Generalized Principal Component Analysis

Karo Solat

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
Esfandiar Maasoumi
Kwok Ping Tsang
Richard A. Ashley
Eric A. Bahel

April 30, 2018
Blacksburg, Virginia

Keywords: Factor Model, PCA, Elliptically Contoured Distributions, Exchange Rate, Forecasting, Monte Carlo Simulation, Statistical Adequacy

Copyright 2018, Karo Solat

Generalized Principal Component Analysis

Karo Solat

(ABSTRACT)

The primary objective of this dissertation is to extend the classical Principal Components Analysis (PCA), aiming to reduce the dimensionality of a large number of Normal interrelated variables, in two directions. The first is to go beyond the static (contemporaneous or synchronous) covariance matrix among these interrelated variables to include certain forms of temporal (over time) dependence. The second direction takes the form of extending the PCA model beyond the Normal multivariate distribution to the Elliptically Symmetric family of distributions, which includes the Normal, the Student's t, the Laplace and the Pearson type II distributions as special cases. The result of these extensions is called the Generalized principal component analysis (GPCA).

The GPCA is illustrated using both Monte Carlo simulations as well as an empirical study, in an attempt to demonstrate the enhanced reliability of these more general factor models in the context of out-of-sample forecasting. The empirical study examines the predictive capacity of the GPCA method in the context of Exchange Rate Forecasting, showing how the GPCA method dominates forecasts based on existing standard methods, including the random walk models, with or without including macroeconomic fundamentals.

Generalized Principal Component Analysis

Karo Solat

(GENERAL AUDIENCE ABSTRACT)

Factor models are employed to capture the hidden factors behind the movement among a set of variables. It uses the variation and co-variation between these variables to construct a fewer latent variables that can explain the variation in the data in hand. The principal component analysis (PCA) is the most popular among these factor models.

I have developed new Factor models that are employed to reduce the dimensionality of a large set of data by extracting a small number of independent/latent factors which represent a large proportion of the variability in the particular data set. These factor models, called the generalized principal component analysis (GPCA), are extensions of the classical principal component analysis (PCA), which can account for both contemporaneous and temporal dependence based on non-Gaussian multivariate distributions.

Using Monte Carlo simulations along with an empirical study, I demonstrate the enhanced reliability of my methodology in the context of out-of-sample forecasting. In the empirical study, I examine the predictability power of the GPCA method in the context of "Exchange Rate Forecasting". I find that the GPCA method dominates forecasts based on existing standard methods as well as random walk models, with or without including macroeconomic fundamentals.

Dedication

I would like to dedicate this dissertation to Professor Phoebus J. Dhrymes (1932-2016) and Professor Theodore Wilbur Anderson (1918-2016).

Acknowledgments

First and foremost, I would like to express my deepest appreciation and gratitude toward my advisor Prof. Aris Spanos for his guidance, continuous support, and immense knowledge. Without his guidance and persistent help this dissertation would not have been possible.
Also, I would like to thank my committee member, Prof. Kwok Ping Tsang, whose guidance and knowledge played a crucial part of the empirical study of this dissertation.
My sincere thanks also goes to Prof. Esfandiar Maasoumi for accepting to be a member of my advisory committee and offering me his valuable guidance.
I would like to thank the rest of my committee members Prof. Richard A. Ashley, and Prof. Eric A. Bahel, for their encouragement, insightful comments, and hard questions.
Last but not the least, I would like to thank my family: my mother Khadijeh Shahveysi, my sister Delnia Solat and my brother Kaveh Solat, for supporting me spiritually throughout my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 An Overview

The method of Principal Component Analysis (PCA) is a multivariate technique widely used to reduce the dimensionality of data summarized in the form of a variance-covariance matrix ellipsoid by rotating the coordinate system to render the resulting components uncorrelated. The classical PCA models use eigenvalue decomposition methods on the contemporaneous data covariance matrix to extract the uncorrelated principal components. This allows the modeler to retain only the components that cover a significantly high portion of the variation in the data.

The origin of the PCA method is not easy to trace back historically because the mathematics for the spectral decomposition of a matrix have been known since the late 19^{th} century and the initial application of Singular Value Decomposition (SVD) to a data matrix. The reason is that statistical analysts up until the 1920s did not distinguish between the variance-covariance parameters ($\mathbf{\Sigma}$) and their estimates ($\widehat{\mathbf{\Sigma}}$). The first to point out this important distinction is Fisher [1922]. The SVD method, which is considered as the building blocks of PCA, and its connection to the components of a correlation ellipsoid, have been presented in Beltrami [1873], Jordan [1874], and Galton [1889]. However, it is widely accepted that the full description of PCA method was first introduced in Pearson [1901] and Hotelling [1933].

This dissertation proposes a twofold extension of the classical PCA. The first replaces the Normal distribution with the Elliptically Symmetric family of distributions, and the second allows for the existence of both contemporaneous and temporal dependence. It is shown that the Maximum Likelihood Estimators (MLEs) for the Generalized PCA (GPCA) are both unbiased and consistent. In the presence of temporal dependence, the unbiasedness of the MLEs depends crucially on the nature of the non-Gaussian distribution and the type of temporal dependence among the variables involved.

Section 1.2 briefly summarizes the classical PCA with a view to bring out explicitly all the underlying probabilistic assumptions imposed on the data, as a prelude to introducing the GPCA and proposing a parameterization of the GPCA as a regression-type model. This is motivated by the fact that oftentimes discussions of the PCA emphasize the mathematical/geometric aspects of this method with only passing references to the underlying probabilistic assumptions. Chapter 2 introduces the definition and notation of a Matrix Variate Elliptically Contoured distribution along with a few representative members of this family. Chapter 3 presents the Generalized Principal Component Analysis (GPCA) model together with its underlying probabilistic assumptions and the associated estimation results. Chapter 4, presents two Monte Carlo simulations associated with the Normal vector autoregressive (Normal VAR) and the Student's t vector autoregressive (StVAR) models, to illustrate the predictive capacity of the GPCA when compared to the PCA. We show that when there is temporal dependence in the data, the GPCA dominates the PCA in terms of out-of-sample forecasting.

Chapter 5 illustrates the estimation results associated with GPCA model by applying the method to a panel of 17 exchange rates of OECD countries and use the deviations from the components to forecast future exchange rate movements, extending the results in Engel et al. [2015]. We find that the GPCA method dominates on forecasting grounds several existing standard methods as well as the random walk model, with or without including macroeconomic.

1.2 Principal Component Analysis

Let $\mathbf{X}_t{:=}(X_{1t},...,X_{mt})^\top\sim\ \mathsf{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}),\ \ t{\in}\mathbb{N}{:=}(1,...,T,...)$[1], be a $m\times 1$ random vector, and $\mathbf{A}_p{:=}(\mathbf{v}_1,...,\mathbf{v}_p)$, be a $m\times p$ matrix $(p\leq m)$, which consists of p ordered[2] orthonormal[3] eigenvectors of the contemporaneous covariance matrix $\boldsymbol{\Sigma}{=}E((\mathbf{X}_t-\mu)(\mathbf{X}_t-\boldsymbol{\mu})^\top)$.

Therefore, the matrix of p $(p\leq m)$ principal components, $\mathbf{F}_t^{pc}{:=}(f_{1t}^{pc},...,f_{pt}^{pc})^\top$, $t{\in}\mathbb{N}$, can be constructed as follows:

$$\mathbf{F}_t^{pc}{=}\mathbf{A}_p^\top(\mathbf{X}_t-\boldsymbol{\mu})\sim\mathsf{N}(0,\boldsymbol{\Lambda}_p), \tag{1.1}$$

where $\boldsymbol{\Lambda}_p{=}diag(\lambda_1,...,\lambda_p)$ is a diagonal matrix with the diagonal elements equal to the first p eigenvalues of $\boldsymbol{\Sigma}$ arranged in a descending order. Table 1.1 summarizes the assumptions imposed to the joint distribution of PCs together with the statistical Generating Mechanism (GM).

[1] $\mathbf{A}^\top$ denotes the transpose of a matrix $\mathbf{A}$ which means every ij^{th} element of $\mathbf{A}$ is equal to the ji^{th} element of $\mathbf{A}^\top$.

[2] Ordered based on the descending order of the corresponding eigenvectors $\lambda_1\geq...\geq\lambda_p$.

[3] Mutually orthogonal and all of unit length.

Table 1.1: Normal Principal Components model

Statistical GM	$\mathbf{F}_t^{pc}{=}\mathbf{A}_p^\top(\mathbf{X}_t-\boldsymbol{\mu})+\boldsymbol{\epsilon}_t,\ t{\in}\mathbb{N},$
[1] Normality	$\mathbf{F}_t^{pc}\sim\mathsf{N}(.,.),$
[2] Linearity	$E(\mathbf{F}_t^{pc}){=}\mathbf{A}_p^\top(\mathbf{X}_t-\boldsymbol{\mu}),$
[3] Constant covariance	$Cov(\mathbf{F}_t^{pc}){=}\boldsymbol{\Lambda}_p{=}diag(\lambda_1,...,\lambda_p),$
[4] Independence	$\{\mathbf{F}_t^{pc},t{\in}\mathbb{N}\}$ is an independent process,
[5] t-invariance	$\theta{:=}(\boldsymbol{\mu},\mathbf{A}_p,\boldsymbol{\Lambda}_p)$ is not changing with t.

