-19 pandemic period caused underestimation of uncertainty.

[14] Various thresholds will be considered in the results that follow.

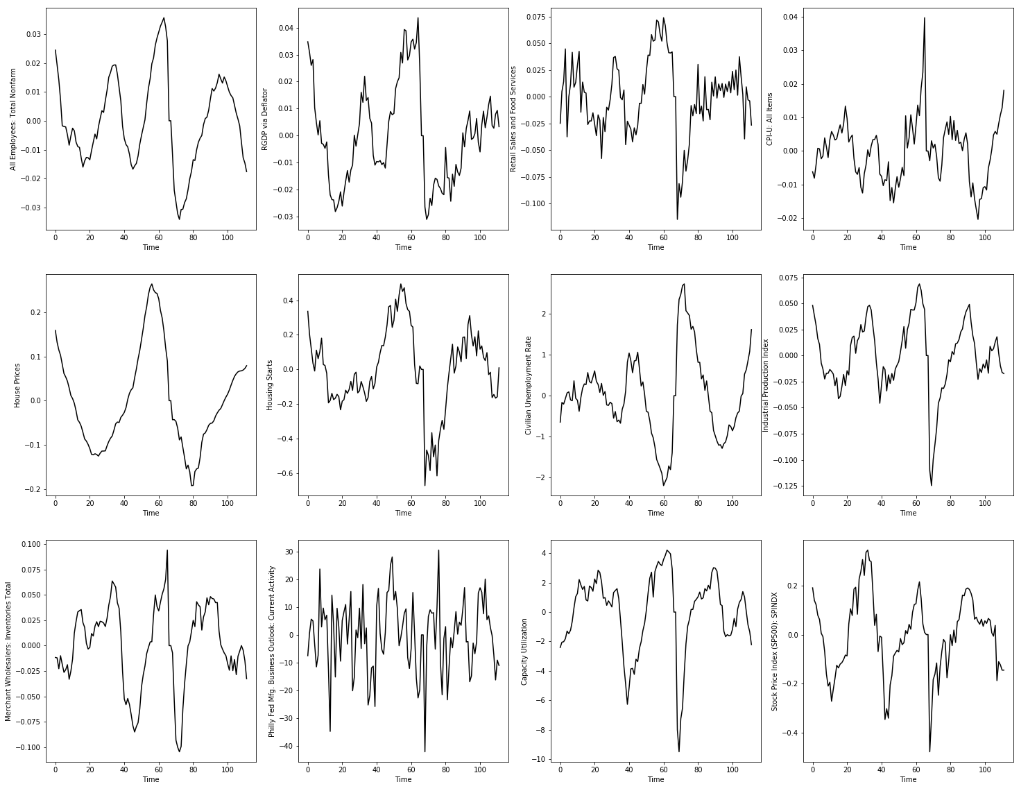


Figure 3.2: Time Series Plots: Detrended and De-seasonalized Predictors

percentage of total variance explained will be higher than the number needed when using the covariance matrix; see Jolliffe (2002) for details. Principal components are then calculated via the method described in Section 3.2.2.

Since the units of measurement for the various predictors were different, the principal components derived from the covariance matrix were dominated by the predictor with the largest sample variance (Jolliffe, 2002). In this case, it was the Philly Manufacturing Survey which had the largest variance, resulting in a strong correlation between this predictor and the first few principal components. Hence, from this point forward, this chapter only considers the

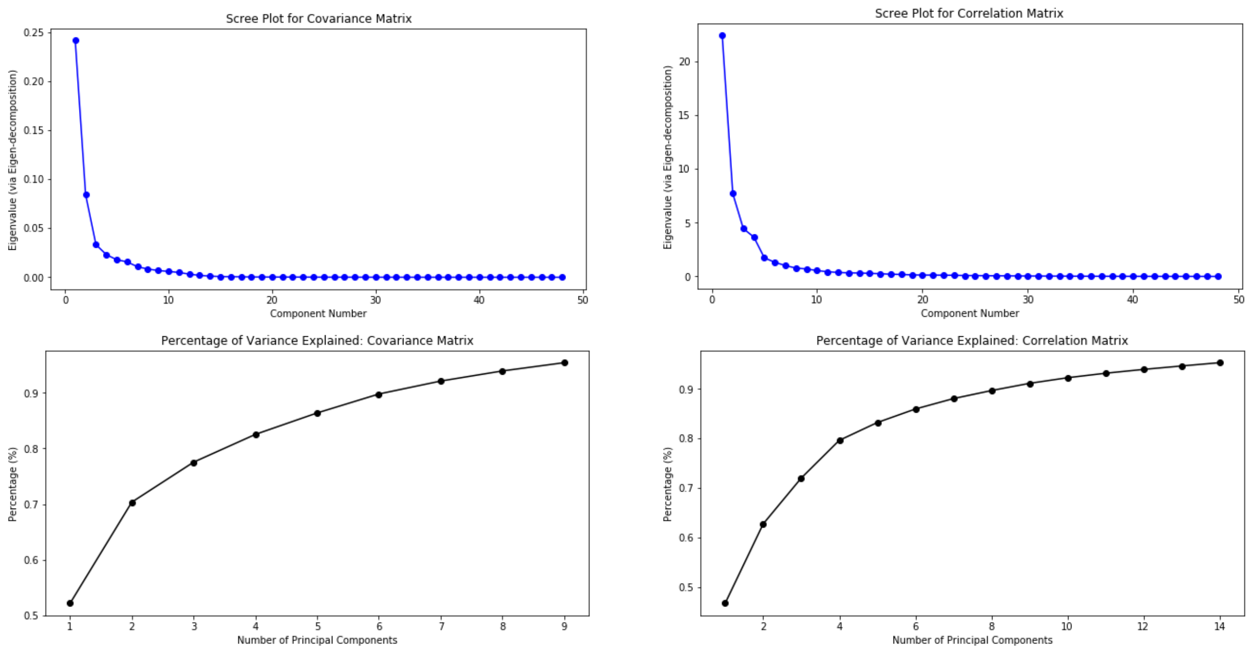


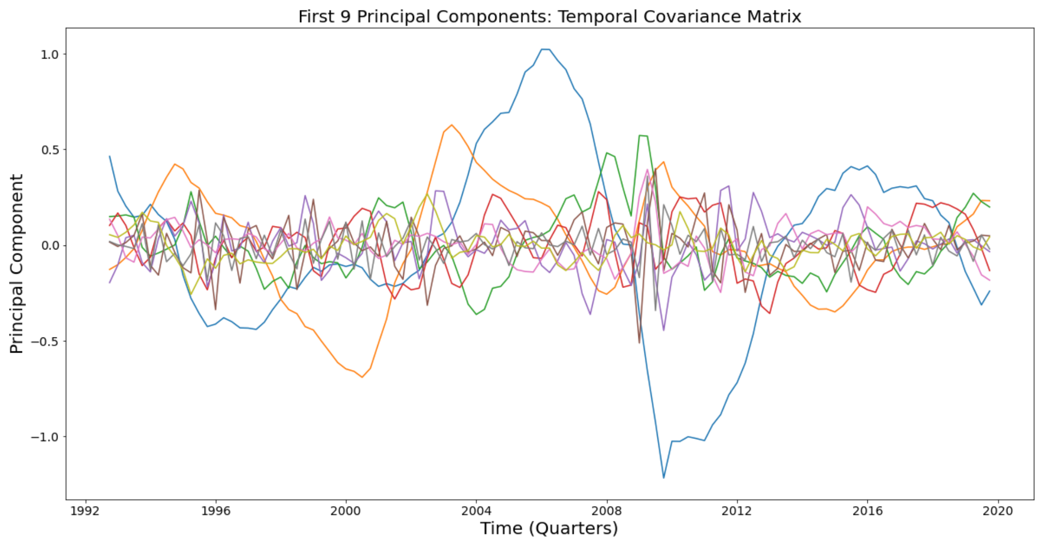
Figure 3.3: Scree Plots and Explained Variance Plots for Covariance and Correlation Matrices at 95% Threshold

standardized correlation matrix when units of measurement vary. The differences between the principal components derived from the covariance matrix and those derived from the standardized correlation matrix are clearly depicted in Figure 3.4.

3.5.3 Dropping Insignificant Principal Components

Next, these principal components are used as the explanatory variables in the dynamic factor augmented autoregressive model described above, where the dependent variable is the data series to be predicted. Principal components are kept if the p-value from the two-sided test indicates that the corresponding coefficient is statistically significant at the 2.5% level.[15] For the remainder of this chapter, the target variable will be the natural log of GDP, but the framework can be applied to all of the variables in the system. In the case of the natural

[15] Other significant levels could be selected depending on how strict one desires the threshold and cutoff to be.



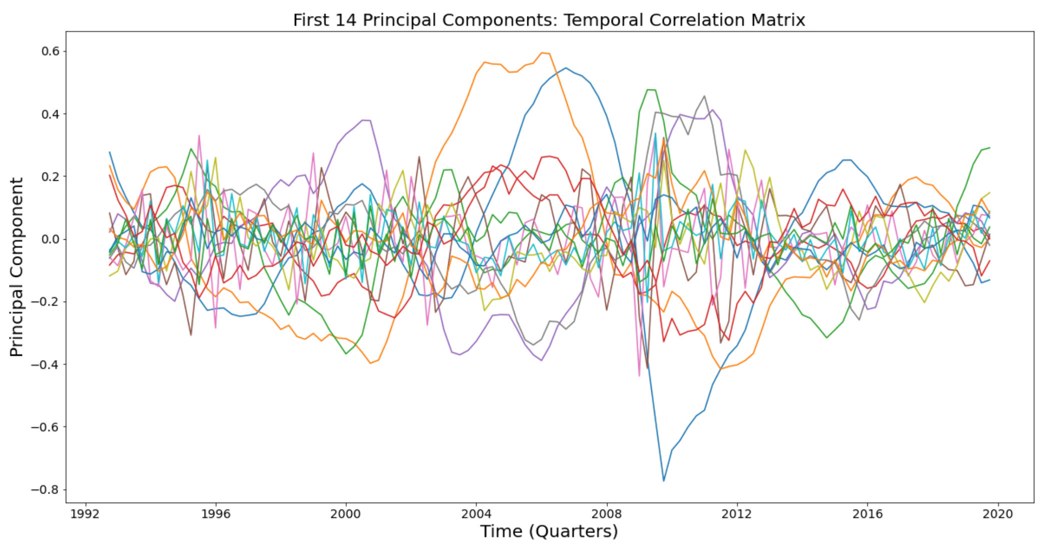


Figure 3.4: Time Series Plots of Principal Components

log of GDP, Table 3.6 and Table 3.7 show how altering the threshold for explained variance affects which principal components should be dropped.

The results in these tables indicate that, at the 2.5% significance level, various principal components are dropped for both the covariance matrix and the standardized correlation

Table 3.7: FAR(1,3) Dropping Statistically Insignificant Principal Components: Covariance Matrix

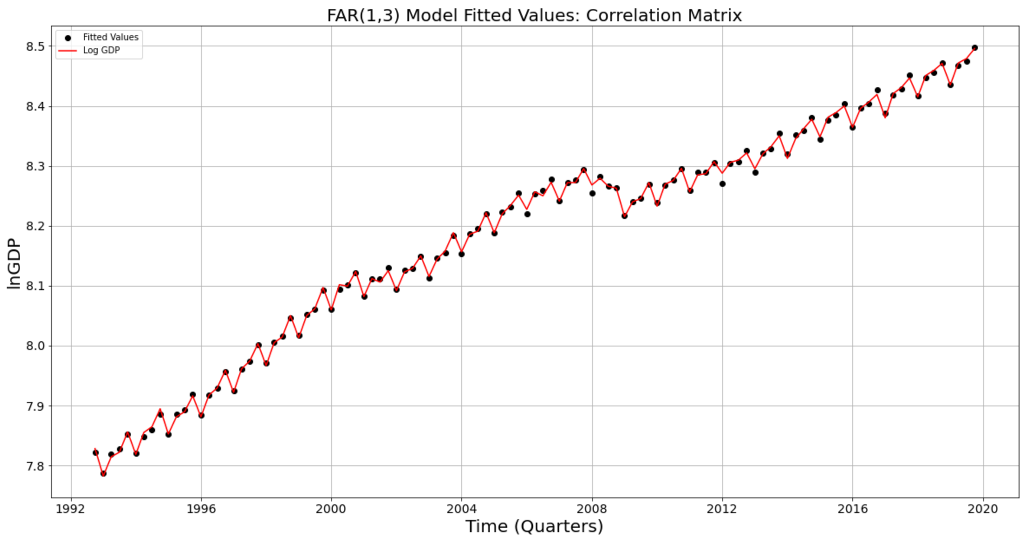
Threshold	Number of PCs Needed	Dropped PCs (2.5% Level)
70%	2	PC2
75%	3	PC2, PC3
80%	4	PC2, PC3, PC4
85%	5	PC2, PC3, PC4, PC5
90%	7	PC3
95%	9	PC3, PC8, PC9

Table 3.8: FAR(1,3) Dropping Statistically Insignificant Principal Components: Correlation Matrix

Threshold	Number of PCs Needed	Dropped PCs (2.5% Level)
70%	3	PC1, PC3
75%	4	PC1, PC2, PC3, PC4
80%	5	PC3, PC4
85%	6	PC3, PC4
90%	9	PC3, PC4, PC6, PC8
95%	14	PC4, PC6, PC8, PC11, PC14

matrix. As mentioned above,however, since the units of measurement differ, the standardized correlation matrix should be considered. Moreover, these results imply that the statistically significant principal components, i.e. the ones which remain, will differ depending on the threshold for explained variance.

Predicted values and the corresponding residual plots for the cases where statistically insignificant principal components are dropped are shown in Figure 3.5, in the case where the threshold for explained variance is 95%. Moreover, plots have been included below to emphasize the impact that different thresholds for explained variance have on the fitted values and corresponding residuals in the case where principal components have been dropped. Note that the thresholds for the proportion of explained variance are the same as those included when dropping principal components in Table 3.6 and Table 3.7.



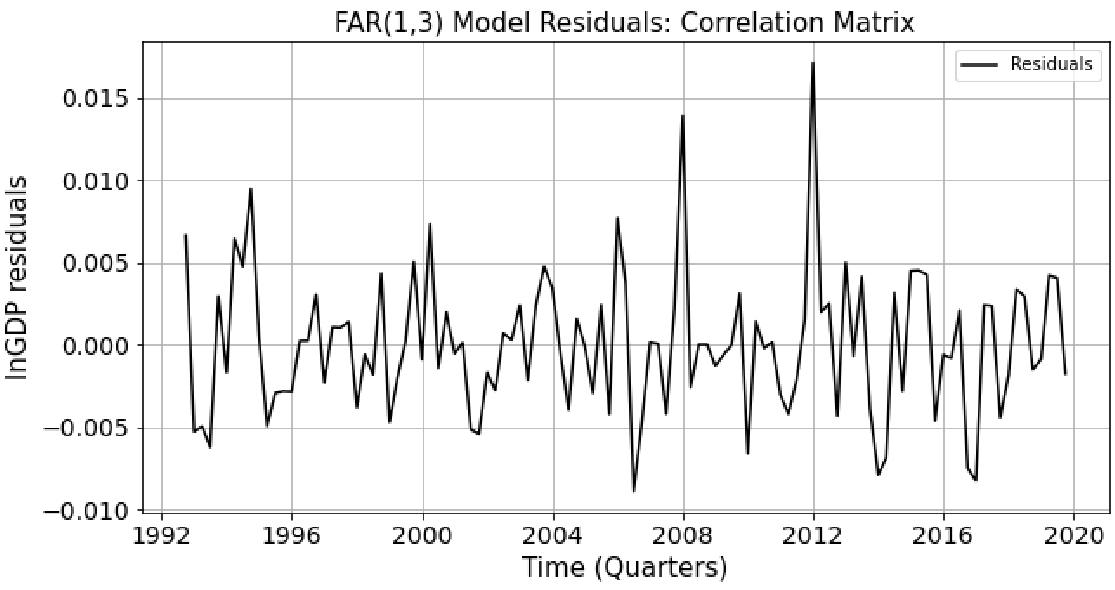


Figure 3.5: Fitted Values and Residuals using Factor Augmented Autoregressive FAR(1,3) Model: Insignificant Principal Components Omitted

A comparison of fitted values and residuals across the different thresholds for total variance explained is given in Figure 3.6. The figure only contains the results for the standardized correlation matrix and is enlarged to highlight the differences in each scenario.

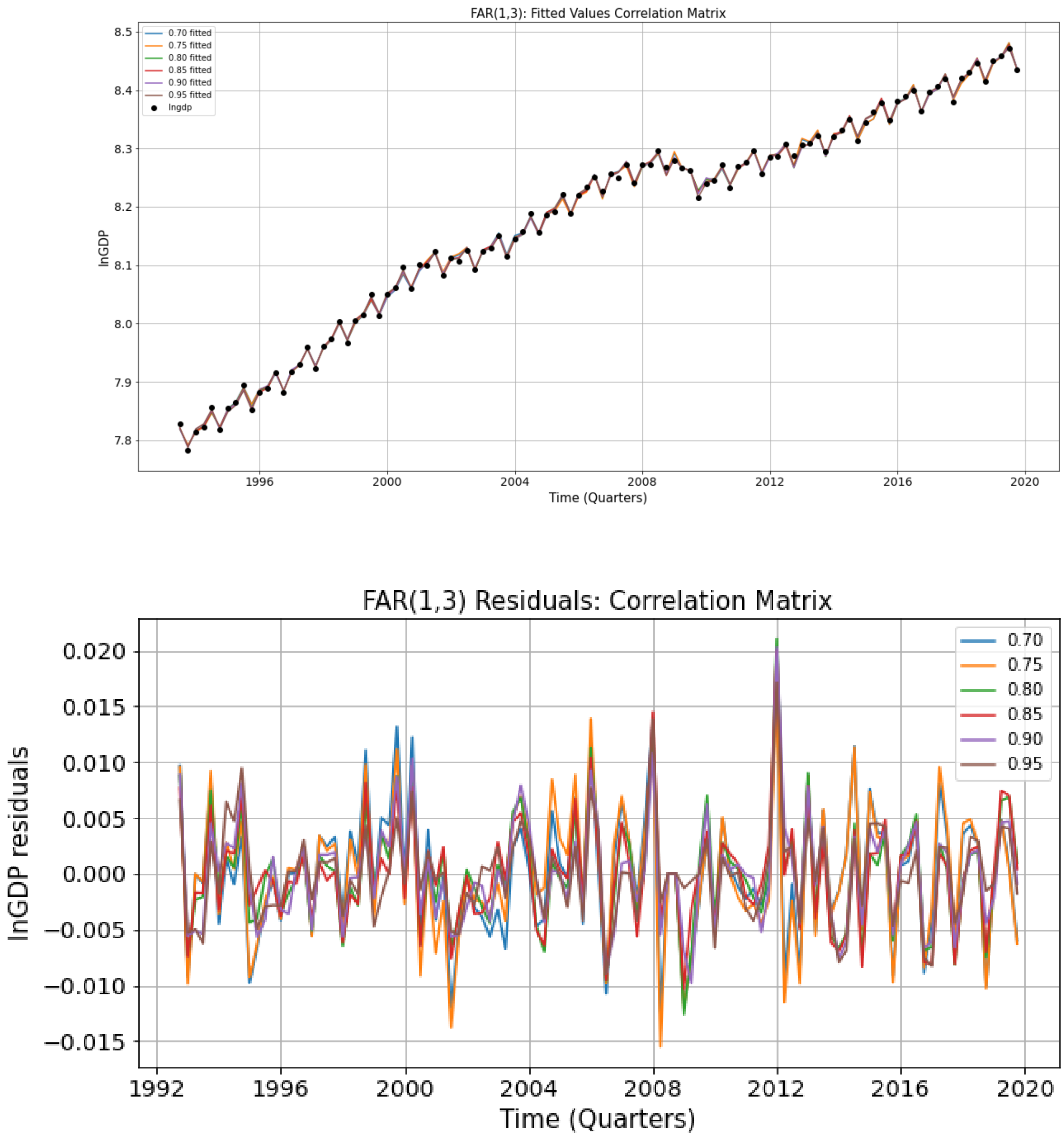


Figure 3.6: FAR(1,3) Forecasting Results by Proportion of Variance Explained

The comparisons in Figure 3.6 show that although the fitted values do not appear to change drastically from one threshold to the next, minor differences are observed in the residuals. This indicates that the autoregressive model is capturing slightly different information when the threshold for explained variance changes. Whether this presents an issue in terms of statistical adequacy requires extensive M-S testing as described in Section 3.3.7 in the forecasting methodology. Moreover, Figure 3.7 illustrates the impact that dropping principal components has on the estimated residuals. In the end, these plots demonstrate that the estimated model, and subsequently the forecasts, will be affected by the decision of which principal components to retain in the model; different methodologies for selection will alter the final forecasting results.

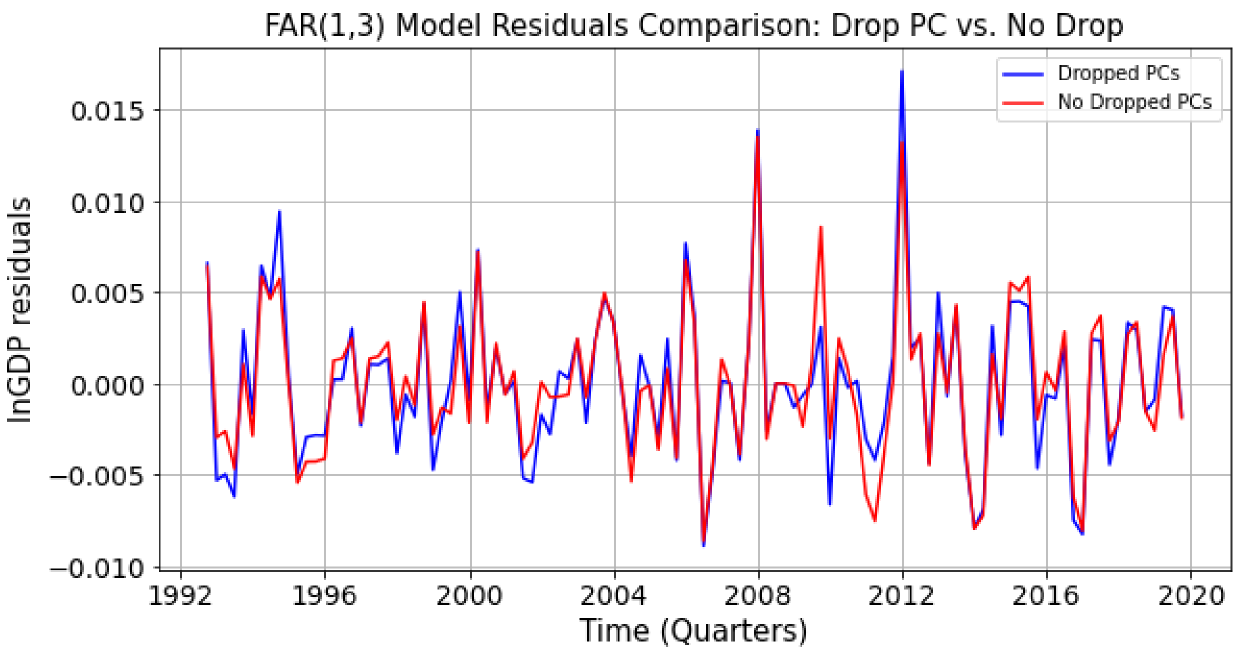


Figure 3.7: FAR(1,3) Residual Comparisons: Dropping Principal Components vs. Not Dropping, 95% Threshold

3.5.4 M-S Testing Results

Now that the FAR(1,3) has been estimated, M-S testing is conducted on the residuals via auxiliary regressions to ensure that the inclusion of the temporal principal components have not introduced any model mis-specification. The probabilistic assumptions being tested are those found in Table 3.9. Moreover, M-S tests results are displayed for two different cases:

the case where principal components from the correlation matrix are not dropped, and the case where principal components from the correlation matrix are dropped based on statistical significance. The auxiliary regressions for the FAR(1,3) model residuals for the statistically significant principal components formed from the standardized correlation matrix at the 95% threshold are:

$$\widehat{u}_t = \overbrace{a_0 + \sum_{i=1}^{3}\delta_i t_*^i + \sum_{i=1}^{2}(\alpha_i cos_{it} + \beta_i sin_{it}) + \psi_1 D67 + \psi_2 D68 + \sum_{i=1, i\neq 4,6,8,11.14}^{14} \gamma_i f_{it} + \zeta_1 lngdp_{t-1}}^{\text{original specification}}$$
$$+ \overbrace{a_1 t_*^4 + a_2 t_*^5}^{[5]} + \overbrace{\beta_3 \hat{y}_t^2}^{[2]} + \overbrace{\beta_4 \hat{u}_{t-1} + \beta_5 \hat{u}_{t-2}}^{[4]} + v_{1,t}$$

$$\widehat{u}_t^{\,2} = \delta_0 + \overbrace{\delta_1 t_*^1 + \delta_2 t_*^2}^{[5]} + \overbrace{\beta_1 \hat{y}_t^2}^{[3]} + \overbrace{\gamma_1 \hat{u}_{t-1}^2 + \gamma_2 \hat{u}_{t-2}^2}^{[4]} + v_{2,t}$$

Here, PC4, PC6, PC8, PC11 and PC14 are dropped from the model, as these terms were insignificant in the previous specification; see Table 3.8. Alternatively, to reduce the number of terms included in the testing equation, one can estimate the first auxiliary regression equation using fitted values as shown Poudyal and Spanos (2022) in the context of the Student's t Vector Autoregressive model:

$$\widehat{u}_t = \alpha_0 + \beta_1 \hat{y}_t + \overbrace{\beta_2 \hat{y}_t^2}^{[2]} + \overbrace{\beta_3 \hat{u}_{t-1} + \beta_4 \hat{u}_{t-2}}^{[4]} + \overbrace{a_1 t_*^4 + a_2 t_*^5}^{[5]} + v_{1,t}$$

The subsequent testing will utilize the abbreviated approach, i.e. it will use the fitted values, in order to reduce the number of terms in the auxiliary regression. Table 3.9 below summarizes the M-S tests in terms of the assumptions being tested and the explicit null hypothesis. Note that the third column shows the null hypothesis when the fitted values in the mean equation are substituted in for the original specification, and that the second variance auxiliary regression equation remains unchanged in both instances.

Table 3.9: M-S Tests: Factor Augmented Autoregression, Correlation Matrix

Assumptions	Null Hypothesis	Null Hypothesis (Fitted Values)
Joint (Regression Function)	H_0: $a_1 = a_2 = \beta_3 = \beta_4 = \beta_5 = 0$	H_0: $a_1 = a_2 = \beta_2 = \beta_3 = \beta_4 = 0$
Linearity	H_0: $\beta_3 = 0$	H_0: $\beta_2 = 0$
Dependence (Mean)	H_0: $\beta_4 = \beta_5 = 0$	H_0: $\beta_3 = \beta_4 = 0$
1st t-invariance	H_0: $a_1 = a_2 = 0$	H_0: $a_1 = a_2 = 0$
Joint (Skedastic Function)	H_0: $\delta_1 = \delta_2 = \beta_1 = \gamma_1 = \gamma_2 = 0$	H_0: $\delta_1 = \delta_2 = \beta_1 = \gamma_1 = \gamma_2 = 0$
Heteroskedasticity	H_0: $\beta_1 = 0$	H_0: $\beta_1 = 0$
Dependence (Variance)	H_0: $\gamma_1 = \gamma_2 = 0$	H_0: $\gamma_1 = \gamma_2 = 0$
2nd t-invariance	H_0: $\delta_1 = \delta_2 = 0$	H_0: $\delta_1 = \delta_2 = 0$

F-statistics and corresponding p-values are given in Table 3.10 below for the cases of both dropping and not dropping insignificant principal components. This is done for three different thresholds for the proportion of variance explained: 75%, 85%, and 95%. Bold values are those which are significant at the 2.5% level or lower. Note that an asterisk by the p-value indicates that one of the terms in the F-test had a statistically significant t-test result while the other did not. For example, if the additional time variable probing for mean heterogeneity t_*^4 was statistically insignificant but t_*^5 was not, then this will be noted as an asterisk in the table, specifically in the row "1st t-invariance." Ultimately, the results indicate that the specification for the factor autoregressive model is statistically adequate. This was true at the 95% thresholds for standardized correlation matrix. Due to the differing units of measurement, the M-S testing results for the covariance matrix are unreliable, as the principal components are dominated by the predictor with the largest sample variance. Hence, they are not included in Table 3.10.

Additionally, Table 3.10 shows the summary M-S test results for the FAR(1,3) at the 85% and 75% proportion of total explained variance, respectively. In general, these results are

Table 3.10: M-S Test Results for FAR(1,3) - Based on Proportion of Variance Explained (%)

	Not Dropping PC			Dropping PC		
Assumptions	95%	85%	75%	95 %	85 %	75%
Joint (regression)	0.58[.7150]	0.28[.9209]	1.50[.1974]	0.55[.7401]	0.58[.7189]	0.72[.6087]
Linearity	0.44[.5077]	0.19[.6642]	0.02[.8949]	0.41[.5255]	0.14[.7069]	0.27[.6062]
Dependence (Mean)	1.25[.2904]	0.31[.7363]	2.45[.0913]	0.81[.4475]	0.49[.6138]	1.22[.3010]
1st t-invariance	0.05[.9503]	0.17[.8467]	1.72[.1852]	0.43[.6516]	1.00[.3700]	0.36[.6958]
Joint (skedastic)	1.19[.3202]	1.60[.1682]	0.93[.4665]	0.65[.6596]	1.56[.1771]	0.82[.5354]
Heteroskedasticity	2.22[.1142]	1.44[.2329]	0.17[.6811]	2.40[.1245]	1.48[.2271]	0.09[.7598]
Dependence (Variance)	3.09[.0820]	1.84[.1642]	0.68[.5086]	0.25[.7794]	1.63[.2009]	0.27[.7658]
2nd t-invariance	0.50[.6066]	1.56[.2144]	0.72[.4898]	1.38[.2570]	1.54[.2198]	1.03[.3598]

consistent with the results for the 95% threshold; namely, there is little evidence of statistical mis-specification in both cases of retaining all temporal principal components and dropping statistically insignificant principal components. This comparison is in line with the comparison depicted in Figure 3.6, as changing the proportion of total variance explained did not appear to majorly alter the residuals of the estimated FAR(1,3) model.

3.6 Forecasting Results

3.6.1 FAR(1,3) Model Forecasts

Having verified the statistical adequacy of the FAR(1,3) model, in the case where the dependent variable is the natural log of GDP, out-of-sample forecasts are conducted using the model. Here, the forecasts are created via the statistical generating mechanism as shown in the forecasting methodology section. To illustrate the ability of the model to forecast over different periods into the future, multiple forecasting horizons are considered. Figure 3.8 below shows the forecast for the natural log of GDP with a forecasting horizon $h = 12$, i.e. the last twelve quarters. Note that the dashed lines on the graph refer to forecasts where principal components were constructed based on a particular proportion of variance explained.

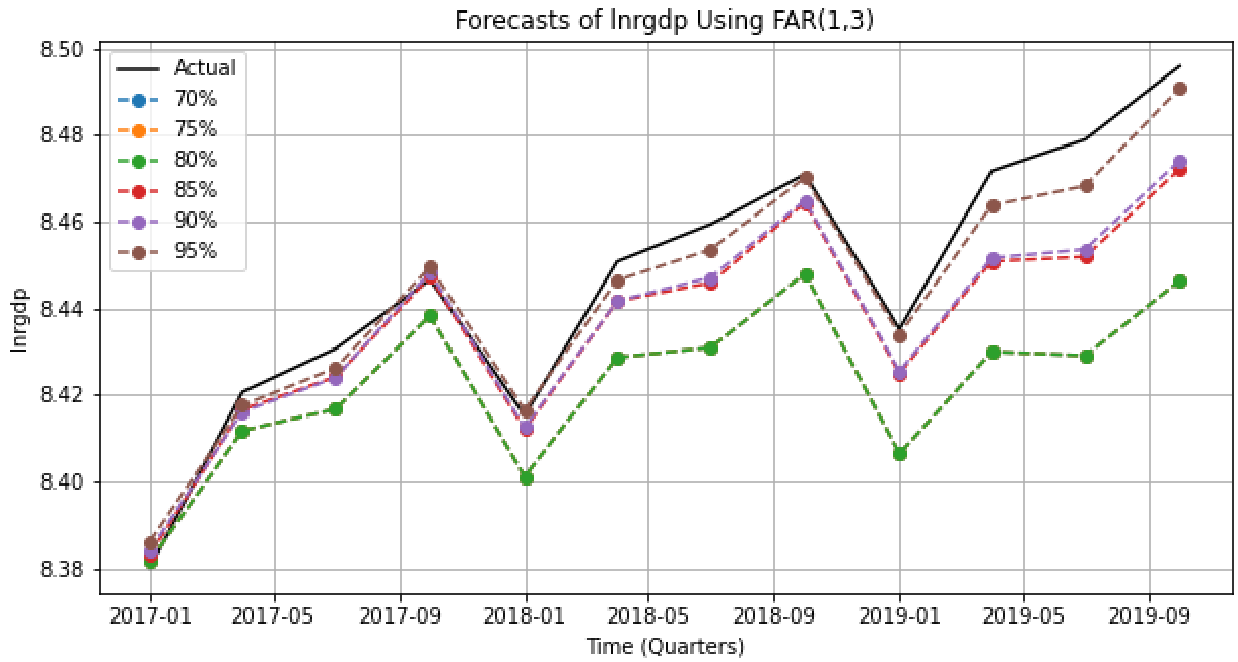


Figure 3.8: FAR(1,3) Forecasting Results by Proportion of Variance Explained

Unsurprisingly, increasing the proportion of variance explained corresponds to forecasts with improved predictability; forecasting accuracy will formally be evaluated in the subsequent section.

3.6.2 Benchmark Models: AR(p) Model

The natural choice of benchmark to compare the forecasting results of the FAR(p,L) with is the AR(p) model. This is due to the inherent relationship between the AR(p) and the FAR(p,L) model. In the case of the natural log of GDP, the AR(1) model was found to be statistically adequate in the methodology above; see Appendix D for details on the estimated model and mis-specification tests. Using Step 8 of the forecasting methodology section, the

AR(1) forecasts were constructed from the statistical generating mechanism. Figure 3.9 shows the forecasts over the forecasting horizon $h = 12$.

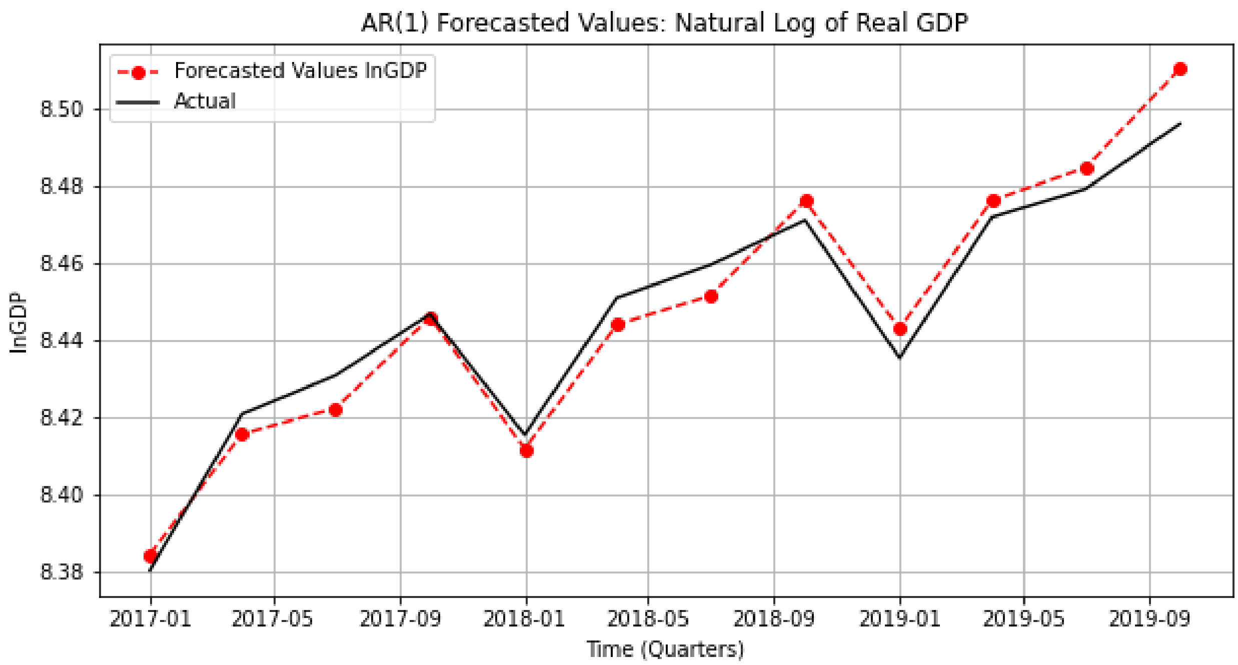


Figure 3.9: AR(1) Benchmark Forecasting Results

3.6.3 Forecasting Accuracy Evaluation

Figure 3.10 illustrates the difference in forecasts between the FAR(1,3) and the benchmark AR(1) model. From this plot, it is clear that while both models are statistically adequate, the additional information in provided by the temporal principal components results in a more accurate forecast over the specified horizon.

More formally, Table 3.11 below shows comparisons between the FAR(1,3) and AR(1) model in terms of common forecasting accuracy measurements, including mean-squared forecasting error (MSFE), root mean-squared forecasting error (RMSFE), and mean-absolute error (MAE). In this case each of these metrics is defined as:

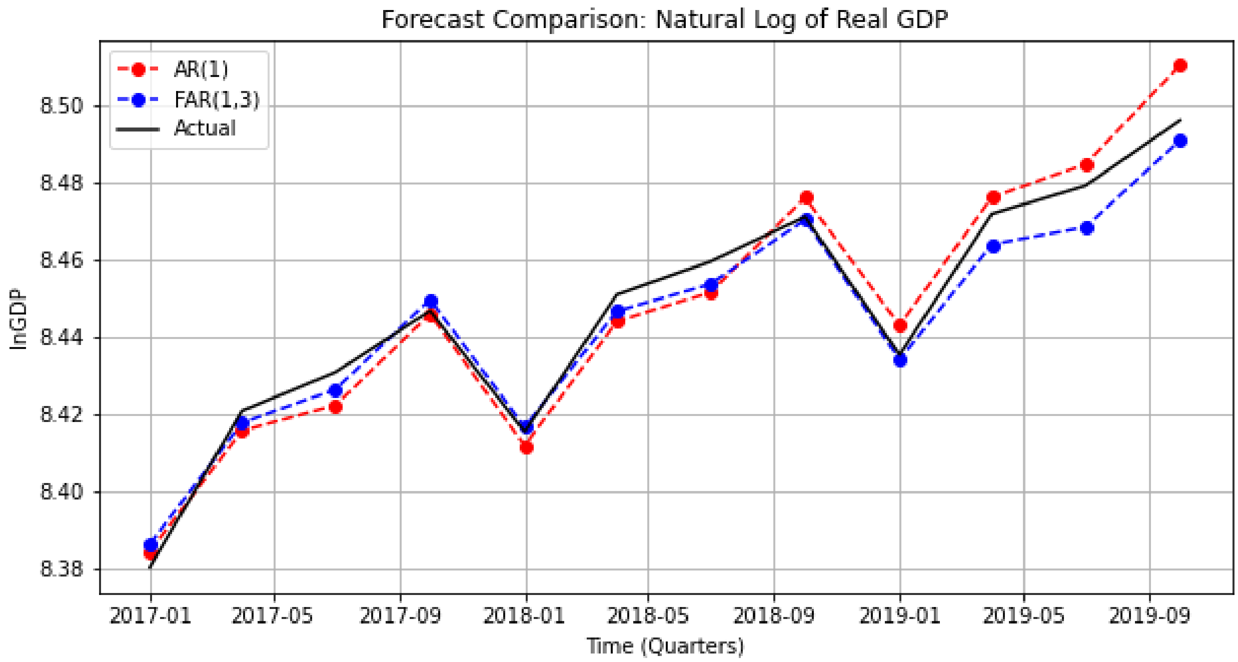


Figure 3.10: FAR(1,3) Forecasting Results Comparison: FAR(1,3) 95% and AR(1) Benchmark

$$MSFE = \frac{1}{n}\sum_{i=1}^{n}(y_{t+i} - \hat{y}_{i+i})^2$$

$$RMSFE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{t+i} - \hat{y}_{t+i})^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_{t+i} - \hat{y}_{t+i}|$$

Other than these common distance metrics, one could also use popular methods, such as the Diebold-Mariano Test, (Diebold and Mariano, 1995) to compare forecasts. See Buturac (2022) for more information on prevalent measures used to assess forecasting results.

Once again, Table 3.11 indicates that, in terms of the chosen forecasting metrics, the

Table 3.11: Natural Log of RGDP Forecast Evaluation

	Horizon				
FAR(1,3) 75%-80%	h = 1	h = 2	h = 4	h = 8	h = 12
MSFE	2.33E-06	4.11E-05	8.63E-05	0.0003	0.0008
RMSFE	0.0015	0.0064	0.0093	0.0172	0.0287
MAE	0.0015	0.0052	0.0082	0.0151	0.0242
FAR(1,3) 85%	h = 1	h = 2	h = 4	h = 8	h = 12
MSFE	8.88E-06	1.25E-05	1.69E-05	4.85E-05	0.0002
RMSFE	0.0030	0.0035	0.0041	0.0070	0.0136
MAE	0.0030	0.0035	0.0035	0.0058	0.0107
FAR(1,3) 90%	h = 1	h = 2	h = 4	h = 8	h = 12
MSFE	1.58E-05	1.82E-05	2.09E-05	4.53E-05	0.0002
RMSFE	0.0040	0.0043	0.0046	0.0067	0.0129
MAE	0.0040	0.0043	0.0042	0.0059	0.0103
FAR(1,3) 95%	h = 1	h = 2	h = 4	h = 8	h = 12
MSFE	3.60E-05	2.27E-05	1.85E-05	1.61E-05	2.79E-05
RMSFE	0.0060	0.0048	0.0043	0.0040	0.0053
MAE	0.0060	0.0045	0.0041	0.0036	0.0045
Benchmark AR(1)	h = 1	h = 2	h = 4	h = 8	h = 12
MSFE	5.65E-05	3.81E-05	1.61E-05	6.46E-05	4.82E-05
RMSFE	7.52E-03	6.17E-03	4.02E-03	8.04E-03	6.94E-03
MAE	0.0075	0.0049	0.0027	0.0073	0.0062

FAR(1,3) outperforms the benchmark AR(1) model at the 95% threshold. Ultimately, not only has the dependence information been summarized in the temporal principal components, but the addition of those temporal principal components has further augmented the forecasting accuracy of the standard autoregressive model. To reiterate, the forecasting methodology created above can be applied to any of the predictors in the model. For these cases, all that needs to be done is to switch out the dependent variable and use the statistically adequate autoregressive model as the baseline. This is due to the fact that the temporal principal components are constructed from a statistically adequate *vector* autoregressive model, and changing the dependent variable will not have methodological implications.

3.6.4 Benchmark Models: Dynamic Factor Models (DFM) and the Factor Augmented Vector Autoregressive Model (FAVAR)

The Dynamic Factor model (DFM) was first explored in Sargent and Sims (1977) and Geweke (1977), extending the classical factor model to cases where the factors evolve over time. However, as noted in the introduction to this chapter, the use of an index to summarize the information in several macroeconomic variables was posited as early as Burns and Mitchell (1946). Recently, the literature on DFMs has expanded, and now, DFMs are widely used to forecast (Stock and Watson, 2002a,b) and nowcast(Bok et al., 2018, Giannone et al., 2008) macroeconomic indicators; see Stock and Watson (2016) for a comprehensive summary of the literature surrounding this model. For this exercise, the second benchmark stems from an extension of the DFM, the factor augmented VAR model (FAVAR) model defined in Bernanke et al. (2005). The FAVAR model in Bernanke et al. (2005) is specified in the two following transition and observation equations:

$$\begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \Phi(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + u_t$$

$$X_t = \Lambda^f F_t + \Lambda^y Y_t + \epsilon_t$$

Here, Y_t is a $(K \times 1)$ vector denoting the macroeconomic variables which move the economy. F_t denotes an $(R \times 1)$ vector of unobserved factors, where R is less than the number of original data series, and X_t is the vector of predictors of length $(S \times 1)$ (2005). Moreover, $\Phi(L)$ is a lag polynomial similar to that found in Hamilton (1994) and Λ^f denotes the factor loadings. Estimation in Bernanke et al. (2005) is accomplished in two ways: a two-step procedure, where the unobserved factors are first estimated via principal components, and then the estimated factors are used to estimate a VAR model, and a one-step procedure, where the unobserved factors and the VAR model are estimated simultaneously (pp. 398-399). The former approach shares similarity with the forecasting methodology in Stock and Watson (2002b) and is the one considered for the benchmark. For more information on both the DFM in Stock and Watson (2002b) and the FAVAR in Bernanke et al. (2005), as well as the method for constructing the benchmark, see Appendix I. Figure 3.11 plots the forecasted FAVAR values against the seasonally adjusted data.