It is important emphasize that the above assumptions [1]-[5] provide an internally consistent and complete set of probabilistic assumptions pertaining to the observable process $\{X_{it}:t=1,...,T,i=1,...,N\}$ that comprise the statistical model underlying the PCA. In practice, one needs to test these assumptions thoroughly using effective Mis-Specification (M-S) tests to probe for any departures from these assumptions before the model is used to draw inferences. If any departures from the model assumptions are detected, one needs to respecify the original model to account for the overlooked statistical information in question. In deriving the inference procedures in the sequel, we will assume that that assumptions [1]-[5] are valid for the particular data. This is particularly crucial in the evaluation of the forecasting capacity of different statistical models as well as in the case of the empirical example in chapter 5.

For more details see Jolliffe [1986], Jackson [1993] and Stock and Watson [2002].

Chapter 2

Family Of Elliptically Contoured Distributions

The family of Elliptically Contoured Distributions is introduced by Kelker [1970], Gupta et al. [1972], Cambanis et al. [1981], and Anderson and Fang [1982]. The properties of matrix variate elliptically contoured distributions is also presented in Gupta et al. [2013].

Definition 2.1. Let matrix $\mathbf{X}$, $m \times T$, be a Random Matrix. We say $\mathbf{X}$ has a matrix-variate elliptically contoured distribution ($m.e.c.d.$), written

$$\mathbf{X}_{(m\times T)}=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1T} \\ x_{21} & x_{22} & \cdots & x_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mT} \end{pmatrix} \sim E_{m,T}(\mathbf{M}, \mathbf{\Sigma} \otimes \mathbf{\Phi}; \psi). \tag{2.1}$$

where $\otimes$ denotes the Kronecker product and $\psi(.)$ is an scalar function called *characteristic generator*, if the characteristic function is of the form

$$\phi_X(\mathbf{S})=etr(i\mathbf{S}^\top \mathbf{M})\psi(tr(\mathbf{S}^\top \mathbf{\Sigma} \mathbf{S} \mathbf{\Phi})), \text{ [1]} \tag{2.2}$$

Where $\mathbf{S}$: $m \times T$, $\mathbf{M}$: $m \times T$, $\mathbf{\Sigma} \geq 0$: $m \times m$, $\mathbf{\Phi} \geq 0$: $T \times T$ and ψ: $[0, \infty) \to \mathbb{R}$. Also, the

[1] $tr(\mathbf{S})$=$trace(\mathbf{S})$ is the sum of elements on the diagonal of the square matrix $\mathbf{S}$ and $etr(\mathbf{S})$=$exp(trace(\mathbf{S}))$.

probability density function (when exists) is of the form

$$f(\mathbf{X}){=}k_{mT}|\mathbf{\Sigma}|^{-\frac{T}{2}}|\mathbf{\Phi}|^{-\frac{m}{2}}h[tr((\mathbf{X}-\mathbf{M})^{\top}\mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{M}))\mathbf{\Phi}^{-1}] \tag{2.3}$$

where k_{mT} denotes the normalizing constant and the non-negative function $h(.)$ is called *density generator*. Note that the characteristic function and the probability density function (when exists) are functions of first two moments.

To simplify, we assume that the density function of $\mathbf{X}$ and its first two moments exist and are finite. In 2.1, $\mathbf{\Sigma}$ represents the contemporaneous covariance matrix of $\mathbf{X}$ and $\mathbf{\Phi}$ represents the temporal covariance matrix of $\mathbf{X}$.

The first and second moments are of the form

- $E(\mathbf{X}){=}\mathbf{M}$;
- $\mathbf{\Sigma}{=}\begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & \ddots & \vdots \\ \sigma_{m1} & \cdots & \sigma_{mm} \end{pmatrix}{=}E\Big((\mathbf{X}{-}E(\mathbf{X}))(\mathbf{X}{-}E(\mathbf{X}))^{\top}\Big)$;
- $\mathbf{\Phi}{=}\begin{pmatrix} \phi_{11} & \cdots & \phi_{1T} \\ \vdots & \ddots & \vdots \\ \phi_{T1} & \cdots & \phi_{TT} \end{pmatrix}{=}E\Big((\mathbf{X}{-}E(\mathbf{X}))^{\top}(\mathbf{X}{-}E(\mathbf{X})\Big)$;
- $Cov(\mathbf{X}){=}Cov(vec(\mathbf{X}^{\top})){=}c\mathbf{\Sigma}{\otimes}\mathbf{\Phi}$[2] where $c=-2\psi'(0)$ is an scalar. [3]

Also, $Cov(x_{it},x_{js}){=}-2\psi'(0)\sigma_{ij}\phi_{ts}$ where $i,j{\in}\{1,...,m\}$ and $t,s{\in}\{1,...,T\}$. Also, the i^{th} row

[2] $vec(\mathbf{X}^{\top})$ denotes the vector $(X_1,...,X_m)^{\top}$ where $X_i, i\in\{1,...,m\}$ is the i^{th} row of Matrix $\mathbf{X}$.
[3] Proof can be found in Gupta et al. [2013] pages 24-26.

(i=1, ..., m) of $\mathbf{X}$ has the variance matrix $c\sigma_{ii}\mathbf{\Phi}$ and The t^{th} column (t=1, ..., T) of $\mathbf{X}$ has the variance matrix $c\phi_{tt}\mathbf{\Sigma}$.

Theorem 2.2. *Let $\mathbf{X}$ be an $m \times T$ random matrix and $\boldsymbol{x} = vec(\mathbf{X}^\top)$. Then $\mathbf{X} \sim E_{m,T}(\mathbf{M}, \mathbf{\Sigma} \otimes \mathbf{\Phi}; \psi)$, i.e. the characteristic function of $\mathbf{X}$ is $\phi_{\mathbf{X}}(\mathbf{S}) = etr(i\mathbf{S}^\top\mathbf{M})\psi(tr(\mathbf{S}^\top\mathbf{\Sigma}\mathbf{S}\mathbf{\Phi}))$, iff $\boldsymbol{x} \sim E_{mT}(vec(\mathbf{M}^\top), \mathbf{\Sigma} \otimes \mathbf{\Phi}; \psi)$, i.e. the characteristic function of $\boldsymbol{x}$ is*

$$\phi_{\boldsymbol{x}}(\boldsymbol{s}) = etr(i\boldsymbol{s}^\top vec(\mathbf{M}^\top))\psi(\boldsymbol{s}^\top(\mathbf{\Sigma} \otimes \mathbf{\Phi})\boldsymbol{s})$$

where $\boldsymbol{s} = vec(\mathbf{S}^\top)$.

Proof. Proof can be found in Gupta and Varga [1994b]. □

The matrix form of a multivariate sampling distribution has a desirable property that allows to estimate the covariance matrix by estimating $\mathbf{\Sigma}$ and $\mathbf{\Phi}$, i.e. contemporaneous covariance and temporal covariance matrices, instead of $Cov(vec(\mathbf{X}^\top))$. In other words, to estimate the parameters we can use $\frac{m\times(m+1)}{2} + \frac{T\times(T+1)}{2}$ parameters instead of $\frac{mT\times(mT+1)}{2}$ parameters.

2.1 Gaussian Distribution

Definition 2.3. Assume we have a random matrix $\mathbf{X}$ of order $m \times T$. We say $\mathbf{X}$ has a matrix variate normal distribution, i.e.

$$\mathbf{X}_{(m\times T)} \sim \mathsf{N}_{m,T}(\mathbf{M}, \mathbf{\Sigma} \otimes \mathbf{\Phi}) \tag{2.4}$$

Where $\mathbf{M} = E(\mathbf{X}) : m \times T$, $\mathbf{\Sigma} = E((\mathbf{X}-\mathbf{M})(\mathbf{X}-\mathbf{M})^\top) \geq 0 : m \times m$, $\mathbf{\Phi} = E((\mathbf{X}-\mathbf{M})^\top(\mathbf{X}-\mathbf{M})) \geq 0 : T \times T$, $Cov(\mathbf{X}) = \mathbf{\Sigma} \otimes \mathbf{\Phi}$. The characteristic function is of the form

$$\phi_{\mathbf{X}}(\mathbf{S}) = etr(i\mathbf{S}^\top\mathbf{M} - \frac{1}{2}\mathbf{S}^\top\mathbf{\Sigma}\mathbf{S}\mathbf{\Phi}), \tag{2.5}$$

where $\mathbf{S} : m \times T$. Also, the probability density function is of the form

$$f(\mathbf{X}) = (2\pi)^{-\frac{mT}{2}}|\mathbf{\Sigma}|^{-\frac{T}{2}}|\mathbf{\Phi}|^{-\frac{m}{2}}etr(-\frac{1}{2}(\mathbf{X}-\mathbf{M})^{\top}\mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{M})\mathbf{\Phi}^{-1}). \tag{2.6}$$

Note that the characteristic function and the probability density function are functions of first two moments.

2.2 Student's t Distribution

The random matrix $\mathbf{X}$ of order $m \times T$ has a student's t distribution with degree of freedom ν, i.e.

$$\mathbf{X}_{(m\times T)} \sim \mathsf{St}_{m,T}(\mathbf{M}, \mathbf{\Sigma}\otimes\mathbf{\Phi}; \nu) \ , \tag{2.7}$$

The characteristic function is of the form

$$\phi_{\mathbf{X}}(\mathbf{S}) = etr(i\mathbf{S}^{\top}\mathbf{M} - \frac{1}{2}\mathbf{S}^{\top}\mathbf{\Sigma}\mathbf{S}\mathbf{\Phi}), \tag{2.8}$$

where $\mathbf{S} : m \times T$, $\mathbf{M} = E(\mathbf{X}) : m \times T$, $\mathbf{\Sigma} = E((\mathbf{X}-\mathbf{M})(\mathbf{X}-\mathbf{M})^{\top}) \geq 0 : m \times m$, $\mathbf{\Phi} = E((\mathbf{X}-\mathbf{M})^{\top}(\mathbf{X}-\mathbf{M})) \geq 0 : T \times T$, $Cov(\mathbf{X}) = \mathbf{\Sigma} \otimes \mathbf{\Phi}$.