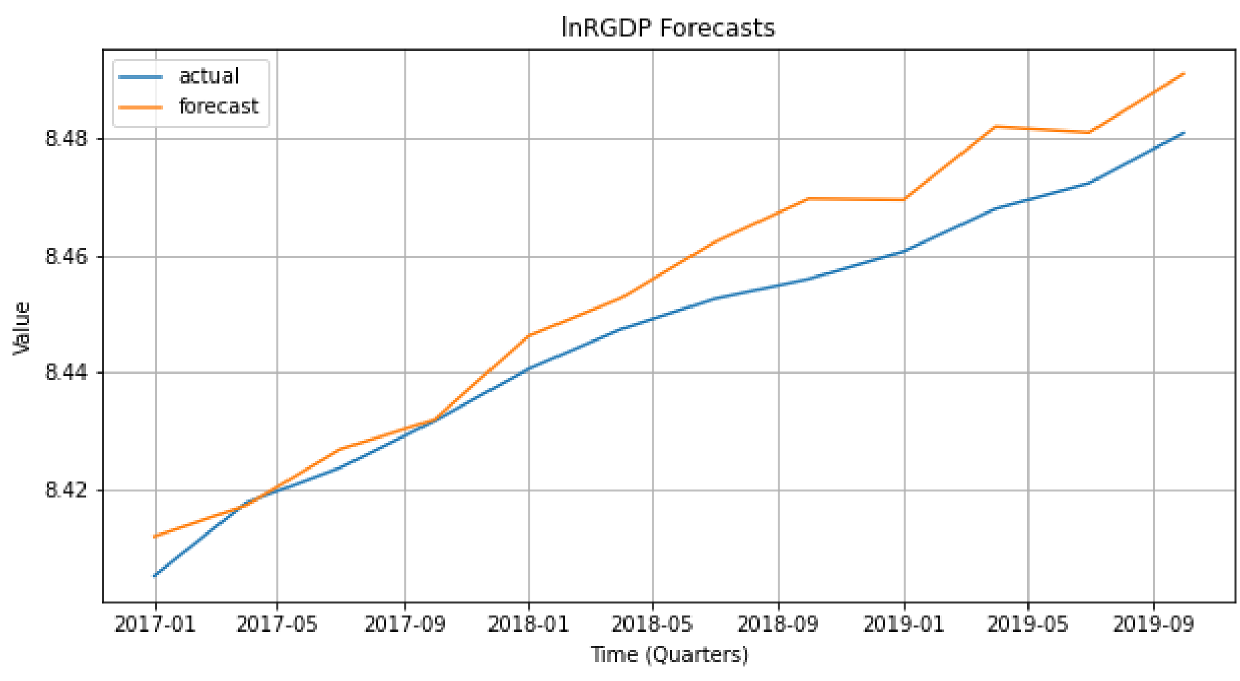


Figure 3.11: Benchmark FAVAR: Log-Difference RGDP Forecast

As shown in the plot, the model generally overpredicts, and the deviation between the forecasts and actual values increases over the forecasting horizon. This is indicative of a statistically inadequate model, where the model assumptions do not capture characteristics of the data. To emphasize the difference further, a comparison of FAR(1,3) to both benchmark models is provided in Table 3.12.

Table 3.12: Natural Log of RGDP Benchmark Forecast Evaluation

	Horizon				
FAR(1,3) 95%	h = 1	h = 2	h = 4	h = 8	h = 12
MSFE	3.60E-05	2.27E-05	1.85E-05	1.61E-05	2.79E-05
RMSFE	0.0060	0.0048	0.0043	0.0040	0.0053
MAE	0.0060	0.0045	0.0041	0.0036	0.0045
FAVAR 95%	h = 1	h = 2	h = 4	h = 8	h = 12
MSFE	4.52E-05	2.27E-05	1.39E-05	4.99E-05	7.11E-05
RMSFE	0.0067	0.0048	0.0037	0.0071	0.0084
MAE	0.0067	0.0036	0.0027	0.0056	0.0072
Benchmark AR(1)	h = 1	h = 2	h = 4	h = 8	h = 12
MSFE	5.65E-05	3.81E-05	1.61E-05	6.46E-05	4.82E-05
RMSFE	7.52E-03	6.17E-03	4.02E-03	8.04E-03	6.94E-03
MAE	0.0075	0.0049	0.0027	0.0073	0.0062

From the results depicted in the table, it is clear that the FAR(1,3) model produces superior forecasts in terms of accuracy when compared to the FAVAR model, which demonstrates its capability to compete and outperform other, commonly utilized models. As the forecasting horizon expands to twelve quarters, the deviation between the predictions and actual values increases for the FAVAR model to beyond that for the other two models being compared. This likely stems from two reasons, which are elaborated upon in more detail in Appendix I. The first is the seasonal adjustment. Since the purpose of including the FAVAR is to contrast the forecasting results when statistical adequacy is not taken into account, the methodology used to construct the FAVAR begins with the subtraction of the seasonal component, to emulate pre-adjusted data. This is similar to how papers, such as Bok et al. (2018), often begin their forecasts using data that has already been seasonally adjusted. Therefore, in terms of the overall results, the FAVAR cannot capture seasonal cycles, whereas the FAR(1,3) can. Secondly, the results in Table 3.12 indicate that the FAVAR model is statistically inadequate. After the forecasting horizon reaches a certain number of quarters, the forecasting accuracy deteriorates for the FAVAR much more quickly than the base model. Such deterioration may also be due to the large number of predictors used in the estimation step; i.e. the VAR model estimation. In the statistically adequate FAR(1,3), a procedure was outlined for eliminating irrelevant principal components, whereas all of the principal components needed to reach the desired threshold for explained variance are kept when estimating the FAVAR. For additional information on the benchmark FAVAR, see Appendix I.

3.7 Limitations and Future Research

3.7.1 COVID-19 Pandemic and Macroeconomic Forecasts

Although the forecasting results are certainly promising, there are some limitations to the current model and forecasting framework that should be addressed in future research. First and foremost, one potential concern when looking at this analysis is that the date range ends in 2019, conveniently before the first wave of the pandemic in the United States, which took place in spring 2020. Data after Q4 2019 is omitted due to a few reasons, chief among them being the pandemic.[16] As acknowledged in Ho (2021), forecasting macroeconomic variables during and after the start of the pandemic has caused a shift in how macroeconomic forecasts are developed, with no definitive answer consensus on how to incorporate such an unprecedented situation. Moreover, as noted in Ng (2021), there have only been two pandemics since the second world war, and neither one provides much insight on how the

[16] Another reason for the exclusion of more data is indirectly related to the pandemic, namely that the data is pandemic-period economic data is continuously being revised and updated; see the following article for a recent example in revisions of employment data: https://fredblog.stlouisfed.org/2022/01/revisions-to-employment-data-during-2020-and-2021/

shock induced by COVID-19 pandemic affects the macroeconomy.

One suggestion is to simply omit these problematic observations, rendering them effectively outliers. In Lenza and Primiceri (2022), however, the authors point out that while dropping the pandemic observations may be fruitful in estimating model parameters, it is ill-suited for forecasting purposes. Ng (2021) elaborates further on this sentiment, stating that the argument can be made that the "extreme values are not void of economic content" and therefore their effects should not be removed. In both of these papers, the authors develop their own models and methods to augment forecasts to include the pandemic, and thus, from the emerging methodological literature, it is clear that the incorporation of data from the pandemic period into macroeconomic models is no simple feat. Treatment of data from the COVID-19 pandemic has spurred a diverse response from modelers, and the methods employed are varied. In future research, it would be worthwhile to explore how well the model predicts the recession induced by the pandemic.

3.7.2 Going Beyond Normality

Another limitation to the current modelling methodology comes from the model assumptions themselves. Namely, the Dynamic Factor Augmented Autoregressive model assumes that the variables are normally distributed. The logical first step in extending the model beyond normality would be to construct temporal principal components the Student's t Vector Autoregressive model defined in Poudyal and Spanos (2022), as this extension allows for a heteroskedastic conditional variance. Departures from normality are even further expanded upon in Solat and Tsang (2021), where the authors introduce the notion of Generalized Principal Component Analysis. Although not studied in the current version of the FAR(p,L) model, these additions would be worth investigating in subsequent specifications.

Appendix H introduces the Student's VAR model and explores the augmented forecasting methodology in such cases.

3.7.3 Temporal Principal Component Construction

Finally, another potential concern lies in the construction of the temporal principal components. It is a typical assumption for principal components to have be centered, or to have zero mean. While detrending and demeaning of the original data was accomplished at the outset of the forecasting methodology, the lagged data columns no longer have a mean equal to zero. This in turn implies that those data columns are no longer centered for the principal components. To combat this, a constant was added to each of the data columns before the covariance matrix was estimated, thereby eliminating the missing value problem. However, it may be the case that this off-centering will significantly affect the temporal principal components and thus the forecasting results.

3.8 Discussion

In summary, the purpose of this chapter is is to study the use of principal components in the context of macroeconomic forecasting. To extend the traditional definition of principal components to account for not only contemporaneous dependence, but also temporal dependence, temporal principal components are first defined. These temporal principal components make up the baseline for the normal Dynamic Factor Augmented Autoregressive model defined in Table 3.4. Finally, a baseline forecasting methodology is constructed for the implementation of the model. The prescribed forecasting methodology is then applied to common macroeconomic and financial variables; the results of which are compared to commonly used forecasting benchmarks. Ultimately, the results of this exercise indicate that the statistically adequate FAR(1,3) model outperforms the benchmark AR(1) and FAVAR (Bernanke et al., 2005) models in terms of typical forecasting evaluation measures. More specifically, while it is found that the forecasting error grows as the forecasting horizon increases, this error is smaller than that in the found in the benchmark models for over longer horizons.

Chapter 4

Conclusion

4.1 Summary

In the context dynamic factor models, Stock and Watson (2017) acknowledges, "there is some evidence that the parametric restrictions (or priors) that make these methods work discard potentially important information" (p. 83). This statement directly aligns with the notion of statistical adequacy brought forth in Spanos (2009, 2018). If the model being used to forecast macroeconomic variables disregards key, statistical information which produced the data, then the forecasts will be unreliable in predicting the future. Hence, the role of this dissertation is to fill a niche, combining the aspects of one of the most popular dimension reduction techniques with a modeling philosophy that fosters learning from data. More specifically, the research presented here focuses on the assumptions underpinning the Principal Components model. Each chapter examines a different facet of modeling macroeconomic data with PCA, with several contributions to the literature listed below:

- First is the extension of the previous work done on linear regression models in (Spanos and McGuirk, 2001), Spanos (2009) and McGuirk and Spanos (2009) to Vector Autoregressive (VAR) models in cases of various forms of mean heterogeneity. The various simulations in Chapter 2 demonstrated the devastating impact that neglecting systematic information in the data, via a mis-specified model, can have on the reliability of inferences. Similar to the results in Spanos and McGuirk (2001), Spanos (2009) and McGuirk and Spanos (2009), in such cases, it is shown that the actual error probabilities can differ starkly from the nominal error probabilities when failing to model all of the relevant information in the dataset.

- Another key contribution is the establishment of a baseline forecasting framework for PCA which achieves superior forecasts in terms of both accuracy and in terms of capturing the probabilistic assumptions inherent in each macroeconomic data series. This forecasting framework requires statistical adequacy from the outset, which contrasts with other, widely used models such as the Dynamic Factor models found in Stock and Watson (2016). In order to create a statistically adequate model using principal components within this framework, a connection is recognized between PCA and the Normal VAR using the probabilistic reduction approach. This connection stems from the estimated joint covariance matrix of the Normal VAR model and forms what is called the Normal Temporal Principal Components model. Furthermore, the Dynamic Factor Augmented Autoregressive model (FAR(p,L)) is defined to incorporate the temporal Principal Components. This model is fully described with testable model assumptions.

- The forecasting methodology in Chapter 3 is then applied to a dataset consisting of commonly utilized macroeconomic indicators, derived from Bok et al. (2018), and it is found that the statistically adequate models perform better, in terms of forecasting accuracy, than the benchmark autoregressive model and FAVAR model (Bernanke et al., 2005). Hence, another contribution is to demonstrate how statistical adequacy can enhance forecast accuracy in the context of principal components and to highlight the potential consequences of neglecting critical, statistical information when forecasting macroeconomic data using dimension reduction techniques.

- Moreover, a similar forecasting framework which connects the Student's t VAR model and the Temporal Principal Components model is defined, extending the work of Poudyal and Spanos (2022) and Solat and Tsang (2021). While this framework needs to be expanded upon in future research, the initial steps can be found in Appendix H of Chapter 3.

4.2 Future Expansions

To augment the results given in the chapters above, there are several directions that this research can be taken in. First and foremost, the methodology discussed in the dissertation is applied to pre-pandemic data, with ending date to the data being Q4 2019. This was done intentionally, as the pandemic data lacks both precedent and a consensus as to how the macroeconomic data should be treated (Ho, 2021). Testing the forecasting methodology against an updated vintage would further serve to demonstrate the robustness of the methodology presented above. Moreover, given the high volatility in macroeconomic variables surrounding the pandemic, this also would serve as a great example for the Student's t VAR model forecasting methodology utilized in Poudyal and Spanos (2022), which is elaborated upon in Appendix H of Chapter 3.

Another aspect touched upon in Chapter 3 but needs further exploration is the idea of statistical adequacy in the context of seasonal adjustments. As shown in Ghysels and Osborn (2001), estimation of seasonal cycles can be achieved through the use of trigonometric polynomials. These trigonometric polynomials were chosen in the methodology presented in Chapter 3 to directly estimate the trends and avoid the potential pitfall of selecting an erroneous vector autoregressive model. In practice, however, it is often the case that forecasting either begins with data that has already been seasonally adjusted or data that is adjusted using the X-13 ARIMA-SEATS software; see Bok et al. (2018) as an example. At its core, the X-13 ARIMA-SEATS software estimates a seasonal ARIMA(p,d,q)(P,D,Q) model and then produces adjusted data omitting seasonal component. While the program requires expertise of the user to navigate the underpinning model, there is the possibility that the data is produced using a statistically mis-specified model. Spanos (2010b) showed that the common model selection procedures of AIC and BIC, which are both a part of the X-13 software, were unreliable in detecting a statistically adequate model. Therefore, a close examination of the two seasonal adjustment methodologies in the context methodologies may prove beneficial in further improving forecast accuracy.

Moreover, while the Dynamic Factor models (DFM) (Geweke, 1977, Sargent and Sims, 1977) and the Factor Augmented Vector Autoregressive (FAVAR) model (Bernanke et al., 2005) were mentioned in Chapter 3 and the latter was used as a benchmark, more work is required to delve into the statistical assumptions behind the models using the probabilistic reduction approach (Spanos, 2019a). These models have become widespread in their use, and the theory behind them has significantly evolved during the last few decades; see Bai and Ng (2002), Giannone et al. (2008), Stock and Watson (2002a) for example. Depending on the specification and type of dynamic factor model, as well as the number of predictors and observations, principal components can be used in to estimate the latent factors; see Bai and Ng (2002), Stock and Watson (2002a, 2016). Once the set of statistical assumptions has been defined for the DFMs and FAVAR, it would be enlightening to contrast these models and the Dynamic Factor Augmented Autoregressive model (FAR(p,L)) defined in Chapter 3 in terms of the inherent model assumptions. For example, chief among the differences is the treatment of the latent factors, where it is often assumed that the dynamic factors evolve according to an autoregressive or a vector autoregessive process (Forni et al., 2009, Stock and Watson, 2016). In the case of the FAVAR, specifically, although the name of the Dynamic Factor Augmented Autoregressive model from Chapter 3 sounds similar to the FAVAR model, the FAVAR augments a VAR model using the latent factors (Bernanke et al., 2005), which differs from the procedure outlined above.

To summarize, next steps for this research include:

1. Extending the dataset to forecast the COVID-19 pandemic period.
2. Employing the forecasting methodology to financial data series and exploring how different types of data changes the framework.
3. Expanding the framework to handle instances where the normality assumption is no longer valid, e.g. cases where the Student's t Vector Autoregressive model would be appropriate.
4. Enhancing the ability of the forecasting framework to capture various forms seasonality. This includes comparing a statistically adequate de-seasonalizing framework with conventional seasonal adjustment methods.
5. Further exploring the relationship between this framework and DFMs when viewed under the probabilistic reduction approach.

Bibliography

Hirotugu Akaike. Information theory and extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), *Second International Symposium on Information*, pages 267–281, 1973.

Theodore W. Anderson. *An Introduction to Multivariate Statistical Analysis.* Wiley Series in Probability and Statistics, third edition, 2003.

Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

Matteo Barigozzi, Marc Hallin, Stefano Soccorsi, and Rainer von Sachs. Time-varying general dynamic factor models and the measurement of financial connectedness. *Journal of Econometrics*, 222(1):324–343, 2019.

Richard E. Bellman. *Adaptive Control Processes: A Guided Tour.* Princeton University Press, 1961.

Ben S. Bernanke, Jean Boivin, and Piotr Eliasz. Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach. *The Quarterly Journal of Economics*, 120(1):387–422, 2005.

Brandyn Bok, Daniele Caratelli, Domenico Giannone, Argia M. Sbordone, and Andrea Tambalotti1. Macroeconomic nowcasting and forecasting with big data. *Annual Review of Economics*, 10:615–643, 2018.

David R. Brillinger. *Time Series- Data Analysis and Theory.* Holden-Day, San Francisco, 1981.

Arthur F. Burns and Wesley C. Mitchell. *Measuring Business Cycles*. NBER Book Series Studies in Business Cycles, 1946.

Goran Buturac. Measurement of economic forecast accuracy: A systematic overview of the empirical literature. *Journal of Risk and Financial Management*, 15(1):1–28, 2022.

Francis X. Diebold and Robert S. Mariano. Comparing forecast accuracy. *Journal of Business & Economic Statistics*, (13):253–265, 1995.

Mario Forni, Dominco Giannone, Marco Lippi, and Lucrezia Reichlin. Opening the black box: Structural factor models with large cross sections. *Econometric Theory*, (25):1319–1347, 2009.

John Geweke. The dynamic factor analysis of economic time series. *In: Aigner, D.J. and Goldberger, A.S., Eds., Latent Variables in Socio-Economic Models 1, North-Holland, Amsterdam*, 1977.

Eric Ghysels and Denise R. Osborn. *The Econometric Analysis of Seasonal Time Series*. Cambridge University Press, 2001.

Domenico Giannone, Lucrezia Reichlin, and David Small. Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, (55):665–676, 2008.

William H. Greene. *Econometric Analysis*. Pearson Education Inc., eighth edition, 2018.

James D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew

Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL https://doi.org/10.1038/s41586-020-2649-2.

Paul Ho. How macroeconomic forecasters adjusted during the covid-19 pandemic. *Federal Reserve Bank of Richmond: Economic Brief*, (21-19):645–651, 2021.

Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, 1933.

Ian T. Jolliffe. *Principal Component Analysis*. Springer, second edition, 2002.

Ian T. Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions Royal Society*, A(374):1–16, 2016.

Julie Josse and François Husson. Selecting the number of components in principal component analysis using cross-validation approximations. *Computational Statistic & Data Analysis*, 56(6):1869–1879, 2012.

Michele Lenza and Giorgio E. Primiceri. How to estimate a vector autoregression after march 2020. *Journal of Applied Econometrics*, 10(4):688–699, 2022.

Helmut Lüketpohl. *New Introduction to Multiple Time Series Analysis*. Springer, 2005.

Anya McGuirk and Aris Spanos. Revisiting error-autocorrelation correction:common factor restrictions and granger non-causality. *Oxford Bulletin of Economics and Statistics*, 71 (2):273–294, 2009.

Michael Michaelides and Aris Spanos. On modeling heterogeneity in linear models using trend polynomials. *Economic Modelling*, 85:74–86, 2020.

Serena Ng. Modeling macroeconomic variations after covid-19. *Columbia University and NBER Working Paper*, 2021.

Karl Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901.

Niraj Poudyal. *Confronting Theory with Data: the Case of DSGE Modeling*. PhD thesis, Virginia Tech, 2012.

Niraj Poudyal and Aris Spanos. Model validation and dsge modeling. *Econometrics*, 10(17), 2022.

Thomas J. Sargent and Christopher A. Sims. Business cycle modeling without pretending to have too much *a Prior* economic theory. *Federal Reserve Bank of Minneapolis: New Methods in Business Cycle Research*, 1977.

Gideon Schwarz. Estimating the dimension of a model. *Annals of Statistics*, (6):462–464, 1978.

Karo Solat. *Generalized Principal Component Analysis*. PhD thesis, Virginia Tech, 2018.

Karo Solat and Kwok Ping Tsang. Forecasting exchange rates with elliptically symmetric principal components. *International Journal of Forecasting*, 37(3):1085–1091, 2021.

Aris Spanos. *Statistical Foundations of Econometric Modeling*. Cambridge University Press, 1986.

Aris Spanos. On modeling heteroskedasticity: The student's t and elliptical linear regression models. *Econometric Theory*, 10(2):286–315, 1994.

Aris Spanos. Statistical misspecification and the reliability of inference: The simple t-test in the presence of markov dependence. *The Korean Economic Review*, 25(2):165–213, 2009.

Aris Spanos. Statistical adequacy and the trustworthiness of empirical evidence: Statistical vs. substantive information. *Economic Modelling*, 27(6):1436–1452, 2010a.

Aris Spanos. Akaike-type criteria and the reliability of inference: Model selection vs. statistical model specification. *Journal of Econometrics*, (158):204–220, 2010b.

Aris Spanos. Recurring controversies about p values and confidence intervals revisited. *Ecology*, 95:645–651, 2014.

Aris Spanos. Mis-specification testing in retrospect. *Journal of Economic Surveys*, 32(2): 541–577, 2018.

Aris Spanos. *Probability Theory and Statistical Inference: Empirical Modeling with Observational Data*. Cambridge University Press, second edition, 2019a.

Aris Spanos. Near-collinearity in linear regression revisited: The numerical vs. statistical perspective. *Communications in Statistics*, 48(22):5492–5516, 2019b.

Aris Spanos. Statistical modeling and inference in the era of data science and graphical causal modeling. *Journal of Economic Surveys*, 2021.

Aris Spanos and Anya McGuirk. The model specification problem from a probabilistic reduction perspective. *American Journal of Agricultural Economics*, 83(5):1168–1176, 2001.

Aris Spanos and Anya McGuirk. The problem of near-multicollinearity revisited: Erratic vs. systematic volatility. *Journal of Econometrics*, 108:365–393, 2002.

James H. Stock and Mark W. Watson. Forecasting using principal components from a large number of predictors. *Business Cycles, Indicators and Forecasting*, 97(460):1167–1179, 2002a.

James H. Stock and Mark W. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 23(6):405–430, 2002b.

James H. Stock and Mark W. Watson. Chapter 8 - dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. *Handbook of Macroeconomics*, 2A:415–525, 2016.

James H. Stock and Mark W. Watson. Twenty years of time series econometrics in ten pictures. *Journal of Economic Perspectives*, 31(2):59–86, 2017.

Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010. doi: 10.25080/Majora-92bf1922-00a.

Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analsysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286, 2006.

Appendices

Appendix A

Checking for Covariance-Stationarity of Simulated VAR(p)

A.1 Vector Autoregressive Model (VAR(p))

Suppose we have the following VAR(p):

$$\mathbf{y}_t = \mathbf{a} + \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \boldsymbol{\epsilon}_t,$$

where

$$E(\boldsymbol{\epsilon}_t) = \mathbf{0}$$

$$E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_\tau^\intercal) = \begin{cases} \boldsymbol{\Omega} & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases}$$

Note that unlike the specification above, here the VAR(p) model is viewed as a stochastic difference equation; see Hamilton (1994) for details. Much like the AR(p) process, one can rewrite the VAR(p) as the following VAR(1) (Hamilton, 1994; p. 259):

$$\mathbf{Z}_t = \mathbf{F}\mathbf{Z}_{t-1} + \mathbf{v}_t$$

where:

$$\mathbf{Z}_t = \begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-p+1} \end{bmatrix},$$

$$\mathbf{F} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & \dots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I}_n & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_n & \mathbf{0} \end{bmatrix}$$

$$\mathbf{v}_t = \begin{bmatrix} \boldsymbol{\epsilon}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

Note that $\mathbf{Z}_t$ is $(np \times 1)$, $\mathbf{F}$ is $(np \times np)$, and $\mathbf{v}_t$ is $(np \times 1)$. As shown in Hamilton (1994), in order for the VAR(p) to be covariance-stationary, then all of the the eigenvalues of $\mathbf{F}$ must lie within the unit circle, i.e. $|\lambda| < 1$, for all λ satisfying the following proposition (Hamilton, 1994):

Proposition: The eigenvalues of $\mathbf{F}$ satisfy

$$|\mathbf{I}_n\lambda^p - \mathbf{A}_1\lambda^{p-1} - \mathbf{A}_2\lambda^{p-2} - \ldots - \mathbf{A}_p| = 0.$$

Equivalently, the VAR is covariance stationary if all values of z satisfying:

$$|\mathbf{I}_n - \mathbf{A}_1 z - \mathbf{A}_2 z^2 - \ldots - \mathbf{A}_p z^p| = 0.$$

lie outside the unit circle.

A.2 Covariance-Stationary Check Python Code

```
#import packages
import numpy as np
import pandas as pd
import math

def reparam_cov_partition(SIGMA,k,p):

    l = p+1

    S00 = np.hsplit(np.vsplit(SIGMA,l)[0],l)[0]

    S01_dict = {}

    for i in range(1,l):

        S01_dict['S10%d' %(i)] = np.hsplit(np.vsplit(SIGMA,l)[0],l)[i]

    S01 = np.hstack(list(S01_dict.values()))

    S10 = S01.T

    S11_dict = {}

    for m in range(1,l):
        for n in range(1,l):
            S11_dict['S11%d%d' %(m,n)] = np.array(np.hsplit(np.vsplit(SIGMA,l)[m],l)[n])

    S11_values = list(S11_dict.values())

    S11 = [np.hstack(S11_values[int(len(S11_dict)/p)*s:int(len(S11_dict)/p)*(s+1)])
           for s in range(p)]

    S11 = np.concatenate(S11)

    return S00, S01, S10, S11
```

```
    def cov_check_var(SIGMA,k,p, output = False):
    # p is number of lags, k is number of variables, SIGMA must by (p+1)*k x (p+1)*k

        #Obtain partitioned covariance matrix
        S00, S01, S10, S11 = reparam_cov_partition(SIGMA, k, p)

        #Reparameterize to find coefficient matrix
        A = np.array(S01 @ np.linalg.inv(S11))

        #Create the "F" matrix in Hamilton (1994)
        M = np.size(A,0)*(p-1)

        N = np.size(A,1)

        I = np.eye(M,N)

        F = np.vstack((A,I))

        #Check whether the eigenvalues are less than 1
        Eigvals = list(abs(np.linalg.eigvals(F)))

        if output == True:
            return A,I, F, Eigvals

        else:
            print(all(vals < 1 for vals in Eigvals))
```

Appendix B

Simulation Python Code

B.1 Trending VAR Simulation Code

```
# VAR(1) Simulation With Trend
import statsmodels.api as sm
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import scipy as sc
from numpy import size, random
from statsmodels.tsa.api import VAR

random.seed(12345)

#%% Initialize Parameters
k = 2
p = 1
n = 200
r = 20000

S00 = np.array([[1.3,0.7],[0.7,1.8]])
S01 = np.array([[0.7,0.5],[0.4,0.6]])
S10 = S01.T
S11 = S00

SIGMA = np.vstack((np.hstack((S00,S01)),np.hstack((S10,S11))))

alpha_1 = S01 @ np.linalg.inv(S11)

mu_0 = np.array([3,6])
mu_1 = np.array([0.5,0.25]) #Set this to zero if you do not have a trend

alpha_0 = (np.eye(k,k) - alpha_1) @ mu_0 + (alpha_1 @ mu_1)
delta = (np.eye(k,k) - alpha_1) @ mu_1

coefs = np.hstack((np.ndarray.flatten(alpha_0),np.ndarray.flatten(alpha_1)))
coefs = coefs[[0,1,2,4,3,5]]

sig2 = S00 - S01 @ np.linalg.inv(S11) @ S10

#%% Simulate the VAR(1)

C = []
std_err = []

#Simulate r times and collect estimated coefficients and standard errors

i = 1

while i <= r:

```

```
        print(i)

        y_data = []

        y0 = np.random.multivariate_normal((mu_0 + mu_1), S00, 1) #first observation at t = 1

        y_data.append(y0)

        for j in range(p,n):

            #obs = alpha_0 + delta*(j+1) + y_data[j-1] @ alpha_1.T + np.random.multivariate_normal(np.zeros((k,)), sig2,1)

            #y_data.append(obs)

            mean = np.reshape(alpha_0 + delta*(j+1) + y_data[j-1] @ alpha_1.T, (k,))

            y_data.append(np.random.multivariate_normal(mean, sig2, 1))

        y_data = np.array(y_data).reshape((n,k))

        model = VAR(y_data)
        results = model.fit(p)

        C.append(np.ndarray.flatten(results.params))
        std_err.append(np.ndarray.flatten(results.stderr))

        i = i + 1

    C = np.array(C).reshape((r,6))
    std_err = np.array(std_err).reshape((r,6))

    #%% Compute the test statistics
    test_stat = [abs(C[l,:] - coefs)/std_err[l,:] for l in range(r)]
    test_stat = pd.DataFrame(np.array(test_stat).reshape((r,6)))

    #%% Compute Type I Error Probabilities

    #Count the number of times that the test statistic exceeds the
    #critical point for the t-test with (n - k - 1) df
    alpha_data = test_stat.copy()
    alpha_data[alpha_data < 1.9721] = 0 #1.9721 for n = 200
    alpha_data[alpha_data >= 1.9721] = 1

    err = alpha_data.sum(axis = 0)/r
    print(err)

    #Histograms   z
    for m in range(size(C,1)):
        fig, ax = plt.subplots(figsize = (10,5))
        ax.hist(C[:,m])

        sm.qqplot(C[:,m], fit = True, line = "45")

        print("The mean:", np.mean(C[:,m]))
        print("The SD:", np.std(C[:,m]))

        print("The mean of the test statistic:", np.mean(test_stat.iloc[:,m]))
        print("Average error:", err[m])

    #%% Function to compute Type 2 Error Probabilities and Power

    def power_curve(r, coefs, std_err, discrep_range, discrep_interval, type_1_actual, df):

        power = []
        power_corrected = []

        #type_2_err = []

        for j in range(0,discrep_range):

```

```
        beta_1_alt = coefs + discrep_interval*j

        #Calculate non-centrality parameter- note that numerator could be replaced with discrepency!
        delta_beta_1 = [(abs(coefs - beta_1_alt))/std_err[l] for l in range(r)]

        #Find the area under the non-central t distribution to the left of alpha (type 2 error probability )
        crit_value = sc.stats.t.ppf(0.975, df) #times 0.5 for the two-sided test
        beta_1 = sc.stats.nct.cdf(crit_value, df, delta_beta_1)

        crit_value_corrected = sc.stats.t.ppf(1-(type_1_actual*0.5), df) #times 0.5 for the two-sided test
        beta_1_corrected = sc.stats.nct.cdf(crit_value_corrected, df, delta_beta_1)

        #1 minus the type 2 error probability -power
        pi_1 = 1 - beta_1
        pi_1_c = 1 - beta_1_corrected

        power.append(np.mean(pi_1))
        power_corrected.append(np.mean(pi_1_c))

    power = np.array(power)
    power_corrected = np.array(power_corrected)
    #type_2_err = np.array(type_2_err)

    x = np.linspace(0, discrep_interval*discrep_range, num = discrep_range, endpoint = False)

    fig, ax = plt.subplots(figsize = (10,5))
    ax.plot(x, power,'k--')
    ax.plot(x, power_corrected,'r--')

    ax.plot(-x, power,'k--')
    ax.plot(-x, power_corrected,'r--')

    ax.set_xlabel("Discrepency")
    ax.set_ylabel("Power")
    ax.legend(['Power','Size-Corrected Power'])

#%% Print power curves
#alpha_01
power_curve(r, coefs[0], std_err[:,0], 100, .1, err[0], 196)

#alpha_02
power_curve(r, coefs[1], std_err[:,1], 100, .1, err[1], 196)

#alpha_11
power_curve(r, coefs[2], std_err[:,2], 100, 0.01, err[2], 196)

#alpha_12
power_curve(r, coefs[4], std_err[:,4], 100, 0.01, err[4], 196)

#alpha_21
power_curve(r, coefs[3], std_err[:,3], 100, 0.01, err[3], 196)

#alpha_22
power_curve(r, coefs[5], std_err[:,5], 100, 0.01, err[5], 196)


```

B.2 Seasonal VAR Simulation Code

```
# VAR(1) Simulation With Seasonality
import statsmodels.api as sm
import numpy as np
import pandas as pd
```

```
import matplotlib.pyplot as plt
import scipy as sc
from numpy import size, pi, random
from statsmodels.tsa.api import VAR

random.seed(12345)

#%% Initialize Parameters
k = 2
p = 1
n = 200
r = 20000

S00 = np.array([[1.3,0.7],[0.7,1.8]])
S01 = np.array([[0.7,0.5],[0.4,0.6]])
S10 = S01.T
S11 = S00

SIGMA = np.vstack((np.hstack((S00,S01)),np.hstack((S10,S11))))

alpha_1 = S01 @ np.linalg.inv(S11)

mu_0 = np.array([3,6])
gamma_1 = np.array([0.9,0.25]) #Set these to zero if no seasonality
gamma_2 = np.array([0.4,0.65])
psi_1 = np.array([0.2,0.85])

#gamma_1 = np.array([0.,0.]) #Set these to zero if no seasonality
#gamma_2 = np.array([0.,0.])
#psi_1 = np.array([0.,0.])

alpha_0 = (np.eye(k,k) - alpha_1) @ mu_0
delta_1 = (gamma_1 + alpha_1 @ psi_1)
delta_2 = (psi_1 - alpha_1 @ gamma_1)
delta_3 = (gamma_2 + alpha_1 @ gamma_2)

coefs = np.hstack((np.ndarray.flatten(alpha_0),np.ndarray.flatten(alpha_1)))
coefs = coefs[[0,1,2,4,3,5]]

sig2 = S00 - S01 @ np.linalg.inv(S11) @ S10

#%% Simulate the VAR(1)

C = []
std_err = []

#Simulate r times and collect estimated coefficients and standard errors

i = 1

while i <= r:

    print(i)

    y_data = []

    y0 = np.random.multivariate_normal(mu_0 + gamma_1*np.cos(pi/2) + psi_1*np.sin(pi/2) + gamma_2*np.cos(pi), S00, 1)

    y_data.append(y0)

    for j in range(p,n):

        #obs = alpha_0 + delta_1*np.cos(pi*(j+1)/2) + delta_2*np.sin(pi*(j+1)/2) + delta_3*np.cos(pi*(j+1)) + y_data[j-1] @ alpha_1.T + np.random.multivariate_normal(np.zeros((k,)), sig2,1)

        #y_data.append(obs)

        mean = np.reshape(alpha_0 + delta_1*np.cos(pi*(j+1)/2) + delta_2*np.sin(pi*(j+1)/2) + delta_3*np.cos(pi*(j+1)) + y_data[j-1] @ alpha_1.T , (k,))

```

```
        y_data.append(np.random.multivariate_normal(mean, sig2, 1))

    y_data = np.array(y_data).reshape((n,k))

    model = VAR(y_data)
    results = model.fit(p)

    C.append(np.ndarray.flatten(results.params))
    std_err.append(np.ndarray.flatten(results.stderr))

    i = i + 1

C = np.array(C).reshape((r,6))
std_err = np.array(std_err).reshape((r,6))

#%%Compute the test statistics
test_stat = [abs(C[l,:] - coefs)/std_err[l,:] for l in range(r)]
test_stat = pd.DataFrame(np.array(test_stat).reshape((r,6)))

#%% Compute Type I Error Probabilities

# Count the number of times that the test statistic exceeds the critical point for the t-test with (n - k - 1) df
alpha_data = test_stat.copy()
alpha_data[alpha_data < 1.9721] = 0 #Change this based on the number of observations, "n"
alpha_data[alpha_data >= 1.9721] = 1

err = alpha_data.sum(axis = 0)/r
print(err)

#Time Series Plots
fig, ax = plt.subplots(figsize=(10, 5))
ax.plot(y_data)
ax.set_xlabel("Time (Quarters)")
ax.set_ylabel("y")
plt.show()

#Histograms
for m in range(size(C,1)):
    fig, ax = plt.subplots(figsize = (10,5))
    ax.hist(C[:,m])

    sm.qqplot(C[:,m], fit = True, line = "45")

    print("The mean:", np.mean(C[:,m]))
    print("The SD:", np.std(C[:,m]))

    print("The mean of the test statistic:", np.mean(test_stat.iloc[:,m]))
    print("Average error:", err[m])

#%% Function to compute Type 2 Error Probabilities and Power

def power_curve(r, coefs, std_err, discrep_range, discrep_interval, type_1_actual, df):

    power = []
    power_corrected = []

    #type_2_err = []

    for j in range(0,discrep_range):

        beta_1_alt = coefs + discrep_interval*j

        #Calculate non-centrality parameter- note that numerator could be replaced with discrepency!
        delta_beta_1 = [(abs(coefs - beta_1_alt))/std_err[l] for l in range(r)]

        #Find the area under the non-central t distribution to the left of alpha (type 2 error probability)
        crit_value = sc.stats.t.ppf(0.975, df) #times 0.5 for the two-sided test
        beta_1 = sc.stats.nct.cdf(crit_value, df, delta_beta_1)

```

```
            crit_value_corrected = sc.stats.t.ppf(1-(type_1_actual*0.5), df) #times 0.5 for the two-sided test
            beta_1_corrected = sc.stats.nct.cdf(crit_value_corrected, df, delta_beta_1)

            #1 minus the type 2 error probability -power
            pi_1 = 1 - beta_1
            pi_1_c = 1 - beta_1_corrected

            power.append(np.mean(pi_1))
            power_corrected.append(np.mean(pi_1_c))

        power = np.array(power)
        power_corrected = np.array(power_corrected)
        #type_2_err = np.array(type_2_err)

        x = np.linspace(0, discrep_interval*discrep_range, num = discrep_range, endpoint = False)

        fig, ax = plt.subplots(figsize = (10,5))
        ax.plot(x, power,'k--')
        ax.plot(x, power_corrected,'r--')

        ax.plot(-x, power,'k--')
        ax.plot(-x, power_corrected,'r--')

        ax.set_xlabel("Discrepency")
        ax.set_ylabel("Power")
        ax.legend(['Power','Size-Corrected Power'])

    #%% Print power curves

    #alpha_01
    power_curve(r, coefs[0], std_err[:,0], 100, .1, err[0], 196)

    #alpha_02
    power_curve(r, coefs[1], std_err[:,1], 100, .1, err[1], 196)

    #alpha_11
    power_curve(r, coefs[2], std_err[:,2], 100, 0.01, err[2], 196)

    #alpha_12
    power_curve(r, coefs[4], std_err[:,4], 100, 0.01, err[4], 196)

    #alpha_21
    power_curve(r, coefs[3], std_err[:,3], 100, 0.01, err[3], 196)

    #alpha_22
    power_curve(r, coefs[5], std_err[:,5], 100, 0.01, err[5], 196)

```

Appendix C

Data and Transformations

Data Series	Source	Frequency	Aggregation Method	Units (Original)	Transformation	Seasonal	Dates Available
All Employees: Total nonfarm	U.S. BLS	M	End of Period (Quarter)	Thousands of Persons	Natural Logarithm	NSA and SA available	1939-01-01 to 10/1/2020
Real gross domestic product	U.S. BEA	Q	N/A	Billions of Chained 2012 Dollars	Natural Logarithm	SA only	1947-01-01 to 7/1/2020
Retail Sales and Food Services	U.S. Census Bureau	M	End of Period (Quarter)	Millions of Dollars	Natural Logarithm	NSA and SA	1992-01-01 to 9/1/2020
CPI-U: All items	U.S. BLS	M	End of Period (Quarter)	Index 1982-1984 = 100	Natural Logarithm	NSA and SA Available	1939-13-01 to 10/1/2020
House prices (S&P/Case-Shiller U.S. National Home Price Index)	S&P Dow Jones Indices LLC	M	End of Period (Quarter)	Index Jan 2000 = 100	Natural Logarithm	NSA and SA available	1987-01-01 to 8/1/2020
Housing starts	U.S. Census Bureau	M	End of Period (Quarter)	Thousands of Units	Natural Logarithm	NSA and SA available	1959-01-01 to 9/1/2020
Civilian Unemployment Rate	U.S. BLS	M	End of Period (Quarter)	Percent	N/A	NSA and SA Available	1948-01-01 to 10/1/2020
Industrial Production Index	Board of Governors of the Federal Reserve System	M	End of Period (Quarter)	Index 2012 = 100	Natural Logarithm	NSA and SA Available	1919-02-01 to 9/1/2020
Merchant wholesalers: Inventories Total	U.S. Census Bureau	M (end of month)	End of Period (Quarter)	Millions of dollars	Natural Logarithm	NSA and SA Available	1992-01-01 to 9/1/2020
Philly Fed Mfg. business outlook: Current activity	Fed. Reserve Bank of Philadelphia	M	End of Period (Quarter)	Index	N/A	NSA and SA available	1968-05-01 to 10/1/2020
Capacity Utilization	Board of Governors of the Federal Reserve System	M	End of Period (Quarter)	Percent of Capacity	N/A	SA only	1967-01-01 to 9/1/2020
Stock Price Index (SP500): SPINDX	Wharton	Q	End of Period (Quarter)	Level	Natural Logarithm	NSA	Q41925-Q42019

Figure C.1: Preliminary Data Information and Transformations

The figure above contains preliminary information on each of the data series used in the project. Note that in the case of real gross domestic product (GDP), the data from the Bureau of Economic Analysis was only available as seasonally adjusted for the period of interest. For reasons discussed in Section 3.3.1, seasonally adjusted data is not ideal for analysis. As such, nominal GDP data was obtained from the Bureau of Economic Analysis and adjusted for inflation manually via the GDP deflator with base year equal to 2012. The conversion from nominal to real GDP was accomplished via the following transformation:

$$\text{Real GDP (billions)} = \frac{\text{Nominal GDP (billions)}}{\text{GDP Deflator } (2012 = 100)} \times 100$$

The converted Real GDP data was then plotted and compared with the unadjusted values on the FRED database [1]. The plot of the transformed data for the period available on FRED is given in Figure C.2.

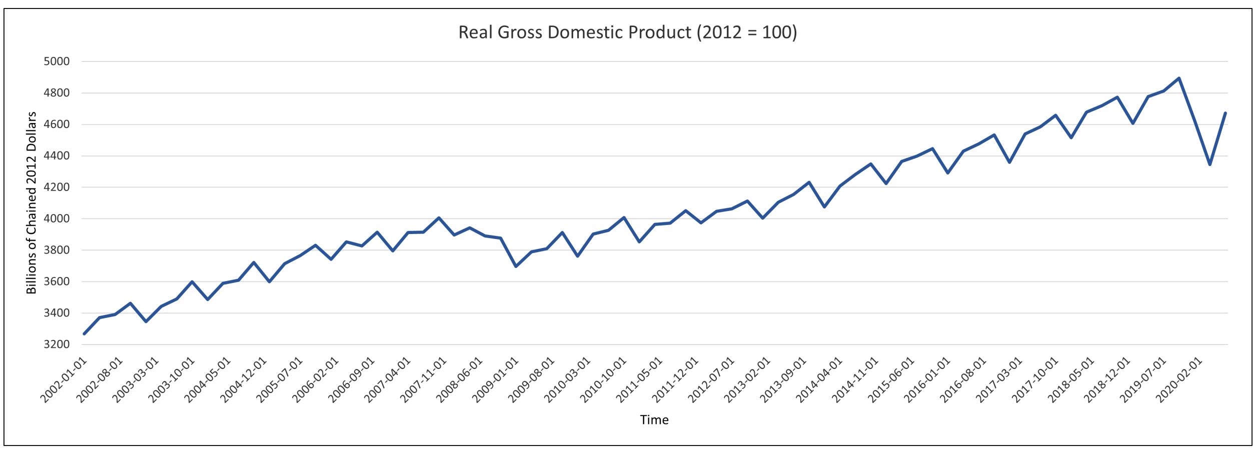


Figure C.2: Plot of Real GDP via GDP Deflator

[1]The data on FRED can be found here: https://fred.stlouisfed.org/series/ND000334Q

Appendix D

AR(1) Model and M-S Testing: Real GDP

D.0.1 Estimated Model and Auxiliary Regressions

Model mis-specification testing is conducted on the estimated residuals from the normal AR(1) model in order to ensure its statistical adequacy prior to selecting the number of trends and lags in the forecasting framework. Auxiliary regressions on the estimated residuals, as shown in Spanos (2018, 2019a), detect departures from the probabilistic assumptions invoked in the model. In the context of the normal, autoregressive model for the natural log of gross domestic product, auxiliary regressions take the following form: AR(1) Specification:

$$lngdp_t = a_0 + \sum_{i=1}^{3} \delta_i t_*^i + \sum_{i=1}^{2} (\alpha_i cos_{it} + \beta_i sin_{it}) + \psi_1 D68 + \gamma_1 lngdp_{t-1} + u_t.$$

Mean auxiliary regression:

$$\widehat{u_t} = \overbrace{a_0 + \sum_{i=1}^{3} \delta_i t_*^i + \sum_{i=1}^{2} (\alpha_i cos_{it} + \beta_i sin_{it}) + \psi_1 D68 + \gamma_1 lngdp_{t-1}}^{\text{original specification}} + \overbrace{a_1 t_*^4 + a_2 t_*^5}^{[5]}$$
$$+ \overbrace{\beta_3 lngdp_{t-1}^2}^{[2]} + \overbrace{\beta_4 lngdp_{t-2} + \beta_5 lngdp_{t-3}}^{[4]} + v_{1,t}$$

Variance auxiliary regression:

$$\widehat{u}_t^{\,2} = \delta_0 + \overbrace{\delta_1 t_*^1 + \delta_2 t_*^2}^{[5]} + \overbrace{\beta_1 lngdp_{t-1}^2}^{[3]} + \overbrace{\gamma_1 \hat{u}_{t-1}^2 + \gamma_2 \hat{u}_{t-2}^2}^{[4]} + v_{2,t}$$

Table D.1 shows the null hypotheses for testing each assumption explicitly. Note that these mis-specification tests detect departures from the probabilistic assumptions by testing the additional terms, not the terms from the original model specification.