The p.d.f. is given by

$$f(\mathbf{X}) = \frac{\Gamma_m[\frac{1}{2}(\nu+m+T-1)]}{\pi^{\frac{mT}{2}}\Gamma_m[\frac{1}{2}(\nu+m-1)]}|\mathbf{\Sigma}|^{-\frac{T}{2}}|\mathbf{\Phi}|^{-\frac{m}{2}} \times |\mathbf{I}_m + \mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{M})\mathbf{\Phi}^{-1}(\mathbf{X}-\mathbf{M})^{\top}|^{-\frac{\nu+m+T-1}{2}} \tag{2.9}$$

Note that the characteristic function and the probability density function are functions of first two moments.

2.3 Laplace Distribution

Definition 2.4. Assume we have a random matrix $\mathbf{X}$ of order $m \times T$. We say $\mathbf{X}$ has a matrix variate Laplace distribution, i.e.

$$\mathbf{X}_{(m\times T)} \sim \mathsf{Lap}_{m,T}(\mathbf{M}, \mathbf{\Sigma} \otimes \mathbf{\Phi}) \tag{2.10}$$

Where $\mathbf{M} = E(\mathbf{X}) : m \times T$, $\mathbf{\Sigma} = E((\mathbf{X}-\mathbf{M})(\mathbf{X}-\mathbf{M})^\top) \geq 0 : m \times m$, $\mathbf{\Phi} = E((\mathbf{X}-\mathbf{M})^\top(\mathbf{X}-\mathbf{M})) \geq 0 : T \times T$, $Cov(\mathbf{X}) = \mathbf{\Sigma} \otimes \mathbf{\Phi}$, if the characteristic function has the form of:

$$\phi_{\mathbf{X}}(\mathbf{S}) = etr(i\mathbf{S}^\top\mathbf{M})(1 + \frac{1}{2}tr(\mathbf{S}^\top\mathbf{\Sigma}\mathbf{S}\mathbf{\Phi}))^{-1}, \tag{2.11}$$

where $\mathbf{S} : m \times T$.

Note that the characteristic function is a function of first two moments.

2.4 Pearson Type II Distribution

Definition 2.5. Assume we have a random matrix $\mathbf{X}$ of order $m \times T$. We say $\mathbf{X}$ has a matrix variate Pearson Type II Distribution (matrix-variate inverted T distribution), i.e.

$$\mathbf{X}_{(m\times T)} \sim \mathsf{PII}_{m,T}(\beta, \nu) \tag{2.12}$$

if the probability density function is in the form of:

$$f(\mathbf{X})= \frac{\Gamma_m^\beta[\frac{1}{2}(\nu+T)\beta]}{\pi^{\frac{1}{2}mT\beta}\Gamma_m^\beta[\frac{1}{2}(\beta\nu)]}|\mathbf{I_m}-\mathbf{X}\mathbf{X}^\top|^{\frac{\beta(\nu-m+1)}{2}-1} \tag{2.13}$$

2.5 Pearson Type VII Distribution

Definition 2.6. Assume we have a random matrix $\mathbf{X}$ of order $m \times T$. We say $\mathbf{X}$ has a matrix variate Pearson Type VII distribution, i.e.

$$\mathbf{X}_{(m\times T)} \sim \mathsf{PVII}_{m,T}(\mathbf{M}, \mathbf{\Sigma}; \beta, \nu) \tag{2.14}$$

Where $\mathbf{M} = E(\mathbf{X}) : m \times T$, $\mathbf{\Sigma} = E((\mathbf{X}-\mathbf{M})(\mathbf{X}-\mathbf{M})^\top) \geq 0 : m \times m$, If the probability density function has the form of

$$f(\mathbf{X})= \frac{\Gamma_m^\beta}{(\pi\nu)^{\frac{1}{2}mT}\Gamma_m^\beta[\frac{1}{2}(\beta-m)]}|\mathbf{\Sigma}|^{-\frac{1}{2}}|\mathbf{I}_m+\frac{1}{\nu}\mathbf{\Sigma}^{-1}(X-\mathbf{M})\mathbf{\Phi}^{-1}(X-\mathbf{M})^\top|^{-\beta} \tag{2.15}$$

Note that the probability density function is a function of first two moments.

2.6 Exponential Power Distribution

Definition 2.7. Assume we have a random matrix $\mathbf{X}$ of order $m \times T$. We say $\mathbf{X}$ has a matrix variate Exponential Power distribution, i.e.

$$\mathbf{X}_{(m\times T)} \sim \mathsf{EP}_{m,T}(\mathbf{M}, \mathbf{\Sigma}; r, s) \tag{2.16}$$

Where $\mathbf{M} = E(\mathbf{X}) : m \times T$, $\mathbf{\Sigma} = E((\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})^\top) \geq 0 : m \times m$. If, the probability density function has the form of

$$f(\mathbf{X}) = \frac{s\Gamma_m(\frac{m}{2})}{(\pi)^{\frac{1}{2}mT}\Gamma_m(\frac{m}{2s})} r^{\frac{m}{2s}} |\mathbf{\Sigma}|^{-\frac{1}{2}} etr(-r[(X - \mathbf{M})\mathbf{\Sigma}^{-1}(X - \mathbf{M})^\top]^s) \tag{2.17}$$

Note that the probability density function is a function of first two moments.

Chapter 3

Statistical Models

3.1 Generalized Principal Component Analysis

Principal component analysis focuses primarily on the contemporaneous covariation in the data by assuming temporal independence i.e. it implicitly assumes that the temporal covariance matrix is an identity matrix ($\mathbf{\Phi}$=$\mathbf{I}_T$). In contrast, the GPCA accounts for both contemporaneous and temporal covariation in the data as well as allowing for a non-Gaussian distribution.

Zhang et al. [1985] show that a matrix variate elliptically symmetric contoured distribution can be viewed as a multivariate distribution by a simple transformation in the characteristic generator function. Let $\mathbf{X} \sim E_{m,T}(\mathbf{M}, \mathbf{\Sigma} \otimes \mathbf{\Phi}; \psi)$. The characteristic function can be written in two form:

$$\begin{aligned} \phi_{\mathbf{X}}(\mathbf{S})= & \quad etr(i\mathbf{S}^\top \mathbf{M})\psi(tr(\mathbf{S}^\top \mathbf{\Sigma} \mathbf{S} \mathbf{\Phi})), \\ \phi_{\mathbf{X}}(\mathbf{S})= & \quad etr(i\mathbf{S}^\top \mathbf{M})\psi_0(\mathbf{S}^\top \mathbf{\Sigma} \mathbf{S})) \end{aligned} \tag{3.1}$$ [1]

where $\psi_0(\mathbf{K}) = \psi(tr(\mathbf{K}\mathbf{\Phi}))$. Therefore, a matrix-variate elliptically symmetric contoured distribution (m.e.c.d.) of order $m \times T$ can be used to describe a vector-variate elliptically contoured distribution (v.e.c.d.) consists of m variables and T observations (for more details,

[1] $tr(\mathbf{S})$=$trace(\mathbf{S})$ is sum of the elements on the diagonal of a square matrix $\mathbf{S}$ and $etr(\mathbf{S})$=$exp(trace(\mathbf{S}))$.

see Siotani [1985], Gupta and Varga [1994b] and Gupta and Varga [1994c]).

Let $\mathbf{X}$ be the sampling matrix of order $m \times T$ with joint distribution:

$$\mathbf{X}=(\mathbf{X}_1, ..., \mathbf{X}_T) \sim E_{m,T}(\boldsymbol{\mu}\mathbf{e}_{T\times 1}^{\top}, \boldsymbol{\Sigma}\otimes\boldsymbol{\Phi}; \psi), \tag{3.2}$$

where $\mathbf{e}_{T\times 1}=(1, ..., 1)^{\top}$, $\boldsymbol{\mu} = (\mu_1, ..., \mu_m)^{\top}$, and $\mu_i, i \in \{1, ..., m\}$ is the expected value of i^{th} row of the sampling matrix $\mathbf{X}$. When $\psi(.)$ and $\boldsymbol{\Phi}$ are known, the MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (say $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$) are of the form (see Anderson [2003b] and Gupta et al. [2013]):

$$\hat{\boldsymbol{\mu}}=\mathbf{X}\tfrac{\boldsymbol{\Phi}^{-1}\mathbf{e}_{T\times 1}}{\mathbf{e}_{T\times 1}^{\top}\boldsymbol{\Phi}^{-1}\mathbf{e}_{T\times 1}} \tag{3.3}$$

$$\hat{\boldsymbol{\Sigma}}=\tfrac{1}{2(T-1)\psi'(0)}\mathbf{X}(\boldsymbol{\Phi}^{-1} - \tfrac{\boldsymbol{\Phi}^{-1}\mathbf{e}_{T\times 1}\mathbf{e}_{T\times 1}^{\top}\boldsymbol{\Phi}^{-1}}{\mathbf{e}_{T\times 1}^{\top}\boldsymbol{\Phi}^{-1}\mathbf{e}_{T\times 1}})\mathbf{X}^{\top} \tag{3.4}$$

where $(\boldsymbol{\Phi}^{-1} - \tfrac{\boldsymbol{\Phi}^{-1}\mathbf{e}_{T\times 1}\mathbf{e}_{T\times 1}^{\top}\boldsymbol{\Phi}^{-1}}{\mathbf{e}_{T\times 1}^{\top}\boldsymbol{\Phi}^{-1}\mathbf{e}_{T\times 1}})$ is the weighted average matrix imposed by a certain form of temporal dependence. A special case of the weighted average matrix is when $\boldsymbol{\Phi} = \mathbf{I}_T$ which the weighted average matrix would reduce to the *deviation from the mean* matrix $(\mathbf{I}_T - \mathbf{e}_{T\times 1}(\mathbf{e}_{T\times 1}^{\top}\mathbf{e}_{T\times 1})^{-1}\mathbf{e}_{T\times 1}^{\top})$.