Table D.1: M-S Tests: Normal AR(1) Model: Natural Log of GDP

Assumptions	**Null Hypothesis**
Joint (Regression Function)	H_0: $a_1 = a_2 = \beta_3 = \beta_4 = \beta_5 = 0$
Linearity	H_0: $\beta_3 = 0$
Dependence (Mean)	H_0: $\beta_4 = \beta_5 = 0$
1st t-invariance	H_0: $a_1 = a_2 = 0$
Joint (Skedastic Function)	H_0: $\delta_1 = \delta_2 = \beta_1 = \gamma_1 = \gamma_2 = 0$
Heteroskedasticity	H_0: $\beta_1 = 0$
Dependence (Variance)	H_0: $\gamma_1 = \gamma_2 = 0$
2nd t-invariance	H_0: $\delta_1 = \delta_2 = 0$

D.0.2 Estimated Equation and M-S Testing Results

The estimated AR(1) equation is given below:

$$\widehat{lngdp}_t = \underset{(0.277)}{0.689} + \underset{(0.012)^*}{0.027}\,t_*^1 - \underset{(0.002)}{0.003}t_*^2 + \underset{(0.001)^*}{0.003}\,t_*^3 - \underset{(0.007)^*}{0.024}\,D68 - \underset{(0.001)^*}{0.003}\,cos_{1t} - \underset{(0.001)^*}{0.020}\,sin_{1t}$$
$$+ \underset{(0.001)^*}{0.020}\,cos_{2t} + \underset{(0.034)^*}{0.916}\,lngdp_{t-1}$$

Standard errors are given under each estimated coefficient in parentheses. Note that an asterisk next to the standard error in the estimated equation indicates statistical significance at the 2.5% significance level. Table D.2 provides the M-S test results for the AR(1) model specification above. Here, the F-statistics are given alongside the corresponding p-values in brackets. In light of these results, one can conclude that the current specification is statistically adequate, as there is no evidence that the assumptions listed in Table D.1 are invalid.

Table D.2: M-S Test Results for AR(1) Model: Real GDP

Assumptions	
Joint (regression)	0.93[.4641]
Linearity	0.92[.3387]
Dependence (Mean)	0.49[.6145]
1st t-invariance	1.66[.1962]
Joint (skedastic)	1.29[.2752]
Heteroskedasticity	2.65[.1069]
Dependence (Variance)	0.73[.4840]
2nd t-invariance	1.60[.2059]

Appendix E

Testing for Non-Normality: VAR(L) Model Residuals

As mentioned previously, the forecasting methodology in Chapter 3 incorporates the rigorous testing of the model assumptions. It is important to highlight that testing for non-normality is conducted *after* testing the other model assumptions. That is because, as noted in Spanos (2019a), if the other model assumptions are invalid, then so too is the distributional assumption. Non-normality is assessed in two ways: first through the use of the Anderson-Darling test on the VAR(3) model residuals and second through distributional plots.

E.1 Anderson-Darling Test

The null hypothesis for the Anderson-Darling (A-D) test is that the data follows a specified distribution, in this case a normal distribution. In this case, it is the VAR(3) model residuals from Chapter 3 that are being tested, as this model was deemed to be statistically adequate in the last chapter. Table E.1 below shows the results for the test, with bold font indicating test statistics significant at the 2.5% level or lower.

Table E.1: Anderson-Darling Test Results for VAR(3) Model

Equation	A-D Test Statistic
$lngdp_t$	0.6318
$lnnonfarm_t$	0.3233
$lnretail_t$	0.4056
$lncpi_t$	0.3624
$lnhp_t$	0.2321
$lnhs_t$	0.3687
$lnip_t$	0.4278
$lninv_t$	0.3751
$lnsp500_t$	0.7604
$outlook_t$	0.7198
$cap.util_t$	0.3941
ur_t	0.2264

Note that the critical values are 0.761, 0.888, 1.056 for the 5%, 2.5% and 1% levels, respectively.

Distributional Plots

Another check to the distributional assumption of normality comes in the form distributional plots, specifically QQ-plots. Figure E.1 shows the plots for statistically adequate VAR(3) model residuals.

For each of the equations, the quantile plots show that a majority of the points lie on the standardized red line. However, toward the tail ends, the points start to deviate from the line, which is indicative of larger tails for the distribution. Longer, or "fatter", distribution tails are characteristic of a distribution more like the Student's t distribution, and, as such, these results ultimately demonstrate that the Student's t model may be more appropriate. However, since the Normal VAR(3) model was established to be statistically adequate in Chapter 3, i.e. no evidence detecting a violation of model assumptions, the Normal VAR(3) was selected over a Student's t version of the VAR model.

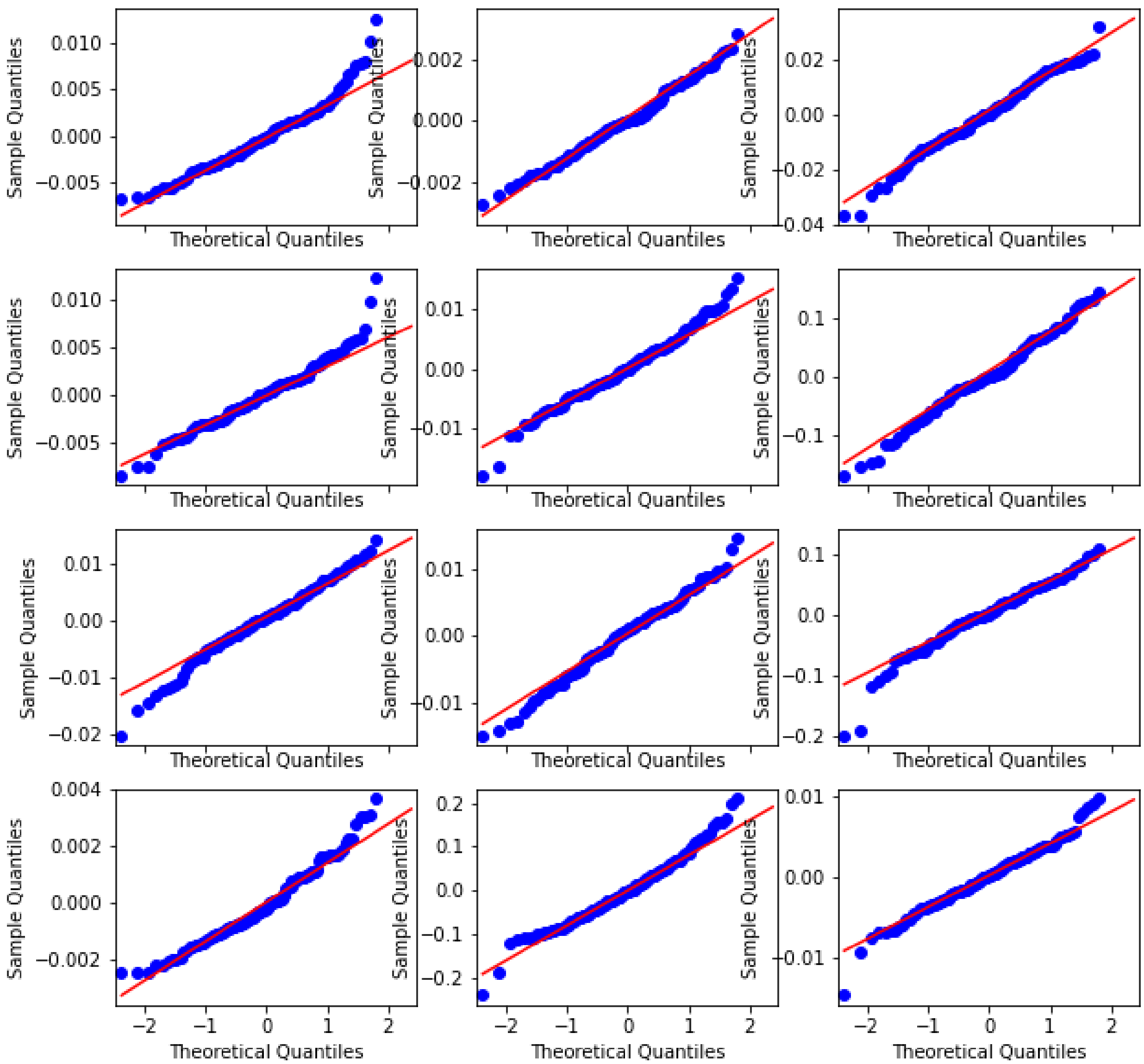


Figure E.1: VAR(3) QQ-Plots

Appendix F

Forecasting Python Code

F.1 Importing Macroeconomic Data

```
import statsmodels.api as sm
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import datetime
import warnings
import scipy as sc

#pip install pandas-datareader
import pandas_datareader as pdr

#pip install yfinance
import yfinance as yf

warnings.filterwarnings('ignore')

#%% Import the Updated Data
start_date = datetime.datetime(1992, 1, 1)
end_date = datetime.datetime(2019, 12, 1)

#Import Macroeconomic Predictors Other than GDP
FRED = ['PAYNSA#0', 'RSAFSNA', 'CPIAUCNS#0', 'CSUSHPINSA#0', 'HOUSTNSA#0', 'UNRATENSA', 'IPB50001N#0',
        'I42IMNM144NCEN','GACDFNA066MNFRBPHI', 'TCU']

macro_df = pdr.DataReader(FRED, 'fred', start_date, end_date)

#Import Nominal GDP and GDP Deflator; Note that GDP Deflator is seasonally adjusted
GDP = ['NA000334Q']

gdp_data = pdr.DataReader(GDP, 'fred', start_date, end_date)
gdp_deflator = pdr.DataReader(['GDPDEF'], 'fred', start_date, end_date)

#Import S&P500 Closing Prices
SP500 = pd.DataFrame(yf.download('^GSPC', start = start_date, end = end_date, interval = "3mo")['Adj Close
 '])

#Convert data to quarterly data using end of period values
clean_data = macro_df.copy()
clean_data =clean_data.drop(clean_data.index[clean_data.index.month.isin([1,2,4,5,7,8,10,11])])
clean_data.index = gdp_data.index

#Merge the DataFrames by columns
Data_Original = pd.concat([gdp_data, clean_data, SP500], axis = 1)

#%% Transform the Dataset
Transformed_Data = Data_Original.copy()

#Convert Units
```

```
    Transformed_Data['NA000334Q'] = Transformed_Data['NA000334Q']/1000 #Millions to Billions
    Transformed_Data['UNRATENSA'] = Transformed_Data['UNRATENSA']/100 #Convert percent to decimal
    Transformed_Data['TCU'] = Transformed_Data['TCU']/100 #Convert percent to decimal

    #Convert Nominal GDP to Real GDP
    Transformed_Data['NA000334Q'] = Transformed_Data['NA000334Q'].div(gdp_deflator['GDPDEF'], axis = 0)*100

    #Rename Data Columns
    Transformed_Data.columns = ['gdp','nonfarm', 'retail','cpi','hs','hp','ur','ip','inv','business','capacity_util', 'sp500']

    #Take Natural Logs of Data
    Transformed_Data[['gdp','nonfarm', 'retail','cpi','hs','hp','ip','inv','sp500']] = np.log(Transformed_Data[['gdp','nonfarm', 'retail','cpi','hs','hp','ip','inv','sp500']])

```

F.2 Factor Augmented Autoregressive Model Python Code

```
#%% Import required modules
import statsmodels.api as sm
import numpy as np
from numpy import size
import pandas as pd
import math
import matplotlib.pyplot as plt
import sympy as sym
from numpy.linalg import eig
from sklearn.metrics import mean_squared_error, mean_absolute_error
from sklearn.linear_model import LinearRegression
#from statsmodels.tsa.deterministic import DeterministicProcess, Fourier

#%% Functions for Scaling t and Chebyshev

"""
The functions below were created in response to Michaelides and Spanos (2020)
and Spanos (2019), where it was found that the inclusion of trend polynomials
in estimation can lead to misleading results due to near-multicollinearity.
Hence, as suggested in the paper, a combination of Chebyshev Polynomials of the
First Kind and scaling are used to detrend the data.

"""

def scaled_t(data):
    '''
    Parameters
    ----------
    data : pandas dataframe or array containing data

    Returns
    -------
    t_scaled : list containing scaled time variable
    '''

    n = size(data,0)
    t_scaled = []

    for t in range(1,n+1):
        t_s = (2*t - n - 1)/(n-1)
        t_scaled.append(t_s)
```

```

    return t_scaled

#Chebyshev Polynomial of the First Kind
def Chebyshev_poly(t_s, trend):
    '''
    Parameters
    ----------
    t_s: is a list containing the scaled time variable
    trend: is an int describing the order of the trend (trend > 0)

    Returns
    -------
    t_poly: dictionary containing scaled, orthogonal time polynomials

    '''
    t = sym.symbols('t')

    t_poly = {}

    for k in range(1,trend+1):
        constant = ((-1)**k*2**k*math.factorial(k))/(math.factorial(2*k))
        w = (1 - t**2)**(0.5)
        der = (1 - t**2)**(k-0.5)
        ddt = sym.diff(der,t,k)
        poly = sym.expand(constant*w*ddt)

        f = sym.lambdify(t,poly)

        t_poly['t_%d' %(k)] = []

        for i in range(0, size(t_s,0)):
            t_poly['t_%d' %(k)].append(f(t_s[i]))

    return t_poly

#Function defining seasonal terms
def trig_seasonal(data, s):

    '''
    Parameters
    ----------
    data: dataframe containing data
    s: is an int denoting the number of seasons

    Returns
    -------
    seasonal_vars: dictionary containing trig seasonal terms

    '''

    terms = int(s*0.5)
    t = np.array(np.arange(1,np.size(data,0)+1))

    seasonal_vars = {}

    for i in range(1, terms + 1):

        seasonal_vars['cos_%d' %(i)] = np.round(np.cos(2*i*math.pi*t/s),4)
        seasonal_vars['sin_%d' %(i)] = np.round(np.sin(2*i*math.pi*t/s),4)

    return seasonal_vars

#The adjustments below make the seasonal dataframe equivalent to the
#Fourier produced in the Deterministic Process module

# terms = int(4*0.5)
# t = np.array(np.arange(0,np.size(Transformed_Data,0))) #Index vs Time (start with 0 or 1, respectively)
# seasonal_vars_test = {}

# for i in range(1, terms + 1):
```

```python
#     seasonal_vars_test['cos_%d' %(i)] = np.round(np.cos(2*i*math.pi*t/4),4)
#     seasonal_vars_test['sin_%d' %(i)] = np.round(np.sin(2*i*math.pi*t/4),4)
# seasonal_df = pd.DataFrame(seasonal_vars_test).apply(pd.Series.explode).reset_index(drop = True)

#%% Detrending and De-Seasonalizing Regression Results

'''
Construct the independent and dependent variables. The independent variables
in the regression are the scaled time polynomials, seasonality terms, and
the dummy variables. Note that the intercept and coefficients are saved into
a list and stored for later use. At this step, the intercept is also subtracted
out, but will be added back in.
'''

def detrend_data(data_trend, data):

    trend_intercept = []
    trend_coef = []

    k = size(data, 1)

    X = data_trend

    data_detrended = pd.DataFrame()

    for i in range(0,k):
        y = data.iloc[:,i]

        lm = LinearRegression()

        lm.fit(X,y)

        trend_intercept.append(lm.intercept_)
        trend_coef.append(lm.coef_)

        y_hat = lm.predict(X)

        data_detrended['y%d_d'%(i+1)] = y - y_hat #intercept also included in subtraction

    return data_detrended, trend_intercept, trend_coef

#%% Add lags back to detrended dataset

'''
As mentioned in the paper, the number of lags and time trends were determined
prior to the analysis through the use of vector autoregressive models. Now that
the data has been detrended, lags are constructed and added into a copy of the
detrended dataset.

The lagged data columns now have NAN's in the first "lag" rows. To avoid an
error message when computing the sample variance, the estimated trends are
added back into the copied dataset. Note that the addition of a constant will
not change the sample variance. Moreover, see the paper for additional notes
on how the modules in python deal with the NAN problem.

'''

def lagged_dataset(data, lags, k, trend_intercept = None, intercept = False):

    data_with_lags = data.copy()

    for p in range(1, lags+1):
        for i in range(1,k+1):
            data_with_lags['L%d'%(p) + "." + 'y%d_d'%(i)] = data_with_lags['y%d_d'%(i)].shift(p)

    if intercept == False:

        return data_with_lags

    else:

```

```
        data_with_lags_int = data_with_lags.copy()
        data_with_lags_int = data_with_lags_int.fillna(value=0)

        # for r in range(lags+1):
        #     data_with_lags_int.iloc[:,k*r:k*(r+1)] =  data_with_lags_int.iloc[:,k*r:k*(r+1)]
        #        + np.asarray(trend_intercept) #lag 0

        return data_with_lags_int

#%% Estimated Variance-Covariance Matrix and Standardized Correlation Matrix

def cov_matrix(data, corr):

    SIGMA = np.cov(data, rowvar = False, bias = True) #Bias is for MLE! Divides by N instead of N-1

    if corr == False:

        return SIGMA

    else:

        #Obtain Standard Deviations (From Diagonal of SIGMA)
        D = np.sqrt(np.diag(np.diag(SIGMA)))

        #Invert
        D_inv = np.linalg.inv(D)

        #Standardized Correlation Matrix
        CORR = D_inv @ SIGMA @ D_inv

        return CORR

#%% Functions for Determining the Number of Principal Components to Use

def and_test(values, per):

    '''
    Parameters
    ----------
    values: list of ordered eigenvalues (from smallest to largest)
    per: threshold for percentage of variance explained

    Returns
    -------
    prints the number of principal components needed to reach the pre-
    specified threshold for the proportion of the variance explained

    '''
    count = 1
    sum_ = values[-1]

    sum_all = np.sum(values)

    for i in range(2,size(values)+1):
        if (sum_/sum_all) < per:
            sum_ = sum_ + values[-i]

            count = count + 1

        else:
            print("Sum:", sum_)
            print("Trace:", sum_all)
            print( "Proportion of Variance Explained:", sum_/sum_all)
            break

    print("Number of Components Needed:", count)

def exp_var(values, per):

    '''
    Parameters
```

```
    ----------
    values: list of eigenvalues from smallest to largest
    per: threshold for percentage of variance explained

    Returns
    -------
    explained_var: list of variance explained by each additional eigenvalue
    count: # of components needed to reach pre-specified threshold
    '''
    explained_var = []

    count = 1
    sum_ = values[-1]
    sum_all = np.sum(values)

    explained_var.append(sum_/sum_all)

    for i in range(2,size(values)+1):
        if (sum_/sum_all) < per:

            sum_ = sum_ + values[-i]

            count = count + 1

            explained_var.append(sum_/sum_all)

        else:
            break
    print('Proportion of Variance Explained',(sum_/sum_all))
    print('Count:', count)
    return explained_var, count

#%% Decomposition of SIGMA Matrix and CORR Matrix and Testing Number of Components Needed

def eig_decomp(Matrix):

    val, vec = eig(Matrix)

    idx = np.argsort(val)

    val = val[idx]
    vec = vec[:, idx]

    return val, vec

#%% Function to compute Principal Components

#Function to compute principal components
def pc(k, p, vector, data):

    PC = pd.DataFrame()

    Comp = {}

    for i in range(p):
        Comp['PC_%d'%(i+1)] = np.matmul(vector[:,k-1-i].T, np.asarray(data.transpose()))
        PC['PC_%d'%(i+1)] = Comp['PC_%d'%(i+1)]

    return PC

#%% Function for Dropping Principal Components

def drop_pc(X, y, count, display = True):

    lm = sm.OLS(y.values, X.values).fit()

    p_val = lm.pvalues

    for r in range(count):
        if p_val[r] >= 0.025:
            X = X.drop(['PC_%d'%(r+1)], axis = 1)
```

```
            if display == True:
                print(p_val[r])
                print("Dropped", 'PC_%d'%(r+1))

            else:
                continue
        else:
            continue

    return X

#%% Temporal Principal Components Function

def temp_pc(data, data_trend, a, lags, drop_contemp, SIGMA, pc_intercept = False):

    k = np.size(data,1)

    #Detrend data
    data_d, trend_intercept, trend_coef = detrend_data(data_trend, data)

    #Drop the contemporaneous data from the model, only considering lagged values for forecasting!
    if drop_contemp == True:
        #If no columns need to be added (additional lags for the VAR(p) model)
        #data_with_lags = lagged_dataset(data_d, lags, k, trend_intercept = None, intercept = False).iloc[:,k:]
        #data_with_lags_int = lagged_dataset(data_d, lags, k, trend_intercept, intercept = True).iloc[:,k:]

        #Adding another lag and dropping contemporaneous columns
        data_with_lags = lagged_dataset(data_d, int(lags+1), k, trend_intercept = None, intercept = False).iloc[:,k:]
        data_with_lags_int = lagged_dataset(data_d, int(lags+1), k, trend_intercept, intercept = True).iloc[:,k:]

    else:
        data_with_lags = lagged_dataset(data_d, lags, k, trend_intercept = None, intercept = False)
        data_with_lags_int = lagged_dataset(data_d, lags, k, trend_intercept, intercept = True)

    #Compute the sample covariance matrix
    if SIGMA == True:
        CORR = cov_matrix(data_with_lags_int, corr = False)
    else:
        CORR  = cov_matrix(data_with_lags_int, corr = True)

    #Spectral decomposition
    val, vec = eig_decomp(CORR)
    ex_val, count = exp_var(val, a)

    #Compute temporal principal components (with or without intercept added back in)
    if pc_intercept == False:
        PC = pc(size(data_with_lags_int, 1), count, vec, data_with_lags)
    else:
        PC = pc(size(data_with_lags_int, 1), count, vec, data_with_lags_int)

    return PC, count

#%% In sample forecasting method- similar to benchmark drop pc changed - there was an error in the ordering of the columns

def forecast_far(data, test, q, lags, a, k, trend, s, column, lags_ar, scaling, droppc, drop_contemp, SIGMA, pc_intercept):

    '''
    Parameters
    ----------
    data : pandas dataframe or array containing data
    q: number of observations in test set
    lags: number of lags for VAR(q) model
    a: percentage of variance explained threshold
    k: number of data series
    trend: degree of trend polynomial
```

```
    s: number of seasons (e.g. 4 for quarterly data)
    column: location of dependent variable in dataframe
    lags_ar: number of lags in the AR(p) model
    scaling: True uses the scaled time variables (and orthogonal polynomials), False does not
    droppc: if True, drops statistically insignificant principal components
    drop_contemp: if True, drops contemporaneous columns from the model, primarily for forecasting
    SIGMA: if True, uses the covariance matrix. if False, uses the standardized correlation matrix
    pc_intercept: if True, adds intercept to dataset in PC computation

    Returns
    -------
    predictions_arry: array containing the specified predictions
    '''

    #Construct the time variables dataset for detrending
    data_trend = pd.DataFrame()

    if scaling == True:
        #applies scaling and orthogonal polynomials
        t_scaled = scaled_t(data)
        t_poly = pd.DataFrame(Chebyshev_poly(t_scaled, trend)).apply(pd.Series.explode).reset_index(drop = True)

        data_trend = pd.concat([data_trend, t_poly])

    else:
        #scaling is not applied
        t = np.array(np.arange(1,np.size(data,0)+1))

        for i in range(trend):
            data_trend['t_%d'%(i+1)] = t**(i+1)

    seasonal_vars = trig_seasonal(data, s)

    for i in range(0, int(s/2)):

        data_trend['cos_%d'%(i+1)] = seasonal_vars['cos_%d'%(i+1)]
        data_trend['sin_%d'%(i+1)] = seasonal_vars['sin_%d'%(i+1)]

    PC, count = temp_pc(data, data_trend, a, lags, drop_contemp, SIGMA, pc_intercept)

    #Create the lags for the autoregressive (p) process
    Lags_AR = pd.DataFrame()
    for j in range(1, int(lags_ar + 1)):
        Lags_AR['L%dy'%(j)] = data.iloc[:,column].shift(j)

    #Specify the data matrix for FAR(1,3) independent variables
    X = pd.concat([Lags_AR.iloc[:,:],PC.iloc[:,:],data_trend.iloc[:,:]],axis = 1) #Create new dataframe with trends

    #Add constant to the dataframe for OLS
    X = sm.add_constant(X)

    #Include all data except last q observations (q observations)
    X_train = X.iloc[:-test,:]
    y_train = data.iloc[:-test, column]

    #Dropping insignificant principal components
    if drop_contemp == True:

        if droppc == True:

            if pc_intercept == True:
                #Fit the model and record p-values
                lm = sm.OLS(y_train.iloc[lags_ar:], X_train.iloc[lags_ar:,:]).fit()
            else:
                #Fit the model and record p-values
                lm = sm.OLS(y_train.iloc[int(lags+1):], X_train.iloc[int(lags+1):,:]).fit()

            p_val = lm.pvalues

```

```
            #Drop insignificant principal components
            for r in range(int(lags_ar+1), int(lags_ar + 1 + count)):
                if p_val[r] >= 0.025:
                    X_train = X_train.drop(['PC_%d'%(int(r-lags_ar))], axis = 1)
                else:
                    continue

            list_cols = list(X_train.columns.values)
            X = X.drop(columns = X.columns.difference(list_cols))

            #Reestimate the Model Without Insignificant PC's
            if pc_intercept == True:
                #Fit the model and record p-values
                lm = sm.OLS(y_train.iloc[lags_ar:], X_train.iloc[lags_ar:,:]).fit()
            else:
                #Fit the model and record p-values
                lm = sm.OLS(y_train.iloc[lags:], X_train.iloc[lags:,:]).fit()

            p_val = lm.pvalues
            params = lm.params

        else:
            #Fit the model
            if pc_intercept == True:
                #Fit the model and record p-values
                lm = sm.OLS(y_train.iloc[lags_ar:], X_train.iloc[lags_ar:,:]).fit()

            else:
                #Fit the model and record p-values
                lm = sm.OLS(y_train.iloc[lags:], X_train.iloc[lags:,:]).fit()

            params = lm.params
    else:

        if droppc == True:

            if pc_intercept == True:
                #Fit the model and record p-values
                lm = sm.OLS(y_train.iloc[lags_ar:], X_train.iloc[lags_ar:,:]).fit()
            else:
                #Fit the model and record p-values
                lm = sm.OLS(y_train.iloc[lags:], X_train.iloc[lags:,:]).fit()

            p_val = lm.pvalues

            #Drop insignificant principal components
            for r in range(int(lags_ar+1), int(lags_ar + 1 + count)):
                if p_val[r] >= 0.025:
                    X_train = X_train.drop(['PC_%d'%(int(r-lags_ar))], axis = 1)
                else:
                    continue

            list_cols = list(X_train.columns.values)
            X = X.drop(columns = X.columns.difference(list_cols))

            #Reestimate the Model Without Insignificant PC's
            if pc_intercept == True:
                #Fit the model and record p-values
                lm = sm.OLS(y_train.iloc[lags_ar:], X_train.iloc[lags_ar:,:]).fit()
            else:
                #Fit the model and record p-values
                lm = sm.OLS(y_train.iloc[lags:], X_train.iloc[lags:,:]).fit()

            p_val = lm.pvalues
            params = lm.params

        else:
            #Fit the model
            if pc_intercept == True:
                #Fit the model and record p-values
                lm = sm.OLS(y_train.iloc[lags_ar:], X_train.iloc[lags_ar:,:]).fit()
```

```

            else:
                #Fit the model and record p-values
                lm = sm.OLS(y_train.iloc[lags:], X_train.iloc[lags:,:]).fit()

            params = lm.params

    #Record estimated coefficients and intercept
    intercept = params[0] #intercept is last column in X
    coefficient = params[1:] #evertyhing but intercept, makes separating trend terms less confusing

    #Break coefficients into respective FAR(p,q) parts
    coefficient_ar = coefficient[0:lags_ar]
    coefficient_trends = coefficient[-int(np.size(data_trend,1)):]
    coefficient_pcs = coefficient[lags_ar:-int(np.size(data_trend,1))]

    #prediction = intercept + trends + lagged dep. var + principal components
    predictions = list(y_train.iloc[-lags_ar:])

    predictions.append(intercept + np.dot(coefficient_trends, data_trend.iloc[np.size(X_train,0),:])
                                 + np.dot(list(reversed(coefficient_ar)), y_train.iloc[-lags_ar:])
                                 + np.dot(coefficient_pcs, X_train.iloc[-1,int(1+lags_ar):-int(np.size(data_trend,1))])
                                 )

    for r in range(1,q):
        prediction = (intercept +
                                 np.dot(coefficient_trends, data_trend.iloc[np.size(X_train,0) + r,:]) + #intercept + trends
                                 np.dot(list(reversed(coefficient_ar)), predictions[r:]) + #lagged dep var
                                 np.dot(coefficient_pcs, X.iloc[np.size(X_train,0) + r,
                                                                int(1+lags_ar):-int(np.size(data_trend,1))])
                                 )#principal components

        predictions.append(prediction)

    predictions_array = np.asarray(predictions)

    return predictions_array[lags_ar:]

#%% Forecast Evaluation Metrics
def far_forecast_eval(predictions_dictionary, data, test, q, column, evaluation = True, plot_forecasts = True, plot_errors = False):

    '''
    predictions_dictionary: dictionary containing forecast results
    data: dataframe containing data to compare
    q: the number of observations in the forecasting horizon
    column: data column matching dependent variable
    evaluation: string specifying evaluation metrics. Options are "All", "MSE", "RMSE", "MAE"
    plot_forecasts: if True, plots forecasts for each dictionary key
    plot_errors: if True, plots the forecasting errors
    '''

    if plot_forecasts == True:

        name_list = data.columns.values.tolist()

        #Plot the Forecasts against the actual values-last q quarters----------------
        fig, ax = plt.subplots(figsize=(10, 5))
        ax.plot(np.array(data.iloc[-test:,0]),np.array(data.iloc[-test:,column]), 'k', label = 'Actual')
        for key in predictions_dictionary.keys():
            ax.plot(np.array(data.iloc[int(size(data,0) - test): int(size(data,0)- test +q),0]),
                    np.array(predictions_dictionary[key]), label = key, linestyle = 'dashed', marker = 'o')
        ax.set_xlabel("Time (Quarters)")
        ax.set_ylabel(name_list[column])
        ax.legend()
        ax.grid(True)
        plt.title("Forecasts of " + str(name_list[column]) +  " Using FAR(p,L)")
        plt.show()

```

```
    else:
        pass

    if plot_errors == True:
        forecasting_errors = {}
        dict_keys = list(predictions_dictionary.keys())
        for key in predictions_dictionary.keys():
            forecasting_errors[dict_keys[dict_keys.index(key)]] =\
                np.array(data.iloc[-q:, column]).reshape((len(data.iloc[-q:, column]),1)) \
                    -predictions_dictionary[key]

        #Plot Forecasting Errors
        fig, ax = plt.subplots(figsize=(10, 5))

        for key in forecasting_errors.keys():
            ax.plot(forecasting_errors[key], label = key)

        ax.set_xlabel("Time (Quarters)")
        ax.set_ylabel("Forecasting Errors")
        ax.legend()
        plt.title("FAR(p,L) Forecasting Errors")
        ax.grid(True)

    if evaluation == True:

        MSE = []
        RMSE = []
        MAE = []

        for key in predictions_dictionary.keys():
            MSE_key = mean_squared_error(predictions_dictionary[key].values,
                    data.iloc[int(size(data,0) - test): int(size(data,0)- test +q):,column].values)

            MSE.append(MSE_key)

            RMSE_key = np.sqrt(mean_squared_error(predictions_dictionary[key].values,
                    data.iloc[int(size(data,0) - test): int(size(data,0)- test +q),column].values))

            RMSE.append(RMSE_key)

            MAE_key = mean_absolute_error(data.iloc[int(size(data,0) - test):
                                                    int(size(data,0)- test +q),column].values,
                                                    predictions_dictionary[key].values)
            MAE.append(MAE_key)

        return np.array(MSE), np.array(RMSE), np.array(MAE)
```

Appendix G

The 'Exact' Log-Likelihood for the VAR(p) Model

It is typically the case that when estimating the VAR(p) model, the initial p observations are omitted from the sample. To see this, begin by re-examining the probabilistic reduction from the VAR(p) specification under the probabilistic assumptions that the intrinsic vector stochastic process $\{\mathbf{Z}_t, t \in \mathbb{N}\}$, where each $\mathbf{Z}_t$ is a $(k \times 1)$ vector, is (i) multivariate normal, (ii) Markov dependent of order p, and (iii) Second-order stationary:

$$\begin{aligned} f(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n; \boldsymbol{\psi}) &= f(\mathbf{Z}_1^p; \boldsymbol{\phi}_0) \prod_{t=p+1}^{n} f_t(\mathbf{Z}_t | \mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}_t) \\ &\overset{\mathrm{M}}{=} f(\mathbf{Z}_1^p; \boldsymbol{\phi}_0) \prod_{t=p+1}^{n} f_t(\mathbf{Z}_t | \mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}_t) \\ &\overset{\mathrm{M/S}}{=} f(\mathbf{Z}_1^p; \boldsymbol{\phi}_0) \prod_{t=p+1}^{n} f(\mathbf{Z}_t | \mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}), \end{aligned} \tag{G.1}$$

where $\mathbf{Z}_1^p := (\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_p)$, $\mathbf{Z}_{t-1}^{t-p} := (\mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \ldots, \mathbf{Z}_{t-p})$.

Using the product of the conditional distributions, one can then form the VAR(p) model specification in Table 3.1 above, as well as the log-likelihood function. However, notice that in this case, the observations in $\mathbf{Z}_1^p$ have been ignored. Hence, the log-likelihood function omitting the first p observations will approximate the exact log-likelihood function, and the sampling distributions of the 'approximate' maximum likelihood estimators are roughly equal to those in the multivariate least squares estimation for a sufficient number of observations. For another example of the omission of the initial observations, see the derivation of the VAR(p) log-likelihood function in Hamilton (1994), where the author uses recursion to formulate $f(\mathbf{Z}_t|\mathbf{Z}_{t-1},\ldots,\mathbf{Z}_{t-p};\boldsymbol{\theta})$.

Now suppose that one wanted to include the missing observations $\mathbf{Z}_1,\ldots,\mathbf{Z}_p$ in the log-likelihood function. For simplicity, consider the instance where $p=1$, i.e. the vector stochastic process $\{\mathbf{Z}_t, t\in\mathbb{N}\}$ is Markov(1) dependent. Denote the full sample as $\mathbf{Z} := (\mathbf{Z}_1,\mathbf{Z}_2,\ldots,\mathbf{Z}_n)$. Note that from this assumption $\mathbf{Z}_1^p = \mathbf{Z}_1$. Under the assumption that the stochastic process follows a multivariate normal distribution, and that the $\mathbf{Z}_1^p$ is an element of the stochastic process, the marginal distribution $f(\mathbf{Z}_1^p;\boldsymbol{\phi}_0)$, where $\boldsymbol{\phi}_0 = (\boldsymbol{\mu},\boldsymbol{\Sigma}_{00})$, is:

$$f(\mathbf{Z}_1^p;\boldsymbol{\phi}_0) = 2\pi^{-\frac{k}{2}}|\boldsymbol{\Sigma}_{00}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\mathbf{Z}_1^p-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}_{00}^{-1}(\mathbf{Z}_1^p-\boldsymbol{\mu})\right).$$

The remainder of the sample, $\mathbf{Z}_{-1} := (\mathbf{Z}_2,\mathbf{Z}_3,\ldots,\mathbf{Z}_n)$ has the joint distribution:

$$f(\mathbf{Z}_{-1};\boldsymbol{\phi}) = \prod_{t=2}^{n} f(\mathbf{Z}_t|\mathbf{Z}_{t-1};\boldsymbol{\phi}).$$

As shown, normality implies that the distribution $f(\mathbf{Z}_t|\mathbf{Z}_{t-1};\boldsymbol{\phi})$ has the form:

$$\begin{bmatrix}\mathbf{Z}_t\\ \mathbf{Z}_{t-1}\end{bmatrix} \sim N\left(\begin{bmatrix}\boldsymbol{\mu}\\ \boldsymbol{\mu}\end{bmatrix},\begin{bmatrix}\boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01}\\ \boldsymbol{\Sigma}_{10}^\top & \boldsymbol{\Sigma}_{11}\end{bmatrix},\right) \tag{G.2}$$

where $\boldsymbol{\mu}$ is the mean vector of $\mathbf{Z}_t$ and is of vector of length $((p+1)k\times 1)$, $\boldsymbol{\Sigma}_{ij}$ is a $k\times k$ covariance matrix for $i,j=0,1,\ldots,p$. In the context of the VAR(1) model, the density function can be written as:

$$f(\mathbf{Z}_{-1};\boldsymbol{\phi}) = \prod_{t=2}^{n} 2\pi^{-\frac{k}{2}}|\boldsymbol{\Omega}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\mathbf{Z}_t - \mathbf{a}_0 - \mathbf{A}_1^\top \mathbf{Z}_{t-1})^\top \boldsymbol{\Omega}^{-1}(\mathbf{Z}_t - \mathbf{a}_0 - \mathbf{A}_1^\top \mathbf{Z}_{t-1})\right)$$
$$= 2\pi^{-\frac{kn}{2}}|\boldsymbol{\Omega}|^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\sum_{t=2}^{n}(\mathbf{Z}_t - \mathbf{a}_0 - \mathbf{A}_1^\top \mathbf{Z}_{t-1})^\top \boldsymbol{\Omega}^{-1}(\mathbf{Z}_t - \mathbf{a}_0 - \mathbf{A}_1^\top \mathbf{Z}_{t-1})\right).$$

It becomes quickly apparent that in order to construct the distribution of the sample, $f(\mathbf{Z};\boldsymbol{\phi})$, through the inclusion of $\mathbf{Z}_1^p$, the parameters in $\boldsymbol{\phi}_0$ need to be re-parameterized into those parameters found in the VAR(1) model, i.e. $\boldsymbol{\phi} := (\mathbf{a}_0, \mathbf{A}_1, \boldsymbol{\Omega})$. Hence, one can use the inherent parameterizations found in Table 3.1 to solve for the parameters in $\boldsymbol{\phi}_0$. These parameterizations are given for the case of the VAR(1) model below:

$$\begin{aligned}
\mathbf{a}_0 &= (\mathbf{I}_k - \mathbf{A}_1^\top)\boldsymbol{\mu},\\
\mathbf{A}_1^\top &= \boldsymbol{\Sigma}_{01}\boldsymbol{\Sigma}_{11}^{-1},\\
\boldsymbol{\Omega} &= \boldsymbol{\Sigma}_{00} - \boldsymbol{\Sigma}_{01}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{01}^\top.
\end{aligned}$$

Since this example is a VAR(1) model, where it is assumed that the underlying stochastic process is second-order stationary, it follows that $\boldsymbol{\Sigma}_{00} = \boldsymbol{\Sigma}_{11}$. Using the first parameterization equation to solve for the parameter $\boldsymbol{\mu}$ yields the straightforward result that:

$$\mathbf{a}_0 = (\mathbf{I}_k - \mathbf{A}_1^\top)\boldsymbol{\mu} \Rightarrow \boldsymbol{\mu} = (\mathbf{I}_k - \mathbf{A}_1^\top)^{-1}\mathbf{a}_0, \tag{G.3}$$

where $\mathbf{I}_k$ denotes a $(k \times k)$ identity matrix. The more complicated problem is the re-parameterization of $\boldsymbol{\Sigma}_{00}$. This will utilize both the second and third parameterization equations above. Some preliminary transformations:

$$\begin{aligned}
\mathbf{A}_1^\top &= \mathbf{\Sigma}_{01}\mathbf{\Sigma}_{11}^{-1} = \mathbf{\Sigma}_{01}\mathbf{\Sigma}_{00}^{-1} \\
&\Rightarrow \mathbf{\Sigma}_{01}^\top = [\mathbf{A}_1\mathbf{\Sigma}_{00}]^\top = \mathbf{\Sigma}_{00}\mathbf{A}_1^\top
\end{aligned}$$

The last line follows from the fact that $\mathbf{\Sigma}_{00}$ is assumed to be a semi-positive definite, symmetric matrix. Substitution results in:

$$\begin{aligned}
\mathbf{\Omega} &= \mathbf{\Sigma}_{00} - \mathbf{\Sigma}_{01}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{01}^\top \\
&= \mathbf{\Sigma}_{00} - \mathbf{\Sigma}_{01}\mathbf{\Sigma}_{00}^{-1}\mathbf{\Sigma}_{01}^\top \\
&= \mathbf{\Sigma}_{00} - \mathbf{A}_1\mathbf{\Sigma}_{01}^\top \\
&= \mathbf{\Sigma}_{00} - \mathbf{A}_1\mathbf{\Sigma}_{00}\mathbf{A}_1^\top.
\end{aligned}$$

Solving this equation for $\mathbf{\Sigma}_{00}$ requires the use of the *vec* operator and its properties, where the *vec* operator is defined as in Hamilton (1994):

Definition G.1 (‘vec’ Operator)**.** Let $\mathbf{X}$ be an $(k \times l)$ matrix. Then $vec(\mathbf{X})$ results in a $(kl \times 1)$ vector, which is formed by stacking the columns of $\mathbf{X}$. This stacking is done such that the columns are stacked one underneath another, from left to right.[1]

[1]For example, let

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}.$$

Then

$$vec(\mathbf{X}) = \begin{bmatrix} x_{11} \\ x_{21} \\ x_{12} \\ x_{22} \end{bmatrix}.$$

Applying the 'vec' operator to the equation for $\mathbf{\Omega}$ above:

$$\begin{aligned} vec(\mathbf{\Omega}) &= vec(\mathbf{\Sigma}_{00}) - (\mathbf{A}_1 \otimes \mathbf{A}_1)\, vec(\mathbf{\Sigma}_{00}) \\ &= (\mathbf{I}_{kk} - (\mathbf{A}_1 \otimes \mathbf{A}_1))\, vec(\mathbf{\Sigma}_{00}) \\ &\Rightarrow vec(\mathbf{\Sigma}_{00}) = (\mathbf{I}_{kk} - (\mathbf{A}_1 \otimes \mathbf{A}_1))^{-1}\, vec(\mathbf{\Omega}) \end{aligned}$$

Now, let vec^{-1} be defined as the operator such that $vec^{-1}(vec(\mathbf{X})) = \mathbf{X}$ for some matrix $\mathbf{X}$. That is, the operator vec^{-1} reverts a column vector back into a matrix of the original dimensions. Utilizing this operator, it follows that

$$vec^{-1}(vec(\mathbf{\Sigma}_{00})) = vec^{-1}\left((\mathbf{I}_{kk} - (\mathbf{A}_1 \otimes \mathbf{A}_1))^{-1}\, vec(\mathbf{\Omega})\right)$$

and thus,

$$\mathbf{\Sigma}_{00} = vec^{-1}\left((\mathbf{I}_{kk} - (\mathbf{A}_1 \otimes \mathbf{A}_1))^{-1}\, vec(\mathbf{\Omega})\right)$$