These formulae for MLEs $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ indicate that for an operational model in (3.2) we need to know the nature of the distribution (ψ) and the temporal dependence $(\boldsymbol{\Phi})$ in the data. These problems do not arise in the case of the classical PCA because it assumes a Normal distribution and temporal independence $(\boldsymbol{\Phi}=\mathbf{I}_T)$. Note that when $\boldsymbol{\Phi}=\mathbf{I}_T$, under certain conditions, the asymptotic joint distribution of the principal components of $\hat{\boldsymbol{\Sigma}}$ is equivalent to the joint distribution of principal components of $\hat{\boldsymbol{\Sigma}}$ when we assume Normality (Gupta et al. [2013], page 144), but it is not reliable when $\boldsymbol{\Phi} \neq \mathbf{I}_T$. Put differently, under temporal independence $(\boldsymbol{\Phi}=\mathbf{I}_T)$ there is no need to worry about a distributional assumption as long as we retain the family of m.e.c.d. But, if there is any form of temporal dependence ($\mathbf{\Phi} \neq \mathbf{I}_T$), then the distributional assumption is important to secure unbiased and consistent MLEs of parameters.

The above discussion suggests the GPCA uses the extended form of covariance matrix, $\mathbf{\Sigma}\otimes\mathbf{\Phi}$, to extract GPCs. To derive p GPCs ($p < m$), we arrange the eigenvalues of $\mathbf{\Sigma}$ and $\mathbf{\Phi}$ in a descending order $(\lambda_1, ..., \lambda_m)$ and $(\gamma_1, ..., \gamma_T)$, and find their corresponding orthonormal eigenvectors $\mathbf{A}_{m\times m}{=}\,(\mathbf{v}_1, ..., \mathbf{v}_m)$ and $\mathbf{B}_{T\times T}{=}\,(\mathbf{u}_1, ..., \mathbf{u}_T)$, respectively. The first p GPCs ($p < m$) take the form:

$$\mathbf{F}{=}\mathbf{A}_p^\top(\mathbf{X} - \boldsymbol{\mu}\mathbf{e}_{T\times 1}^\top)\mathbf{B} \sim E_{p\times T}(\mathbf{0}_{p\times T}, (\mathbf{A}_p^\top\mathbf{\Sigma}\mathbf{A}_p)\otimes(\mathbf{B}^\top\mathbf{\Phi}\mathbf{B}); \psi) \Longrightarrow$$

$$\Longrightarrow \mathbf{F}{=}(\mathbf{F}_1, ..., \mathbf{F}_T){=}\mathbf{A}_p^\top(\mathbf{X}{-}\boldsymbol{\mu}\mathbf{e}_{T\times 1}^\top)\mathbf{B} \sim E_{p\times T}(\mathbf{0}_{p\times T}, \mathbf{\Lambda}_p\otimes\mathbf{\Gamma}_T; \psi) \tag{3.5}$$

where $\mathbf{F}_t{=}(f_{1t}, ..., f_{pt})^\top$, $\mathbf{A}_p{=}\,(\mathbf{v}_1, ..., \mathbf{v}_p)$, $\mathbf{\Lambda}_p{=}diag(\lambda_1, ..., \lambda_p)$ and $\mathbf{\Gamma}_T{=}diag(\gamma_1, ..., \gamma_T)$.

Why do GPCs account for the maximum variation present in the data? The simple answer is that the first element of matrix $\mathbf{F}$, f_{11} can be derived by the following optimization problem.

Let $\mathbf{v}$ be an $m \times 1$ and $\mathbf{u}$ be a $T \times 1$ vectors where $||\mathbf{v}||{=}1$ and $||\mathbf{u}||{=}1$.[2] Assume $\mathbf{v}$ and $\mathbf{u}$ are optimizing $Var(\mathbf{v}^\top\mathbf{X}\mathbf{u})$ subject to the restrictions $||\mathbf{v}||{=}1$ and $||\mathbf{u}||{=}1$. The Lagrangian function is:

$$\mathscr{L}(\mathbf{v}, \mathbf{u}, \xi_{\mathbf{v}}, \xi_{\mathbf{u}}){=}(\mathbf{v}^\top\mathbf{\Sigma}\mathbf{v} \otimes \mathbf{u}^\top\mathbf{\Phi}\mathbf{u}) - \xi_{\mathbf{v}}(\mathbf{v}^\top\mathbf{v} - 1) - \xi_{\mathbf{u}}(\mathbf{u}^\top\mathbf{u} - 1) \tag{3.6}$$

[2] $||.||$ denotes the length of a vector.

First Order Conditions (F.O.C.) $\Longrightarrow$

$$\mathbf{\Sigma v} - \xi_{\mathbf{v}}\mathbf{v}{=}0 \Longrightarrow \mathbf{\Sigma v}{=}\xi_{\mathbf{v}}\mathbf{v}, \tag{3.7}$$

and

$$\mathbf{\Phi u} - \xi_{\mathbf{u}}\mathbf{u}{=}0 \Longrightarrow \mathbf{\Phi u}{=}\xi_{\mathbf{u}}\mathbf{u}. \tag{3.8}$$

Hence, $\xi_{\mathbf{v}}$ ($\xi_{\mathbf{u}}$) is an eigenvalue for $\mathbf{\Sigma}$ ($\mathbf{\Phi}$) and $\mathbf{v}$ ($\mathbf{u}$) is the corresponding eigenvector. In fact, since $\xi_{\mathbf{v}}$ ($\xi_{\mathbf{u}}$) optimizes the objective function, it is the highest eigenvalue of $\mathbf{\Sigma}$ ($\mathbf{\Phi}$).

By repeating the same process, for kl^{th} element of $\mathbf{F}$, f_{kl}, we solve the same optimization problem by subtracting the first $kl-1$ elements of $\mathbf{F}$ with the following objective function:

$$Var(\mathbf{v}^\top[\mathbf{X} - \sum_{i=1}^{k-1}\sum_{j=1}^{l-1}\mathbf{v}_i\mathbf{v}_i^\top\mathbf{X}\mathbf{u}_j\mathbf{u}_j^\top]\mathbf{u}) \tag{3.9}$$

■

In light of (3.5), the GPCs are contemporaneously and temporally independent. By assuming Normality, Table 3.1 summarizes the assumptions imposed to the joint distribution of GPCs together with the statistical Generating Mechanism (GM).

Table 3.1: Normal Generalized Principal Components model

Statistical GM	$\mathbf{F}=\mathbf{A}_p^\top(\mathbf{X}-\boldsymbol{\mu}\mathbf{e}_{T\times 1}^\top)\mathbf{B}+\boldsymbol{\epsilon}$
[1] Normality	$\mathbf{F}\sim\mathsf{N}(.,.)$,
[2] Linearity	$E(\mathbf{F})=\mathbf{A}_p^\top(\mathbf{X}-\boldsymbol{\mu}\mathbf{e}_{T\times 1}^\top)\mathbf{B}$,
[3] Constant covariance	$Cov(\mathbf{F})=\boldsymbol{\Lambda}_p\otimes\boldsymbol{\Gamma}_T$,
[4] Independence	$\{\mathbf{F}_t, t\in\mathbb{N}\}$ is an independent process,
[5] t-invariance	$\boldsymbol{\theta}:=(\boldsymbol{\mu},\mathbf{A}_p,\boldsymbol{\Lambda}_p)$ is not changing with t.

As argued above, the probabilistic assumptions [1]-[5] in Table 3.1 comprise the statistical model underlying the GPCA. As such, these assumptions need to be tested before the modeler proceeds to use the inference procedures derived in what follows, including the optimal estimators and the procedures used to evaluate the forecasting capacity of this and related models. If any of these assumptions are found wanting, the modeler needs to respecify the original model. All the derivations that follow assume the validity of assumptions [1]-[5].