For compactness, let $\mathbf{J} = ((\mathbf{I}_{kk} - (\mathbf{A}_1 \otimes \mathbf{A}_1))$. Then

$$\mathbf{\Sigma}_{00} = vec^{-1}\left((\mathbf{J})^{-1} vec(\mathbf{\Omega})\right).$$

Putting this altogether, the joint density function $f(\mathbf{Z}_1^p; \boldsymbol{\phi}_0)$ can be rewritten as:

$$\begin{aligned} f(\mathbf{Z}_1^p; \boldsymbol{\phi}) &= 2\pi^{-\frac{k}{2}} |vec^{-1}\left((\mathbf{J})^{-1} vec(\mathbf{\Omega})\right)|^{-\frac{1}{2}} \\ &\times \exp\left(-\frac{1}{2}\left(\mathbf{Z}_1^p - (\mathbf{I}_k - \mathbf{A}_1^\top)^{-1}\mathbf{a}_0\right)^\top \left(vec^{-1}\left((\mathbf{J})^{-1} vec(\mathbf{\Omega})\right)\right)^{-1} \left(\mathbf{Z}_1^p - (\mathbf{I}_k - \mathbf{A}_1^\top)^{-1}\mathbf{a}_0\right)\right) \end{aligned}$$

The distribution of the sample, $f(\mathbf{Z};\ \boldsymbol{\phi})$, therefore equals the product:

$$
\begin{aligned}
f(\mathbf{Z};\boldsymbol{\phi}) &= f(\mathbf{Z}_1^p;\boldsymbol{\phi}) \times f(\mathbf{Z}_{-1};\boldsymbol{\phi}) \\
&= \Bigg[2\pi^{-\frac{k}{2}} |vec^{-1}\left((\mathbf{J})^{-1} vec(\boldsymbol{\Omega})\right)|^{-\frac{1}{2}} \\
&\times \exp\left(-\frac{1}{2}\left(\mathbf{Z}_1^p - (\mathbf{I}_k - \mathbf{A}_1^\top)^{-1}\mathbf{a}_0\right)^\top \left(vec^{-1}\left((\mathbf{J})^{-1} vec(\boldsymbol{\Omega})\right)\right)^{-1} \left(\mathbf{Z}_1^p - (\mathbf{I}_k - \mathbf{A}_1^\top)^{-1}\mathbf{a}_0\right)\right)\Bigg] \\
&\times \left[2\pi^{-\frac{kn}{2}} |\boldsymbol{\Omega}|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{t=2}^{n} (\mathbf{Z}_t - \mathbf{a}_0 - \mathbf{A}_1\mathbf{Z}_{t-1})^\top \boldsymbol{\Omega}^{-1} (\mathbf{Z}_t - \mathbf{a}_0 - \mathbf{A}_1\mathbf{Z}_{t-1})\right)\right] \\
&= \left(2\pi^{-k}\right) |vec^{-1}\left((\mathbf{J})^{-1} vec(\boldsymbol{\Omega})\right)|^{-\frac{1}{2}} |\boldsymbol{\Omega}|^{-\frac{1}{2}} \\
&\times \exp\Bigg(-\frac{1}{2}\left(\mathbf{Z}_1^p - (\mathbf{I}_k - \mathbf{A}_1^\top)^{-1}\mathbf{a}_0\right)^\top \left(vec^{-1}\left((\mathbf{J})^{-1} vec(\boldsymbol{\Omega})\right)\right)^{-1} \left(\mathbf{Z}_1^p - (\mathbf{I}_k - \mathbf{A}_1^\top)^{-1}\mathbf{a}_0\right) \\
&- \frac{1}{2}\sum_{t=2}^{n} (\mathbf{Z}_t - \mathbf{a}_0 - \mathbf{A}_1^\top\mathbf{Z}_{t-1})^\top \boldsymbol{\Omega}^{-1} (\mathbf{Z}_t - \mathbf{a}_0 - \mathbf{A}_1^\top\mathbf{Z}_{t-1})\Bigg)
\end{aligned}
$$

Now that the distribution of the sample is specified, the form of the 'exact' log-likelihood function for the VAR(1) model can be specified:

$$
\begin{aligned}
\log L(\boldsymbol{\phi};\mathbf{Z}) = -k\log(2\pi) &- \frac{1}{2}\log(|vec^{-1}\left((\mathbf{J})^{-1} vec(\boldsymbol{\Omega})\right)|) - \frac{1}{2}\log(|\boldsymbol{\Omega}|) \\
&- \frac{1}{2}\left(\mathbf{Z}_1^p - (\mathbf{I}_k - \mathbf{A}_1^\top)^{-1}\mathbf{a}_0\right)^\top \left(vec^{-1}\left((\mathbf{J})^{-1} vec(\boldsymbol{\Omega})\right)\right)^{-1} \left(\mathbf{Z}_1^p - (\mathbf{I}_k - \mathbf{A}_1^\top)^{-1}\mathbf{a}_0\right) \\
&- \frac{1}{2}\sum_{t=2}^{n} (\mathbf{Z}_t - \mathbf{a}_0 - \mathbf{A}_1^\top\mathbf{Z}_{t-1})^\top \boldsymbol{\Omega}^{-1} (\mathbf{Z}_t - \mathbf{a}_0 - \mathbf{A}_1^\top\mathbf{Z}_{t-1}).
\end{aligned}
$$

Given this log-likelihood function, one can go on to derive the 'exact' MLE of the full sample and their sampling distributions without relying on approximations.

Appendix H

GDP Growth Forecasting: Principal Components and the Student's *t* Vector Autoregressive Model

H.1 Student's *t* Vector Autoregressive Model

In this appendix, the forecasting framework established in Chapter 3 is extended to handle cases where assumptions other than independence do not hold. The main contribution to the literature on Principal Component Analysis is the use of the Student's t VAR model (StVAR), which is based on the results found in Spanos (1994). One clear advantage of the StVAR as opposed to the Normal VAR is that whereas the Normal model assumes that the conditional variance is not a function of the conditioning variables, the Student's t VAR model allows one to relax the homoskedasticity assumption and model heteroskedastic behavior. Moreover, the StVAR model allows the conditional variance to take into account trending means, whereas this is not the case in the Normal VAR model. Poudyal and Spanos (2022) provides additional details as to how the Student's t and Normal models differ.

Furthermore, another contribution that this method provides is the use of the probabilistic reduction approach in Spanos (1994, 2019a). This approach uses a few, key pre-specified assumptions to simplify the derivation of joint, conditional and marginal distributions utilized in the model.

H.1.1 Student's t Distribution

This appendix utilizes a multivariate Student's t distribution to create the corresponding VAR model. However, as explained in Poudyal (2012), the literature on the Student's t distribution is divided into two areas: that which studies the Multivariate Student's t distribution and that which examines the Matrix variate t distribution. The density functions for each of the types are described in more detail below.

Matrix Variate Student's t Distribution

Following the work done in Poudyal and Spanos (2022), let $\mathbf{Z}$ be a $(T \times m)$ random matrix with parameters $\boldsymbol{\mu}, \boldsymbol{\Omega}$, $\mathbf{V}$ and a degree of freedom parameter, ν. This random matrix admits a matrix variate Student's t distribution when the density function takes the form:

$$f(\mathbf{z}) = \frac{\Gamma((\nu+m+T-1)/2)}{\pi^{\frac{1}{2}Tm}\Gamma((\nu+T-1)/2)}|\nu\mathbf{V}|^{-\frac{1}{2}m}|\boldsymbol{\Omega}|^{-\frac{1}{2}T}|\mathbf{I}_T + \frac{1}{\nu}\mathbf{V}^{-1}(\mathbf{z}-\boldsymbol{\mu})^\top\boldsymbol{\Omega}^{-1}(\mathbf{z}-\boldsymbol{\mu})|^{-\frac{1}{2}(\nu+m+T-1)}$$

However, as observed in Poudyal (2012), this version of the Student's t distribution is a special case of the multivariate Student's t where the matrix t distribution heavily restricts the variance-covariance matrix.

Multivariate Student's t Distribution

In order to avoid the restrictions that exist for the matrix variate Student's t distribution, one can use the multivariate Student's t distribution. Let $\mathbf{Z}_t$ be a stochastic vector process with length $(m \times 1)$ and let $\boldsymbol{\psi} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. The density function for the multivariate case is given as:

$$f(\mathbf{Z}_t; \boldsymbol{\psi}) = (\pi\nu)^{-\frac{m}{2}} \frac{\boldsymbol{\Gamma}[\frac{\nu+m}{2}]}{\boldsymbol{\Gamma}[\frac{\nu}{2}]} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} [1 + \frac{1}{\nu}(\mathbf{Z}_t - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Z}_t - \boldsymbol{\mu})]^{-\frac{\nu+m}{2}} \tag{H.1}$$

For the remainder of the appendix, the more general multivariate Student's t will be used to both define the respective vector autoregressive model and the corresponding principal components.

Important Notes on the Student's t Distribution

One addition that the Student's t distribution has over the Normal is the degrees of freedom parameter, ν. The number of moments available, as well as the nature of the distribution depends on the value ν takes. For example, as ν approaches infinity, the Student's t distribution becomes the Normal distribution.

H.1.2 Model Specification: Probabilistic Reduction Approach

Let $\{\mathbf{Z}_t, t \in \mathbb{N}\}$ be a stochastic vector process, where $\mathbf{Z}_t : (k \times 1)$. Here, the model specification is derived via the probabilistic reduction approach, outlined in Spanos and McGuirk (2001) and Spanos (2019a). This approach requires the imposition of three assumptions on the stochastic process, specifically in the categories of (i) Distribution, (ii) Dependence, and (iii) Heterogeneity. In this instance, the it is assumed that $\mathbf{Z}_t$ is (i) multivariate Student's t,

(ii) Markov dependent (p) and (iii) second-order stationary. More specifically:

(i) Distribution: $\mathbf{Z}_t$ follows a multivariate Student's t distribution, i.e. $\mathbf{Z}_t \sim St_k(\boldsymbol{\mu}, \mathbf{V}; \nu)$, where ν is the degrees of freedom and $\mathbf{V} = \nu\boldsymbol{\Sigma}$. Therefore, the density function for $\mathbf{Z}_t$ takes the following form:

$$f(\mathbf{Z}_t; \boldsymbol{\psi}) = (\pi\nu)^{-\frac{m}{2}} \frac{\boldsymbol{\Gamma}[\frac{\nu+m}{2}]}{\boldsymbol{\Gamma}[\frac{\nu}{2}]} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} [1 + \frac{1}{\nu}(\mathbf{Z}_t - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Z}_t - \boldsymbol{\mu})]^{-\frac{\nu+m}{2}} \tag{H.2}$$

Note that in equation (4.2), $\boldsymbol{\psi} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. This assumption is the key distinction from the typical, normal VAR model.

(ii) Dependence: $\mathbf{Z}_t$ is Markov dependent of order p, where Markov dependent of order p is as defined in Spanos (1986, 2019a).

(iii) Heterogeneity: $\mathbf{Z}_t$ is second-order, or covariance, stationary, i.e. $E(\mathbf{Z}_t) = \boldsymbol{\mu}$ and $Cov(\mathbf{Z}_t, \mathbf{Z}_{t-p}) = E((\mathbf{Z}_t - \boldsymbol{\mu})(\mathbf{Z}_{t-p} - \boldsymbol{\mu})^\top) = \boldsymbol{\Sigma}(p)$.

From assumptions (i)-(iii) above, the Probabilistic Reduction approach in Spanos (1994, 2019a) is used to simplify the joint distribution of the sample, $D(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n; \boldsymbol{\psi})$, which takes the form:

$$\begin{aligned} D(\mathbf{Z}_1, \ldots, \mathbf{Z}_n; \boldsymbol{\psi}) &\overset{\text{M}}{=} D(\mathbf{Z}_1^p; \boldsymbol{\phi}_0) \prod_{t=p+1}^{n} D_t(\mathbf{Z}_t | \mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}_{1,t}) D(\mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}_{2,t}) \\ &\overset{\text{M/S}}{=} D(\mathbf{Z}_1^p; \boldsymbol{\phi}_0) \prod_{t=p+1}^{n} D(\mathbf{Z}_t | \mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}_1) D(\mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}_2), \end{aligned} \tag{H.3}$$

where $\mathbf{Z}_1^p := (\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_p)$ and $\mathbf{Z}_{t-1}^{t-p} := (\mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \ldots, \mathbf{Z}_{t-p})$. This reduction is almost identical to that for the Normal VAR model found in Chapter 3. However, notice there is now an additional term, particularly the marginal distribution $D(\mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}_2)$. This inclusion follows directly due to a lack of *weak exogeneity* in Student's t model; see Spanos (1994, 2019a) for the definition of *weak exogeneity* and for more discussion. Once again, the first p observations are omitted, resulting in the approximate version of the log-likelihood function when performing Maximum Likelihood Estimation.

Thus, keeping the notation similar to that found in the previous work of Poudyal (2012), let $\mathbf{x}_t := [\mathbf{Z}_t, \mathbf{Z}_{t-1}, \ldots, \mathbf{Z}_{t-p}]^\top$. Given the reduction presented above, the joint distribution of $\mathbf{x}_t$ takes the form:

$$\mathbf{x}_t \sim St\left[\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_Z \\ \boldsymbol{\mu}_Z \\ \vdots \\ \boldsymbol{\mu}_Z \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01} & \boldsymbol{\Sigma}_{02} & \ldots & \boldsymbol{\Sigma}_{0p} \\ \boldsymbol{\Sigma}_{10}^\top & \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \ldots & \boldsymbol{\Sigma}_{1p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{p0}^\top & \boldsymbol{\Sigma}_{p1}^\top & \boldsymbol{\Sigma}_{p2}^\top & \ldots & \boldsymbol{\Sigma}_{pp} \end{pmatrix}; \nu\right], \tag{H.4}$$

where $\boldsymbol{\mu}_Z$ is the mean vector of $\mathbf{Z}_t$, $\boldsymbol{\mu}$ is of vector of length $((p+1)k \times 1)$, and $\boldsymbol{\Sigma}_{ij}$ is a $k \times k$ covariance matrix for $i, j = 0, 1, \ldots, p$. Since $\mathbf{x}_t$ follows a Student's t distribution, the density function is given as:

$$f(\mathbf{x}_t; \boldsymbol{\psi}) = (\pi\nu)^{-\frac{m}{2}} \frac{\boldsymbol{\Gamma}[\frac{\nu+m}{2}]}{\boldsymbol{\Gamma}[\frac{\nu}{2}]} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} [1 + \frac{1}{\nu}(\mathbf{x}_t - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}_t - \boldsymbol{\mu})]^{-\frac{\nu+m}{2}}. \tag{H.5}$$

In this case $m = (p+1)k$.

Conditional Distribution and Parameterization

Once the mean and variance-covariance matrices are estimated, they can then be used to derive the conditional mean, conditional variance and estimated equation coefficients. Following Poudyal (2012) and Poudyal and Spanos (2022), the vectors $\mathbf{x}_t$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ are partitioned into the following:

$$\mathbf{x}_t = \begin{bmatrix} \mathbf{Z}_t, (k \times 1) \\ \mathbf{Z}_{t-1}^{t-p}, (pk \times 1) \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_Z, (k \times 1) \\ \boldsymbol{\mu}_{pk}, (pk \times 1) \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{00}, (k \times k) & \boldsymbol{\Sigma}_{01}, (k \times pk) \\ \boldsymbol{\Sigma}_{10}^{\top}, (pk \times k) & \boldsymbol{\Sigma}_{11}, (pk \times pk) \end{bmatrix};$$

Using the partitioning above, one can derive the joint, conditional and marginal distributions as done in Poudyal (2012) and Spanos (1994):

$$D(\mathbf{Z}_t, \mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\psi}) = D(\mathbf{Z}_t|\mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}_1) D(\mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}_2) \sim St(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu) \tag{H.6}$$

$$D(\mathbf{Z}_t|\mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}_1) \sim St(\mathbf{a}_0 + \mathbf{A}^{\top}\mathbf{Z}_{t-1}^{t-p}, (\frac{\nu}{\nu + pk - 2})\boldsymbol{\Omega} q(\mathbf{Z}_{t-1}^{t-p}); \nu + pk) \tag{H.7}$$

$$D(\mathbf{Z}_{t-1}^{t-p}; \boldsymbol{\phi}_2) \sim St(\boldsymbol{\mu}_{pk}, \boldsymbol{\Sigma}_{11}; \nu) \tag{H.8}$$

Note that in this context:

$$\mathbf{A}^{\top} = \boldsymbol{\Sigma}_{01}\boldsymbol{\Sigma}_{11}^{-1} \tag{H.9}$$

$$\mathbf{a}_0 = \boldsymbol{\mu}_z - \mathbf{A}^{\top}\boldsymbol{\mu}_{pk} \tag{H.10}$$

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma}_{00} - \boldsymbol{\Sigma}_{01}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{01}^{\top} \tag{H.11}$$

$$q(\mathbf{Z}_{t-1}^{t-p}) = [1 + \frac{1}{\nu}(\mathbf{Z}_{t-1}^{t-p} - \boldsymbol{\mu}_{pk})^{\top}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{Z}_{t-1}^{t-p} - \boldsymbol{\mu}_{pk})] \tag{H.12}$$

and that $\boldsymbol{\phi}_1 := \{\mathbf{a}_0, \boldsymbol{\mu}, \mathbf{A}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}_{11}\}$ and $\boldsymbol{\phi}_2 = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}_{11}\}$. It is also important to consider the lack of variation freeness, and subsequently the violation of weak exogeneity, in this context, since it is clear that both $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ share the same elements. Hence, as mentioned in the reduction of the joint distribution, one cannot ignore the marginal distribution in equation (4.8). With this parameterization, one can fully specify the Student's t VAR model found in Poudyal and Spanos (2022), which is summarized in Table H.1 .

Table H.1: Student's t Vector Autoregressive (VAR(p)) Model

	Statistical GM: $\mathbf{Z}_t = \mathbf{a}_0 + \mathbf{A}_1^\top \mathbf{Z}_{t-1}^{t-p} + \mathbf{u}_t,\ \ t\in\mathbb{N}$,	
[1]	Student's t:	$D(\mathbf{Z}_t\vert\mathbf{Z}_{t-1}^{t-p};\boldsymbol{\theta})$, for $\mathbf{Z}_{t-1}^{t-p}$:=$(\mathbf{Z}_{t-1},...\mathbf{Z}_1)$ is Student's t with d.o.f ν,
[2]	Linearity:	$E(\mathbf{Z}_t\vert\sigma(\mathbf{Z}_{t-1}^{t-p}))$=$\mathbf{a}_0 + \mathbf{A}^\top\mathbf{Z}_{t-1}^{t-p}$,
[3]	Heteroskedasticity:	$Var(\mathbf{Z}_t\vert\sigma(\mathbf{Z}_{t-1}^{t-p}))$=$\left(\frac{\nu}{\nu+pk-2}\right)\Omega q(\mathbf{Z}_{t-1}^{t-p})$,
		$q(\mathbf{Z}_{t-1}^{t-p})$=$1 + \frac{1}{\nu}\left(\ \mathbf{Z}_{t-1}^{t-p} - \boldsymbol{\mu}_{pk}\ \right)^\top \boldsymbol{\Sigma}_{11}^{-1}\left(\ \mathbf{Z}_{t-1}^{t-p} - \boldsymbol{\mu}_{pk}\ \right)$
[4]	Markov(p):	$\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$ is a Markov(p) process,
[5]	t-invariance:	$\boldsymbol{\theta}$:=$(\mathbf{a}_0, \mathbf{A}, \Omega, \boldsymbol{\Sigma}_{11})$ are t-invariant.

Degrees of Freedom Assumption

In this appendix, it is assumed that the degrees of freedom parameter ν is the same for each of the predictors included in the StVAR model. While this is most likely not the case, extending the model to take into account varying degrees of freedom can be considered in future research. This issue was also encountered in Poudyal (2012) and has not been adjusted for here.

H.1.3 StVAR Model Estimation

The method of estimation for the StVAR model parameters is Maximum Likelihood Estimation (MLE). Here, the log-likelihood function for $\mathbf{x}_t$ follows from the density function in equation (H.5) above,

$$L(\boldsymbol{\psi};\mathbf{x}_t) = c - \frac{1}{2}ln|\boldsymbol{\Sigma}| - \frac{1}{2}(\nu+m)ln[1+\frac{1}{\nu}(\mathbf{x}_t-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x}_t-\boldsymbol{\mu})], \tag{H.13}$$

where

$$c = ln\Gamma(\frac{\nu+m}{2}) - ln\Gamma(\frac{\nu}{2}) - \frac{m}{2}ln\pi\nu.$$

Since there are T observations, the log-likelihood function for the sample with an identical distribution assumption is:

$$L(\boldsymbol{\psi}; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T) = Tc - \frac{T}{2}ln|\boldsymbol{\Sigma}| - \frac{1}{2}(\nu+m)\sum_{t=1}^{T} ln[1 + \frac{1}{\nu}(\mathbf{x}_t - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}_t - \boldsymbol{\mu})]. \quad \text{(H.14)}$$

Once the parameters in $\boldsymbol{\psi}$ are estimated, the StVAR model parameters are obtained via the parameterization outlined above. As the number of parameters to estimate increases rather quickly with additional lags or predictors, estimation becomes more computationally challenging. This increases the time needed to optimize the log-likelihood. In this case, an iterative optimization algorithm, such as Sequential Least Squares Programming, can be used to find the maximum.

H.2 The Student's *t* VAR Principal Components: Forecasting Methodology

The goal of this appendix is to connect the Student's t VAR model with conventional principal component analysis. Let $\{\mathbf{Z}_t, t \in \mathbb{N}\}$ be a vector stochastic process, where each random vector $\mathbf{Z}_t$ is of length $(k \times 1)$. Recall the baseline definition of the Normal Principal Components model, detailed in Chapter 1, which is fully described in the following table:

Table H.2: Normal Principal Components Model

Statistical GM:	$\mathbf{C}_t = \boldsymbol{\Gamma}_p^{\top}(\mathbf{Z}_t - \boldsymbol{\mu}) + \mathbf{u}_t,$	
[1]	Normality:	$\mathbf{C}_t \sim N(.,.)$
[2]	Linearity:	$E(\mathbf{C}_t) = \boldsymbol{\Gamma}_p^{\top}(\mathbf{Z}_t - \boldsymbol{\mu}),$
[3]	Constant Covariance:	$Cov(\mathbf{C}_t) = \boldsymbol{\Lambda}_p,$
[4]	Independence:	$\{\mathbf{C}_t, t \in \mathbb{N}\}$ is an independent process
[5]	t-invariance:	$\boldsymbol{\theta} := \{\boldsymbol{\mu}, \boldsymbol{\Lambda}_p, \boldsymbol{\Gamma}_p\}$ does not change with t

In Chapter 3, the Normal Principal Components model was further expanded upon to account for the temporal dependence inherent in many macroeconomic indicators. The result of this extension is the the Normal, Temporal Principal Components model given in Table H.3, which loosens the assumption of Independence instead replaces it with the assumption of Markov Dependence(L). To accommodate the aforementioned change, principal components were derived from the Normal Vector Autoregressive model (VAR), outlined in Chapter 2. The main connection between these two models arises from the estimated temporal covariance matrix of the VAR model.