To illustrate the above, let us assume $\mathbf{X}_t=(X_{1t},...,X_{mt})^\top$, $t=1,...,T$, is a Normal, Markov and Stationary process with expected value $\boldsymbol{\mu}=E(\mathbf{X}_t)=(\mu_1,...,\mu_m)^\top$. The sampling matrix of random vector $\mathbf{X}_t$ with T observations is $\mathbf{X}=(\mathbf{X}_1,...,\mathbf{X}_T)$ where $\mathbf{X}\sim E_{m\times T}(\boldsymbol{\mu}\mathbf{e}_{T\times 1}^\top,\boldsymbol{\Sigma}\otimes\boldsymbol{\Phi};\psi)$ where $\mathbf{e}_{T\times 1}=(1,...,1)^\top$. The parameterization of a Normal, Markov ($\mathbf{M}$) and stationary ($\mathbf{S}$) process $\{\mathbf{X}_t, t\in\mathbb{N}\}$, by using sequential conditioning, implies that (see Spanos [2018]):

$$
\begin{aligned}
f(\mathbf{X}_1,...,\mathbf{X}_T;\theta)= &\ f_1(\mathbf{X}_1;\theta_1).\textstyle\prod_{t=2}^T f_t(\mathbf{X}_t|\mathbf{X}_{t-1},\mathbf{X}_{t-2},...,\mathbf{X}_1;\theta_t)\\
\overset{\mathbf{M}}{=} &\ f_1(\mathbf{X}_1;\theta_1).\textstyle\prod_{t=2}^T f_t(\mathbf{X}_t|\mathbf{X}_{t-1};\theta_t)\\
\overset{\mathbf{M\&S}}{=} &\ f(\mathbf{X}_1;\theta).\textstyle\prod_{t=2}^T f(\mathbf{X}_t|\mathbf{X}_{t-1};\theta)
\end{aligned}
$$

The above derivation enables us to derive the covariance matrix between $\mathbf{X}_t$ and $\mathbf{X}_s$. For simplicity, assume $\boldsymbol{\mu} = \mathbf{0}$. If $t < k < s$ where $t, k, s \in \{1, ..., T\}$, then:

$$\begin{aligned} Cov(\mathbf{X}_t, \mathbf{X}_s) = & \; E(\mathbf{X}_t \mathbf{X}_s) \\ = & \; E\big(E(\mathbf{X}_t \mathbf{X}_s | \mathbf{X}_k)\big) \\ = & \; E\big(E(\mathbf{X}_t | \mathbf{X}_k) E(\mathbf{X}_s | \mathbf{X}_k)\big) \\ = & \; E\Big(\big(\tfrac{Cov(\mathbf{X}_t, \mathbf{X}_k}{Var(\mathbf{X}_k)}\big)\mathbf{X}_k \big(\tfrac{Cov(\mathbf{X}_s, \mathbf{X}_k}{Var(\mathbf{X}_k)}\big)\mathbf{X}_k\Big) \\ = & \; \tfrac{Cov(\mathbf{X}_t, \mathbf{X}_k).Cov(\mathbf{X}_s, \mathbf{X}_k)}{Var(\mathbf{X}_k)} \end{aligned} \tag{3.10}$$

Using 3.10, Spanos (Spanos [1999],page 445-449) shows that:

$$Cov(\mathbf{X}_t, \mathbf{X}_s) \boldsymbol{\Sigma}.\phi(|t-s|) = \boldsymbol{\Sigma}.\phi(0).a^{|t-s|}, t, s \in \{1, ..., T\}$$

where $0 < a \leq 1$ is a real constant. This implies that:

$$Cov(X) = \Sigma \otimes \Phi = \begin{pmatrix} Cov(\mathbf{X}_1, \mathbf{X}_1) & Cov(\mathbf{X}_1, \mathbf{X}_2) & \cdots & Cov(\mathbf{X}_1, \mathbf{X}_T) \\ Cov(\mathbf{X}_2, \mathbf{X}_1) & Cov(\mathbf{X}_2, \mathbf{X}_2) & \cdots & Cov(\mathbf{X}_2, \mathbf{X}_T) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\mathbf{X}_T, \mathbf{X}_1) & Cov(\mathbf{X}_T, \mathbf{X}_2) & \cdots & Cov(\mathbf{X}_T, \mathbf{X}_T) \end{pmatrix} \tag{3.11}$$

$\Longrightarrow$

$$\boldsymbol{\Phi} = \begin{pmatrix} \phi(0) & \phi(1) & \cdots & \phi(T-1) \\ \phi(1) & \phi(0) & \cdots & \phi(T-2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(T-1) & \phi(T-2) & \cdots & \phi(0) \end{pmatrix} = \phi(0) \begin{pmatrix} 1 & a & \cdots & a^{T-1} \\ a & 1 & \cdots & a^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ a^{T-1} & a^{T-2} & \cdots & 1 \end{pmatrix} \neq \mathbf{I}_T \tag{3.12}$$

which means that the temporal covariance matrix $\mathbf{\Phi}$ is a symmetric Toeplitz matrix (see Mukherjee and Maiti [1988]). It is important to emphasize that asssuming temporal independence in the derivation of the classical PCA will result in biased estimators for $\hat{\boldsymbol{\mu}}$ and $\hat{\mathbf{\Sigma}}$; see 3.3 and 3.4.

3.2 Regression Models

Let $\mathbf{X}_t$=$(X_{1t}, ..., X_{mt})^\top$ be a set of m random variables that can be explained by p latent GPCs, $\mathbf{F}_t$=$(f_{1t}, ..., f_{pt})^\top$; for simplicity we assume $E(\mathbf{X}_t)$=$\mathbf{0}_m$. The sampling matrix distribution of $\mathbf{X}$=$(\mathbf{X}_1, ..., \mathbf{X}_T)$ and its derived GPCs $\mathbf{F}$=$(\mathbf{F}_1, ..., \mathbf{F}_T)$ (see 3.5) are:

$$\mathbf{X} \sim E_{m\times T}(\mathbf{0}_{m\times T}, \mathbf{\Sigma}\otimes\mathbf{\Phi}; \psi), \tag{3.13}$$

$$\mathbf{F} \sim E_{p\times T}(\mathbf{0}_{p\times T}, \mathbf{\Lambda}_p\otimes\mathbf{\Gamma_T}; \psi). \tag{3.14}$$

As argued above, the nature of distribution and temporal dependence $\mathbf{\Phi}$ should be specified before one can obtain unbiased MLEs of the unknown parameters. In light of that, the matrices $\mathbf{\Gamma}_T$ and $\mathbf{B}$ are assumed known. To address the time-varying form of the diagonal matrix $\mathbf{\Gamma}_T$ which represents the temporal co-variation matrix of GPCs, we have two different approaches.

Given that $\mathbf{\Phi}$ is an invertible matrix, the constant transformation presented below can adjust for the time variation in the GPCs by replacing it with the adjusted GPCs as follows:

$$\tilde{\mathbf{F}}=\mathbf{F}.\mathbf{\Gamma}_T^{-1/2} \sim E_{p\times T}(\mathbf{0}_{p\times T}, \mathbf{\Lambda}_p \otimes \mathbf{\Gamma}_T^{-1/2}\mathbf{\Gamma}_T\mathbf{\Gamma}_T^{-1/2}; \psi) \tag{3.15}$$

$$\sim E_{p\times T}(\mathbf{0}_{p\times T}, \mathbf{\Lambda}_p \otimes I_T; \psi) \tag{3.16}$$

Empirically, the factor model is useful when the data set can be explained by a few factors (for instance, in the finance literature 3 to 5 factors are usually suggested). Hence, the ratio of the summation of the largest few eigenvalues over the summation of all eigenvalues of the covariance matrix is closed enough to one (usually 95% is the threshold). This means that the rest of the eigenvalues when we have a large number of observations (T) are very small and converging to zero as t grows. Hence, for the sake of the argument we assume that there is no time variations in the $\mathbf{\Gamma}_T$ except for the first few elements on the diagonal.

Let $\mathbf{Z}_t{:=}\begin{pmatrix}\mathbf{X}_t\\ \mathbf{F}_t\end{pmatrix}$, $(m+p)\times 1$ and its sampling matrix $\mathbf{Z}{:=}(\mathbf{Z}_1,...,\mathbf{Z}_T)$, $(m+p)\times T$. The joint distribution of $\mathbf{Z}$ is:

$$\mathbf{Z}{=}\begin{pmatrix}\mathbf{X}\\ \mathbf{F}\end{pmatrix}\sim E_{(m+p)\times T}(\mathbf{0}_{(m+p)\times T},\begin{pmatrix}\mathbf{\Sigma} & \mathbf{\Xi}_{12}\\ \mathbf{\Xi}_{21} & \mathbf{\Lambda}_p\end{pmatrix}\otimes(\mathbf{\Phi}+\mathbf{\Gamma}_T);\psi), \tag{3.17}$$

where $\mathbf{\Xi}_{12}{=}Cov(\mathbf{X},\mathbf{F}){=}\mathbf{\Xi}_{21}^{\top}$. Hence, the conditional distribution $(\mathbf{X}|\mathbf{F})$ is:

$$(\mathbf{X}|\mathbf{F})\sim E_{m\times T}(\mathbf{\Xi}_{12}\mathbf{\Lambda}_p^{-1}\mathbf{F},(\mathbf{\Sigma}-\mathbf{\Xi}_{12}\mathbf{\Lambda}_p^{-1}\mathbf{\Xi}_{21})\otimes(\mathbf{\Phi}+\mathbf{\Gamma}_T);\psi_{q(\mathbf{F})}), \tag{3.18}$$

where $q(\mathbf{F}){=}tr(\mathbf{F}^{\top}\mathbf{\Lambda}_p^{-1}\mathbf{F}\mathbf{\Phi}^{-1})$.

The question that naturally arises at this stage pertains to the crucial differences between the classical PCA and the GPCA. If we assume normality and temporal independence, i.e. $\mathbf{\Phi}{=}\mathbf{I}_T$, in the above derivations, then the matrix of eigenvectors ($\mathbf{B}$) can be assumed as an identity matrix, reducing the GPCA to the classical PCA model. In this case, the conditional distribution in 3.18 can be reduced to:

$$(\mathbf{X}_t|\mathbf{F}_t)\sim \mathsf{N}_m(\mathbf{\Xi}_{12}\mathbf{\Lambda}_p^{-1}\mathbf{F}_t,(\mathbf{\Sigma}-\mathbf{\Xi}_{12}\mathbf{\Lambda}_p^{-1}\mathbf{\Xi}_{21});\ \psi_{q(\mathbf{F})}), \tag{3.19}$$

where $q(\mathbf{F}_t)=\mathbf{F}_t^\top \mathbf{\Lambda}_p^{-1}\mathbf{F}_t$. Not surprisingly, this shows that the classical PCA is an special case of the GPCA when we impose Normality and temporal independence on the data.

To shed additional light on the above derivation, let us focus on particular examples.

3.2.1 Normal, Markov and Stationary Process

Let $\mathbf{X}$, $\mathbf{F}$ and $\mathbf{Z}$ be as defined in section 3.2, but assume that $\mathbf{X}$ is a Normal, Markov and stationary vector process. This implies that the joint distribution of $\mathbf{X}=(\mathbf{X}_1, ..., \mathbf{X}_T)$ where $\mathbf{X}_t=(X_{1t}, ..., X_{mt})^\top$ can be represented by a block 'bivariate' Normal distribution:

$$(\mathbf{X}_{t-1}, \mathbf{X}_t) \sim \mathsf{N}_{m\times 2}\left(\mathbf{0}_{m\times 2},\ \mathbf{\Sigma}\otimes \begin{pmatrix} \phi(0) & \phi(1) \\ \phi(1) & \phi(0) \end{pmatrix}\right) \tag{3.20}$$

The 2×2 temporal covariance matrix in 3.20 is a reduced form of symmetric Toeplitz matrix 3.12 for Normal, Markov and Stationary process. Note that if we replace the Markov assumption with Markov of order P, then reduced form of the temporal covariance matrix would be a matrix of order $(P+1)\times(P+1)$.

This probabilistic structure gives rise to a Normal Vector Autoregressive (VAR) model, as shown in Table 3.2.

Table 3.2: Normal Vector Autoregressive (VAR) model

Statistical GM	$\mathbf{X}_t=\mathbf{B}^\top\mathbf{X}_{t-1}+\mathbf{u}_t,\ t\in\mathbb{N},$
[1] Normality	$(\mathbf{X}_t,\mathbf{X}^0_{t-1})\sim\mathsf{N}(.,.),$
	where $\mathbf{X}_t : m\times 1$ and $\mathbf{X}^0_{t-1} :=(\mathbf{X}_{t-1},...,\mathbf{X}_1),$
[2] Linearity	$E(\mathbf{X}_t\vert\sigma(\mathbf{X}^0_{t-1}))=\mathbf{B}^\top\mathbf{X}_{t-1},$
[3] Homoskedasticity	$Var(\mathbf{X}_t\vert\sigma(\mathbf{X}^0_{t-1}))=\mathbf{\Omega},$
[4] Markov	$\{\mathbf{X}_t, t\in\mathbb{N}\}$ is a Markov process,
[5] t-invariance	$\mathbf{\Theta}:=(\mathbf{B},\mathbf{\Omega})$ is not changing with t.

$$\mathbf{B}=(\mathbf{\Sigma}\phi(0))^{-1}\mathbf{\Sigma}\phi(1)=\frac{\phi(1)}{\phi(0)}I_m,$$

$$\mathbf{\Omega}=\mathbf{\Sigma}\phi(0)-(\mathbf{\Sigma}\phi(1))^\top(\mathbf{\Sigma}\phi(0))^{-1}(\mathbf{\Sigma}\phi(1))=\mathbf{\Sigma}(\phi(0)-\frac{\phi(1)^2}{\phi(0)})$$

Table 3.2 comprises the probabilistic assumptions defining the Normal VAR(1) model, and the same comments given for Tables 1.1 and 3.1 apply to this statistical model.

The joint distribution of GPCs takes the form:

$$\mathbf{F}\sim\mathsf{N}_{p\times T}(\mathbf{0}_{p\times T},(\mathbf{\Lambda}_p\otimes\mathbf{\Gamma}_T)) \tag{3.21}$$

Hence, the joint distribution of $\mathbf{Z}=\begin{pmatrix}\mathbf{X}\\ \mathbf{F}\end{pmatrix}$ presented in (3.17) can be reduced to:

$$(\mathbf{Z}_{t-1},\mathbf{Z}_t)\sim\mathsf{N}_{(p+m)\times 2}(\mathbf{0}_{(p+m)\times 2},(\mathbf{\Sigma}_0\otimes\mathbf{\Omega}_0)), \tag{3.22}$$

where $\mathbf{\Sigma}_0 = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{\Xi}_{12} \\ \mathbf{\Xi}_{21} & \mathbf{\Lambda}_p \end{pmatrix}$, $\mathbf{\Omega}_0 = \begin{pmatrix} \phi(0) & \phi(1) \\ \phi(1) & \phi(0) \end{pmatrix}$ and $\mathbf{\Xi}_{12} = Cov(\mathbf{X}, \mathbf{F}) = \mathbf{\Xi}_{21}^\top$.

Thus, the conditional distribution $(\mathbf{Z}_t|\mathbf{Z}_{t-1})$ would be of the form:

$$(\mathbf{Z}_t|\mathbf{Z}_{t-1}) \sim \mathsf{N}_{m+p}\Big(\tfrac{\phi(1)}{\phi(0)}\mathbf{Z}_{t-1}, \mathbf{\Sigma}_0 \otimes \big(\phi(0) - \tfrac{\phi(1)^2}{\phi(0)}\big)\Big) \tag{3.23}$$

As argued above, apart from a few largest eigenvalues, we can assume the rest of eigenvalues are equal to zero; i.e. for a large set of observations, $\exists t_o < T$ s.t. $\forall t > t_0$: $\gamma_t \simeq 0$, which means that they can be ignored in the bivariate distribution when $t > t_0$.

Further reduction to the form $(\mathbf{Z}_t|\mathbf{Z}_{t-1})$ gives rise to a Normal Dynamic Linear Regression (NDLR) model. Let

$$(\mathbf{X}_t, \mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}) \sim \mathsf{N}_{2(m+p)}(\mathbf{0}_{2(m+p)}, \mathbf{\Omega}_0 \otimes \mathbf{\Sigma}_0), \tag{3.24}$$

$$(\mathbf{\Omega}_0 \otimes \mathbf{\Sigma}_0) = \begin{pmatrix} \phi(0)\mathbf{\Sigma}_0 & \phi(1)\mathbf{\Sigma}_0 \\ \phi(1)\mathbf{\Sigma}_0 & \phi(0)\mathbf{\Sigma}_0 \end{pmatrix} = \left(\begin{array}{c|ccc} \phi(0)\mathbf{\Sigma} & \phi(0)\mathbf{\Xi}_{12} & \phi(1)\mathbf{\Sigma} & \phi(1)\mathbf{\Xi}_{12} \\ \hline \phi(0)\mathbf{\Xi}_{21} & \phi(0)\mathbf{\Lambda}_P & \phi(1)\mathbf{\Xi}_{21} & \phi(1)\mathbf{\Lambda}_p \\ \phi(1)\mathbf{\Sigma} & \phi(1)\mathbf{\Xi}_{12} & \phi(0)\mathbf{\Sigma} & \phi(0)\mathbf{\Xi}_{12} \\ \phi(1)\mathbf{\Xi}_{21} & \phi(1)\mathbf{\Lambda}_P & \phi(0)\mathbf{\Xi}_{21} & \phi(0)\mathbf{\Lambda}_p \end{array}\right)$$
$$= \left(\begin{array}{c|c} \mathsf{V}_{11} & \mathsf{V}_{12} \\ \hline \mathsf{V}_{21} & \mathsf{V}_{22} \end{array}\right)$$

The joint distribution in 3.24, can be decomposed as follow:

$$f(\mathbf{X}_t, \mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}; \Theta) = \quad f(\mathbf{X}_t|\mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}; \Theta_1) \cdot f(\mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}; \Theta_2)$$

So, the joint distribution 3.24 can be viewed as a product of marginal and conditional distributions presented below:

$$(\mathbf{X}_t|\mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}) \sim \mathsf{N}_m(\mathbf{V}_{11}^{-1}\mathbf{V}_{12}(\mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1})^\top,\ \mathbf{V}) \\ \text{where } \mathbf{V} = \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21} \tag{3.25}$$

and,

$$(\mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}) \sim \mathsf{N}_{m+2p}(\mathbf{0}_{m+2p}, \mathbf{V}_{22}) \tag{3.26}$$

The decomposition of bivariate normal distribution in 3.24 to the conditional distribution 3.25 and marginal distribution 3.26 induces a form of re-parameterization as follows:

$$\begin{aligned} \theta &:= \{\mathbf{V}_{11}, \mathbf{V}_{12}, \mathbf{V}_{22}\} \\ \theta_1 &:= \{\mathbf{V}_{22}\} \\ \theta_2 &:= \{\mathbf{B}, \mathbf{V}\} \text{ where } \mathbf{B}{=}\mathbf{V}_{11}^{-1}\mathbf{V}_{12} \text{ and } \mathbf{V}{=}\mathbf{V}_{11}{-}\mathsf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21} \end{aligned}$$

This re-parameterization indicates that the parameter sets θ_1 and θ_2 are variation free[3]; so we have a weak exogeneity with respect to Θ_1 and the marginal distribution can be ignored for the modeling purpose and instead we can model in term of conditional distribution (see Spanos [1999] pages 366-368).