Table H.3: Normal, Temporal Principal Components Model

Statistical GM:	$\mathbf{F}_t = \mathbf{V}_r^{\top}(\mathbf{Y}_t - \boldsymbol{\mu}_y) + \mathbf{u}_t,\ \mathbf{Y}_t = [\mathbf{Z}_t, \mathbf{Z}_{t-1}, \ldots, \mathbf{Z}_{t-q}]^{\top},\ r \leq (L+1)k$	
[1]	Normality:	$\mathbf{F}_t \sim N(.,.)$
[2]	Linearity:	$E(\mathbf{F}_t) = \mathbf{V}_r^{\top}(\mathbf{Y}_t - \boldsymbol{\mu}_y),$
[3]	Constant Covariance:	$Cov(\mathbf{F}_t) = \mathbf{D}_r,$
[4]	Markov:	$\{\mathbf{Z}_t, t \in \mathbb{N}\}$ is an Markov(p) process
[5]	t-invariance:	$\boldsymbol{\theta} := \{\boldsymbol{\mu}_y, \mathbf{D}_r, \mathbf{V}_r\}$ does not change with t

In this appendix, an additional assumption of the baseline Principal Components model is being altered. That is, the assumption of normality is being replaced with a different distributional assumption, in addition the relaxing of independence. In this context, once the temporal covariance matrix, $\hat{\mathbf{\Sigma}}$, from the Student's t VAR model is estimated via the MLE method prescribed above, the eigenvalues and respective eigenvectors are extracted using the spectral decomposition. Then, upon selecting the p largest eigenvalues, p principal components are constructed using the corresponding eigenvectors. These principal components coincide with the "temporal principal components" presented in Chapter 3.

H.2.1 Forecasting Methodology

Since the basis of the principal components used in this chapter follow closely to those defined previously, the forecasting methodology will also be similar to that in Chapter 3. However, the models used in the previous chapter assumed both a normal distribution and homoskedasticity, whereas now those assumptions are being relaxed.

Summary of Steps

In summary, the updated forecasting procedure in this chapter consists of the following steps:

Step 1. Determine the number of trends and lags via statistically adequate AR(p) and VAR(L) models.

Step 2. (*New*) If the assumption of homoskedasticity is violated in the first step, introduce the the Student's t VAR(L) model to ensure statistical adequacy. Recheck the residuals via Mis-specification testing.

Step 3. Detrend and de-seasonalize the data using the estimated trends from the statistically adequate VAR(L) or StVAR(L) model.

Step 4. Estimate the covariance matrix from the VAR(L) or StVAR(L) model.

Step 5. Estimate the temporal principal components using spectral decomposition or SVD.

Step 6. Choose which components to keep based on statistical significance, scree plot, etc.

Step 7. Estimate the Dynamic Factor-Augmented Autoregressive model (FAR(p,L)).

Step 8. Check the statistical adequacy of the FAR(p,L) using M-S testing.

Step 9. Construct out-of-sample forecasts iteratively over the forecast horizon.

Appendix I

Factor Augmented VAR Model Benchmark

I.1 Model Specification

One of the benchmark forecasts presented in Chapter 3 is that of the Factor Augmented Vector Autoregressive model (FAVAR) created in Bernanke et al. (2005). Bernanke et al. (2005) expands on the DFM literature through the definition of the FAVAR model to study the impacts of monetary policy on the macroeconomy. The purpose of this benchmark is to show that the forecasting methodology in Chapter 3 can produce competitive results, in terms of forecasting accuracy, when compared to prevalent methods that ignore statistical adequacy. Here, one of the forecasting methodologies presented in Bernanke et al. (2005) is similar to Stock and Watson (2002b), where principal components are used to estimate the latent factors in the model. For some background, the underlying representation of the DFM in Stock and Watson (2002b) used to produce their forecasts is given as:

$$y_{t+1} = \beta F_t + \gamma(L) y_t + \epsilon_{t+1}$$

$$X_t = \Lambda F_t + u_t,$$

where y_t is the target series, X_t is a vector of predictors of dimension $(N \times 1)$. Here, $F_t = (f'_t, \ldots, f'_{t-q})$ is an $(r \times 1)$ vector of unobserved factors,[1] L is the lag operator,[2] and u_t is the idiosyncratic disturbance of length $(r \times 1)$. Forecasting is carried out in Stock and Watson (2002b) in two-steps. Namely, in the first step, the training data is used to estimate the unobserved factors via the method of principal components. Subsequently, the unobserved factors are used as independent variables in the forecasting regression.

Returning to Bernanke et al. (2005), the FAVAR is specified as follows. Using the same notation found in the paper, let Y_t be a $(K \times 1)$vector denoting the macroeconomic variables which move the economy. Furthermore, let F_t be an $(R \times 1)$ vector of unobserved factors, where R is less than the number of original data series, and let X_t be the vector of predictors of length $(S \times 1)$. The model in Bernanke et al. (2005) is specified via transition and observation equations:

$$\begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \Phi(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + u_t \tag{I.1}$$

$$X_t = \Lambda^f F_t + \Lambda^y Y_t + \epsilon_t \tag{I.2}$$

,

where $\Phi(L)$ is a lag polynomial similar to that found in Hamilton (1994). In the second equation, Λ^f denotes the factor loadings. To estimate this model, the authors provide two potential methodologies. The first, which is utilized in this paper, is the two-step procedure given in Stock and Watson (2002b) above. Namely, principal components of X_t are used

[1] It is assumed that there are a finite number of lags, q for the factors.

[2] In Stock and Watson (2002b), the lag polynomial is denoted as $\gamma(L) = \sum_{j=0}^{p} \gamma_j L^j$, where p is the number of lags of the target variable.

to estimate the unobserved factors (Bernanke et al., 2005). These factor estimates are then used to estimate the VAR model depicted in equation I.1.

I.2 Seasonal Adjustment and Data Trends

To create the benchmark forecast presented in Chapter 3, the data was first run through the X-13 ARIMA-SEATS Seasonal Adjustment Program, maintained by the U.S. Census Bureau.[3] To reiterate, this type of seasonal adjustment was not considered for the FAR(p,L) forecasts in Chapter 3, because the ARIMA model underlying the software could be statistically mis-specified, and the forecasting methodology above directly estimated the seasonal trends. However, many papers, such as Bok et al. (2018), start with the seasonally adjusted data. That is, seasonality is largely assumed to have been corrected prior to the analysis/forecasting. As such, the benchmark model assumes that the seasonal component of the data has been removed prior to analysis. This allows for the same seasonal adjustment across the training and testing periods. Moreover, it is common practice to remove other trends through the use of natural logarithms and differences; see the data methodology outlined in Stock and Watson (2002b) for an example. Following Stock and Watson (2002b), natural logarithms were applied for data not already in percents. Table I.1 below depicts the data transformations applied to each series. The methodology of Stock and Watson (2002b) further assumes that the data is I(0), which implies that the data is subsequently transformed via first differences.

Moreover, as in Stock and Watson (2002b), the data was combed for outliers using the interquartile range, and values above and below ten times the bounds were replaced with missing values. To make the transformed data amenable to principal components, it was

[3] The link for the download is: https://www.census.gov/data/software/x13as.html

Table I.1: Benchmark: Data Transformations

Data Series	Transformation
All Employees: Total Nonfarm	Log-differences
Real Gross Domestic Product	Log-differences
Retail Sales and Food Services	Log-differences
CPI-U: All Items	Log-differences
House Prices (S&P|Case-Shiller U.S. National Home Price Index)	Log-differences
Housing Starts	Log-differences
Civilian Unemployment Rate	Differences
Industrial Production Index	Log-differences
Merchant Wholesalers: Inventories Total	Log-differences
Philly Fed Mfg. Business Outlook: Current Activity	Differences
Capacity Utilization	Differences
Stock Price Index (SP500): SPINDX	Log-differences

then standardized to have mean zero and variance equal to one.

I.3 Benchmark Forecasts

The benchmark model splits the data into training and test sets to align with Chapter 3. That is, the data up until Q4 2016 is considered for the training dataset and the remaining data is used for the testing set. The target variable for this exercise is Real GDP, as these results were depicted in Chapter 3. Thus, the GDP variable was separated from the other predictors. As done in Stock and Watson (2002b) and Bernanke et al. (2005), principal components are used as estimates of the latent factors in the DFM, FAVAR, respectively. In this case, these principal components are estimated from the training set predictors, and the corresponding weights are used to transform the data in the testing dataset. In order to compare the results of the benchmark with those given earlier, a threshold of 95% variance explained was chosen. This amounted to the inclusion of eight principal components, which is depicted in Figure I.1.

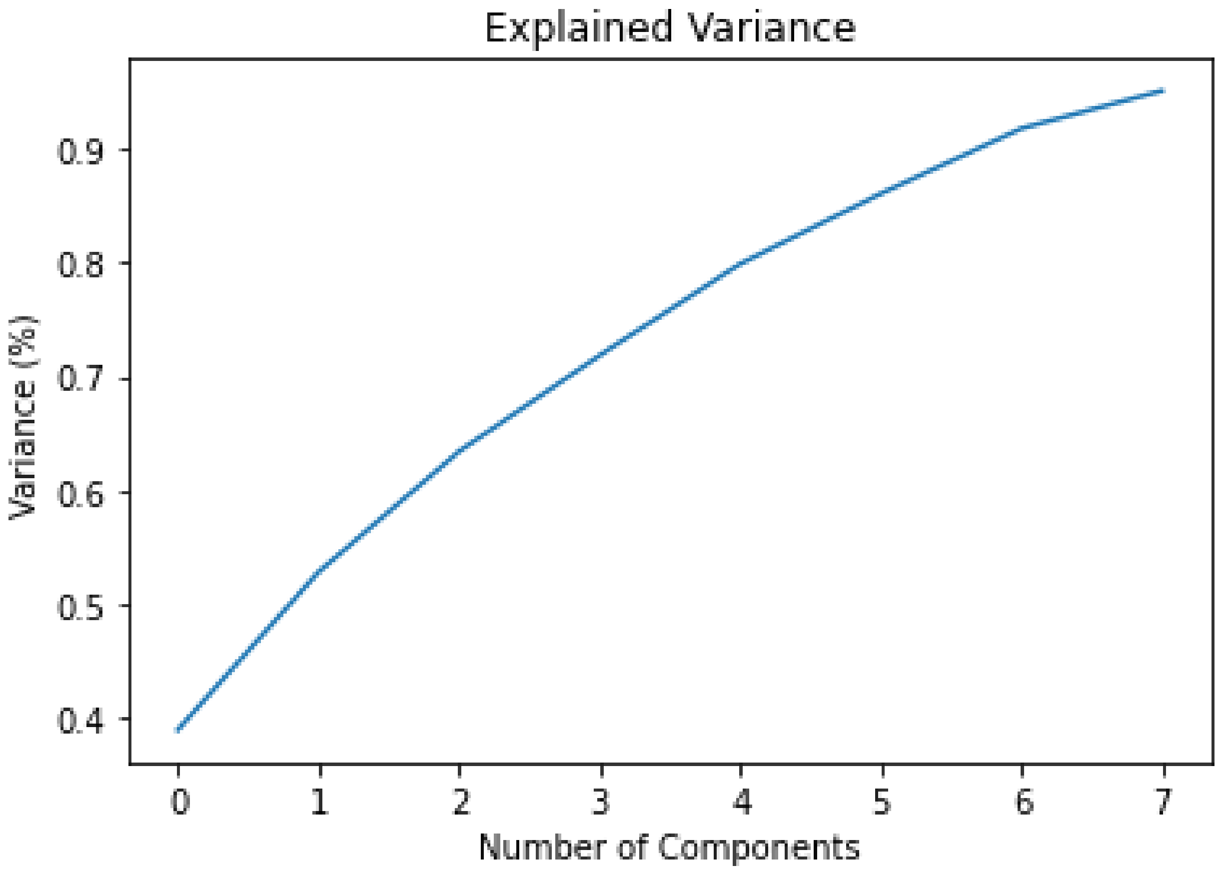


Figure I.1: Benchmark: Explained Variance Plot

The number of lags for the factors were chosen to be three based on the AIC, whereas BIC is used in Stock and Watson (2002b). Coincidentally, this is the same number of lags used in the statistically adequate VAR model presented in Chapter 3. Forecasts are generated using the same recursive methodology applied in Chapter 3, see 3.3.8 for more elaboration on this step. The resulting forecasts are depicted in Figure I.2. Note that the standardization to mean zero and unit variance was first undone by multiplying each forecast value with the sample deviation and then adding the sample mean.

Reversing the log-difference transformation produces the final benchmark forecasts in Figure I.3.[4] When looking at the forecasts closely, a few details are immediately noticed. First,

[4]The inverse of the difference transformation is given as the cumulative sum of the last training period

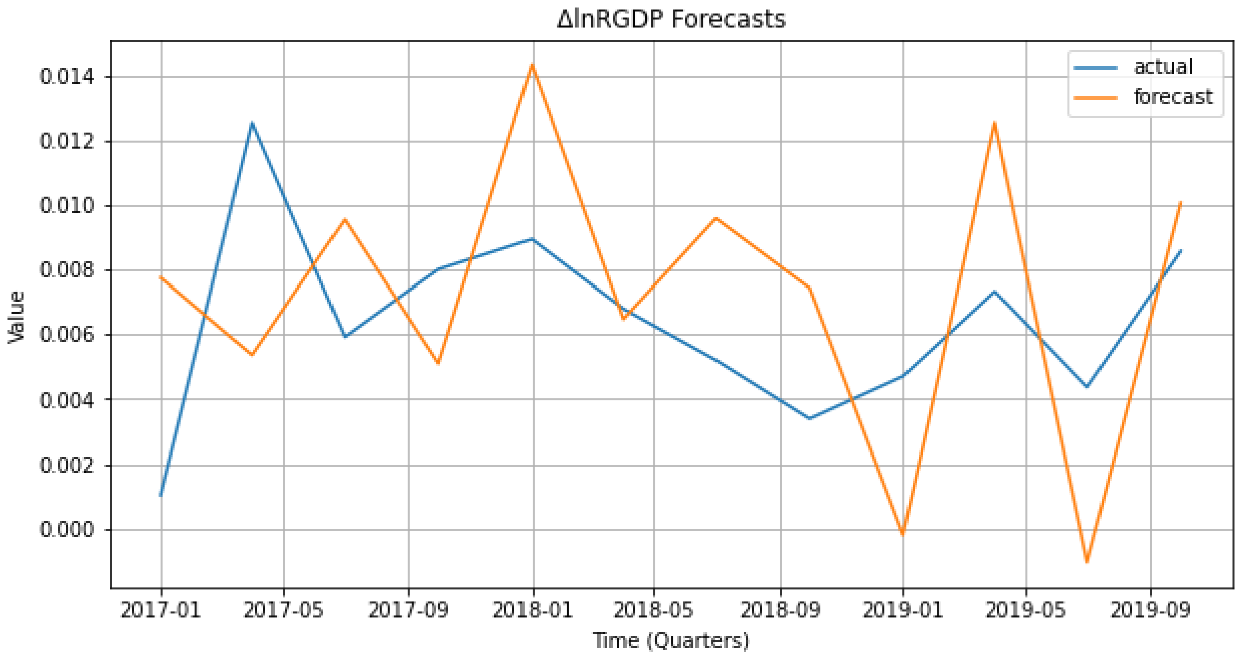


Figure I.2: Benchmark DFM: Log-Difference RGDP Forecast

the benchmark forecasts lack seasonality. This, however, ties back into the assumptions of the benchmark. Namely, the benchmark assumes that the model does not take statistical adequacy into consideration and that the seasonality has been subtracted out of the data prior to the forecasting exercise. Second, the forecast appears to generally overpredict, which results in elevated forecasting errors over the longer horizon. This result is indicative of mis-specified model. It is also important to keep in mind that to achieve the same threshold of the percentage of variance explained, 95%, eight principal components were required, out of a total of eleven predictors. This fact, coupled with the number of lags, resulted in a large number of regressors in the prediction equation, which usually leads to model overfitting. In contrast, the FAR(p,L) model defined in Chapter 3, the number of principal components was reduced based on their statistical significance, keeping in mind the goal of maintaining statistical adequacy.

---

entry and the forecasted values

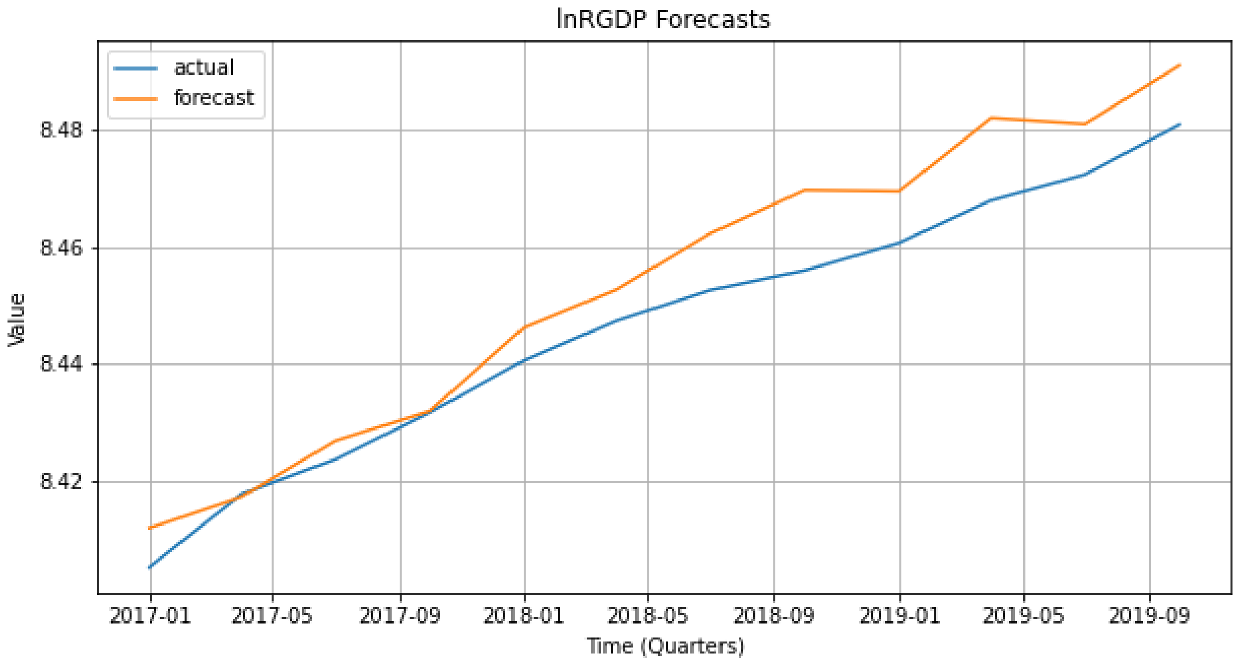


Figure I.3: Benchmark FAVAR: lnRGDP Forecast

I.4 depicts a comparison between the forecasted values and the seasonally unadjusted, original data. Clearly, the adjusted forecasts cannot predict the seasonal cycles in the original data, as that component was subtracted out using X-13 ARIMA-SEATS. In this case, accuracy measures, such as the MSFE, produce elevated levels when compared to those depicted from the statistically adequate forecasts. However, these results are not reported, as it does not make much sense to compare the seasonally adjusted forecasts with the seasonally unadjusted data.

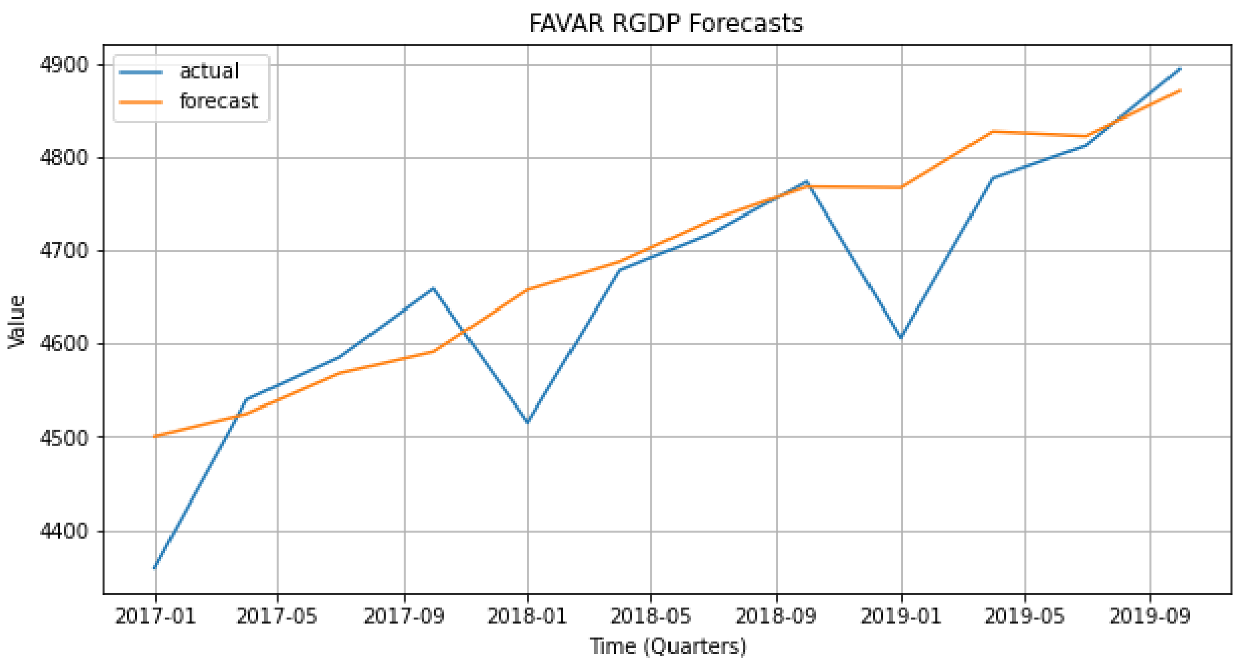


Figure I.4: Benchmark FAVAR: lnRGDP Forecast