3.2.2 Student's t, Markov and Stationary Process

Again, let $\mathbf{X}$, $\mathbf{F}$ and $\mathbf{Z}$ be as defined in section 3.2, but assume that $\mathbf{X}$ is a Student's t, Markov and stationary process. This implies that the joint distribution of $\mathbf{X}{=}(\mathbf{X}_1, ..., \mathbf{X}_T)$

[3] Θ_1 and Θ_2 are variation free if for all values of Θ_1 the range of possible values of Θ_2 doesn't change.

where $\mathbf{X}_t = (X_{1t}, ..., X_{mt})^\top$ can be represented by:

$$(\mathbf{X}_T, \mathbf{X}^0_{T-1}) \sim \mathsf{St}_{m\times T}\left(\mathbf{0}_{m\times T},\ \mathbf{\Sigma}\otimes\mathbf{\Phi} = \begin{pmatrix} \phi_{11} & \mathbf{\Phi}_{12} \\ \mathbf{\Phi}_{21} & \mathbf{\Phi}_{22} \end{pmatrix};\nu\right) \tag{3.27}$$

where ν is the degree of freedom and $\mathbf{X}^0_{t-1} = (\mathbf{X}_{t-1}, ..., \mathbf{X}_1)$.

Table 3.3 presents the probabilistic structure of a Student's t Vector Autoregressive (StVAR) model.

Table 3.3: Student's t Vector Autoregressive (StVAR) model

Statistical GM	$\mathbf{X}_t = \mathbf{B}^\top\mathbf{X}_{t-1} + \mathbf{u}_t,\ t\in\mathbb{N}$,
[1] Student's t	$(\mathbf{X}_t, \mathbf{X}^0_{t-1}) \sim \mathsf{St}(.,.;\nu)$, where $\mathbf{X}_t : m\times 1$ and $\mathbf{X}^0_{t-1} := (\mathbf{X}_{t-1}, ..., \mathbf{X}_1)$,
[2] Linearity	$E(\mathbf{X}_t \vert \sigma(\mathbf{X}^0_{t-1})) = \mathbf{B}^\top\mathbf{X}_{t-1}$,
[3] Heteroskedasticity	$Var(\mathbf{X}_t \vert \sigma(\mathbf{X}^0_{t-1})) = \frac{\nu\phi_{11.2}}{\nu+m-2} q(\mathbf{X}^0_{t-1})$, $q(\mathbf{X}^0_{t-1}) := \mathbf{\Sigma}[\mathbf{I}_m + \mathbf{\Sigma}^{-1}\mathbf{X}^0_{t-1}\mathbf{\Phi}^{-1}_{22}\mathbf{X}^0_{t-1}{}^\top]$ $\phi_{11.2} := \phi_{11} - \mathbf{\Phi}_{12}\mathbf{\Phi}^{-1}_{22}\mathbf{\Phi}_{21}$
[4] Markov	$\{\mathbf{X}_t, t\in\mathbb{N}\}$ is a Markov process,
[5] t-invariance	$\mathbf{\Theta} := (\mathbf{B}, \mathbf{\Sigma}, \mathbf{\Phi})$ is not changing with t.

Table 3.3 specifies the main statistical model for GPCA based on the matrix Student's t distribution. The validity of the probabilistic assumptions [1]-[5] is assumed in the derivations that follow. In practice, this statistical model is adopted only when these assumptions are valid for the particular data; see chapter 5.

Hence, the joint distribution of GPCs takes the form:

$$\mathbf{F} \sim \mathsf{St}_{p\times T}(\mathbf{0}_{p\times T}, (\mathbf{\Lambda}_p\otimes\mathbf{\Gamma}_T);\nu) \tag{3.28}$$

Hence, the joint distribution of $\mathbf{Z}{=}\begin{pmatrix}\mathbf{X}\\ \mathbf{F}\end{pmatrix}$ presented in (3.17) can be reduced to:

$$\begin{pmatrix}\mathbf{Z}_t\\ \mathbf{Z}_{t-1}\end{pmatrix} \sim \mathsf{St}_{(p+m)\times 2}(\mathbf{0}_{(p+m)\times 2}, (\mathbf{\Omega}_0 \otimes \mathbf{\Sigma}_0)\,;\nu) \tag{3.29}$$

where $\mathbf{\Sigma}_0{=}\begin{pmatrix}\mathbf{\Sigma} & \mathbf{\Xi}_{12}\\ \mathbf{\Xi}_{21} & \mathbf{\Lambda}_p\end{pmatrix}$, $\mathbf{\Omega}_0{=}\begin{pmatrix}\phi(0) & \phi(1)\\ \phi(1) & \phi(0)\end{pmatrix}$ and $\mathbf{\Xi}_{12}{=}Cov(\mathbf{X},\mathbf{F}){=}\mathbf{\Xi}_{21}^{\top}$.

Thus, the conditional distribution $(\mathbf{Z}_t|\mathbf{Z}_{t-1})$ would be of the form:

$$\begin{gathered}(\mathbf{Z}_t|\mathbf{Z}_{t-1}) \sim \mathsf{St}_{m+p}\Big(\tfrac{\phi(1)}{\phi(0)}\mathbf{Z}_{t-1}, q(\mathbf{Z}_{t-1}).\big((\phi(0)-\tfrac{\phi(1)^2}{\phi(0)}).\mathbf{\Sigma}_0\big);\nu+m\Big)\\ q(\mathbf{Z}_{t-1}){:=}[1+\tfrac{1}{\nu}\mathbf{Z}_{t-1}^{\top}(\phi(0)\mathbf{\Sigma}_0)^{-1}\mathbf{Z}_{t-1}]\end{gathered} \tag{3.30}$$

Further reduction to the form $(\mathbf{Z}_t|\mathbf{Z}_{t-1})$ gives rise to a Student's t Dynamic Linear Regression (NDLR) model. Let

$$(\mathbf{X}_t, \mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}) \sim \mathsf{St}_{2(m+p)}(\mathbf{0}_{2(m+p)}, \mathbf{\Omega}_0 \otimes \mathbf{\Sigma}_0; \nu) \tag{3.31}$$

$$(\mathbf{\Omega}_0 \otimes \mathbf{\Sigma}_0){=}\begin{pmatrix}\phi(0)\mathbf{\Sigma}_0 & \phi(1)\mathbf{\Sigma}_0\\ \phi(1)\mathbf{\Sigma}_0 & \phi(0)\mathbf{\Sigma}_{\mathbf{0}}\end{pmatrix} = \left(\begin{array}{c|ccc}\phi(0)\mathbf{\Sigma} & \phi(0)\mathbf{\Xi}_{\mathbf{12}} & \phi(1)\mathbf{\Sigma} & \phi(1)\mathbf{\Xi}_{12}\\ \hline \phi(0)\mathbf{\Xi}_{21} & \phi(0)\mathbf{\Lambda}_P & \phi(1)\mathbf{\Xi}_{21} & \phi(1)\mathbf{\Lambda}_p\\ \phi(1)\mathbf{\Sigma} & \phi(1)\mathbf{\Xi}_{12} & \phi(0)\mathbf{\Sigma} & \phi(0)\mathbf{\Xi}_{12}\\ \phi(1)\mathbf{\Xi}_{21} & \phi(1)\mathbf{\Lambda}_P & \phi(0)\mathbf{\Xi}_{21} & \phi(0)\mathbf{\Lambda}_p\end{array}\right)$$

$$= \left(\begin{array}{c|c}\mathsf{V}_{11} & \mathsf{V}_{12}\\ \hline \mathsf{V}_{21} & \mathsf{V}_{22}\end{array}\right)$$

The joint distribution in 3.31, can be decomposed as follows:

$$f(\mathbf{X}_t, \mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}; \theta) = \quad f(\mathbf{X}_t | \mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}; \theta_1) \cdot f(\mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}; \theta_2).$$

So, the joint distribution 3.31 can be viewed as a product of marginal and conditional distributions presented below:

$$\begin{gathered}(\mathbf{X}_t | \mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}) \sim \mathsf{St}_m\big(\mathsf{V}_{11}^{-1}\mathsf{V}_{12}(\mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1})^\top, \mathsf{V}.q(\mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}); \nu + m + 2p\big) \\ \text{where } \mathsf{V}{=}\mathsf{V}_{11}{-}\mathsf{V}_{12}\mathsf{V}_{22}^{-1}\mathsf{V}_{21}, \\ q(\mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}){:=}[1 + \tfrac{1}{\nu}(\mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1})^\top \mathsf{V}_{22}^{-1}(\mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1})]\end{gathered} \tag{3.32}$$

and,

$$(\mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}) \sim \mathsf{St}_{m+2p}(\mathbf{0}_{m+2p}, \mathsf{V}_{22}; \nu) \tag{3.33}$$

The decomposition of bivariate student's t distribution in 3.31 to the conditional distribution 3.32 and marginal distribution 3.33 induces a form of re-parameterization as follows:

$$\begin{aligned}\theta{:=} &\quad \{\mathbf{V}_{11}, \mathbf{V}_{12}, \mathbf{V}_{22}\} \\ \theta_1{:=} &\quad \{\mathbf{V}_{22}\} \\ \theta_2{:=} &\quad \{\mathbf{B}, \mathbf{V}, \mathbf{V}_{22}\} \text{ where } \mathbf{B}{=}\mathbf{V}_{11}^{-1}\mathbf{V}_{12} \text{ and } \mathbf{V}{=}\mathbf{V}_{11}{-}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\end{aligned}$$

This re-parameterization indicates that the parameter sets θ_1 and θ_2 are not variation free because $\mathbf{V}_{22}$ appears in all parameters sets which can directly impose restrictions; so we do not have a weak exogeneity with respect to θ_1 and the marginal distribution cannot be ignored for the modeling purpose and instead we can model in term of conditional distribution.

Chapter 4

Monte Carlo Simulation

4.1 The Normal VAR Simulation

The reason that we choose a Normal VAR for the Monte Carlo simulation is that the Random Walk as a benchmark model has the best chance to survive against factor models when we have a Normal, Markov and Stationary process. The Normal VAR model presented in Table 3.2 can be re-parameterized as a Normal Dynamic Linear Regression (NDLR) model by introducing a different partitions on the bivariate joint distribution presented in 3.20. Let $\mathbf{X}_t=(X_{1t}, ..., X_{mt})^\top$ and $\boldsymbol{\mu}=E(\mathbf{X}_t)$, so,

$$\begin{pmatrix} \mathbf{X}_t \\ \mathbf{X}_{t-1} \end{pmatrix} \sim \mathsf{N}_{2m}\left(\begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \end{pmatrix}, \begin{pmatrix} \phi(0) & \phi(1) \\ \phi(1) & \phi(0) \end{pmatrix} \otimes \boldsymbol{\Sigma}\right) \tag{4.1}$$

$$\sim \mathsf{N}_{2m}\left(\begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \end{pmatrix}, \begin{pmatrix} \phi(0)\boldsymbol{\Sigma} & \phi(1)\boldsymbol{\Sigma} \\ \phi(1)\boldsymbol{\Sigma} & \phi(0)\boldsymbol{\Sigma} \end{pmatrix}\right) \tag{4.2}$$

Let define $\mathbf{X}_t^j=(X_{1t}, ..., X_{(j-1)t}, X_{(j+1)t}, ..., X_{mt})$, $\mathbf{W}_t^j=\begin{pmatrix} \mathbf{X}_t^j \\ \mathbf{X}_{t-1} \end{pmatrix}$ and $E(\mathbf{W}_t^j)=\boldsymbol{\mu}_{W_t^j}$ where $j=1, ..., m$.

For simplicity, assume j=1; the joint distribution in 4.1 can be written as follows:

$$\begin{pmatrix} X_{1t} \\ \mathbf{W}_t^1 \end{pmatrix} \sim \mathsf{N}_{2m} \left(\begin{pmatrix} \mu_1 \\ \boldsymbol{\mu}_{W_t^1} \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right) \tag{4.3}$$

where $\mu_1 = E(X_{1t})$, σ_{11}=$Var(X_{1t})$, $\boldsymbol{\Sigma}_{12}$=$Cov(X_{1t}, \mathbf{W}_t^1)$=$\boldsymbol{\Sigma}_{21}^\top$, and $\boldsymbol{\Sigma}_{22}$=$Cov(\mathbf{W}_t^1)$.

As we have explained in section 3.2.1, the Normal, Markov and Stationary process can be modeled only in term of conditional distribution due to the weak exogeneity. The parameterization of this conditional distribution can be summarized as follows:

$$(X_{1t}|\mathbf{W}_t^1) \sim \mathsf{N}(\alpha + \boldsymbol{\beta}\mathbf{W}_t^1, \sigma_0) \tag{4.4}$$

$$\alpha{=}\mu_1{-}\boldsymbol{\beta}\boldsymbol{\mu}_{W_t^1}, \boldsymbol{\beta}{=}\tfrac{\boldsymbol{\Sigma}_{12}}{\sigma_{11}},\ \sigma_0{=}\sigma_{11}{-}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.$$

4.1.1 Simulation Design

Let $\mathbf{X}_t$=$(X_{1t}, ..., X_{15t})$ where $t \in \{1, ..., 250\}$ and:

$$\begin{pmatrix} \mathbf{X}_t \\ \mathbf{X}_{t-1} \end{pmatrix} \sim \mathsf{N}_{30} \left(\begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \end{pmatrix}, \begin{pmatrix} \phi(0) & \phi(1) \\ \phi(1) & \phi(0) \end{pmatrix} \otimes \boldsymbol{\Sigma} \right)$$

where $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\phi(0)$ and $\phi(1)$ are as follows .

$$\boldsymbol{\Sigma} = \begin{pmatrix}
0.072 & 0.030 & 0.018 & 0.031 & 0.036 & 0.064 & -0.044 & -0.024 & 0.008 & 0.031 & -0.022 & 0.083 & 0.084 & 0.036 & -0.016 \\
0.030 & 0.017 & 0.011 & 0.014 & 0.017 & 0.028 & -0.015 & -0.005 & 0.006 & 0.015 & -0.005 & 0.037 & 0.038 & 0.018 & -0.002 \\
0.018 & 0.011 & 0.033 & 0.020 & 0.023 & 0.028 & 0.017 & 0.022 & 0.031 & 0.034 & 0.023 & 0.037 & 0.034 & 0.023 & 0.025 \\
0.031 & 0.014 & 0.020 & 0.025 & 0.021 & 0.033 & -0.008 & 0.001 & 0.016 & 0.026 & 0.003 & 0.044 & 0.045 & 0.022 & 0.005 \\
0.036 & 0.017 & 0.023 & 0.021 & 0.027 & 0.040 & -0.006 & 0.003 & 0.019 & 0.028 & 0.004 & 0.050 & 0.049 & 0.027 & 0.007 \\
0.064 & 0.028 & 0.028 & 0.033 & 0.040 & 0.070 & -0.030 & -0.010 & 0.020 & 0.039 & -0.008 & 0.088 & 0.087 & 0.043 & -0.003 \\
-0.044 & -0.015 & 0.017 & -0.008 & -0.006 & -0.030 & 0.072 & 0.054 & 0.029 & 0.005 & 0.052 & -0.043 & -0.049 & -0.007 & 0.047 \\
-0.024 & -0.005 & 0.022 & 0.001 & 0.003 & -0.010 & 0.054 & 0.045 & 0.031 & 0.015 & 0.044 & -0.016 & -0.021 & 0.004 & 0.041 \\
0.008 & 0.006 & 0.031 & 0.016 & 0.019 & 0.020 & 0.029 & 0.031 & 0.034 & 0.031 & 0.031 & 0.025 & 0.020 & 0.019 & 0.031 \\
0.031 & 0.015 & 0.034 & 0.026 & 0.028 & 0.039 & 0.005 & 0.015 & 0.031 & 0.038 & 0.016 & 0.052 & 0.049 & 0.028 & 0.019 \\
-0.022 & -0.005 & 0.023 & 0.003 & 0.004 & -0.008 & 0.052 & 0.044 & 0.031 & 0.016 & 0.044 & -0.013 & -0.018 & 0.005 & 0.041 \\
0.083 & 0.037 & 0.037 & 0.044 & 0.050 & 0.088 & -0.043 & -0.016 & 0.025 & 0.052 & -0.013 & 0.117 & 0.116 & 0.053 & -0.005 \\
0.084 & 0.038 & 0.034 & 0.045 & 0.049 & 0.087 & -0.049 & -0.021 & 0.020 & 0.049 & -0.018 & 0.116 & 0.119 & 0.052 & -0.010 \\
0.036 & 0.018 & 0.023 & 0.022 & 0.027 & 0.043 & -0.007 & 0.004 & 0.019 & 0.028 & 0.005 & 0.053 & 0.052 & 0.031 & 0.008 \\
-0.016 & -0.002 & 0.025 & 0.005 & 0.007 & -0.003 & 0.047 & 0.041 & 0.031 & 0.019 & 0.041 & -0.005 & -0.010 & 0.008 & 0.039
\end{pmatrix}$$

$$\boldsymbol{\Phi} = \begin{pmatrix} \phi(0) & \phi(1) \\ \phi(1) & \phi(0) \end{pmatrix} = \begin{pmatrix} \phi(0) & \phi(0).a \\ \phi(0).a & \phi(0) \end{pmatrix} = \begin{pmatrix} 1.8 & 1.8\times 0.8 \\ 1.8\times 0.8 & 1.8 \end{pmatrix}$$

$$\boldsymbol{\mu} = (2.5, 1.9, 0.8, 0.5, 1.3, 0.9, 3.4, 2.3, 0.3, 0.08, 4.5, 3.7, 1.4, 2.9, 0.001)$$

The contemporaneous covariance matrix, $\boldsymbol{\Sigma}$, is based on the contemporaneous covariance matrix of the log exchange rates of 15 OECD countries based on US dollar. Also, reduced form of temporal covariance matrix $\boldsymbol{\Phi}$ is an example of the Normal, Markov and Stationary process explained in 3.12. In addition, the covariance matrix $\boldsymbol{\Phi} \otimes \boldsymbol{\Sigma} > 0$ is a positive definite matrix.

The theoretical coefficients, the t-statistics (brackets) and corresponding p-values (square brackets) associated with the *difference* between the actual (θ^*) and estimated ($\hat{\theta}$) coefficients[1] are:

$$
\begin{aligned}
X_{1t}= & \underset{\substack{(-1.484)\\ [0.138]}}{0.564} + \underset{\substack{(-0.055)\\ [0.956]}}{0.659} X_{2t} - \underset{\substack{(-0.563)\\ [0.574]}}{1.767} X_{3t} - \underset{\substack{(0.253)\\ [0.800]}}{0.059} X_{4t} + \underset{\substack{(0.172)\\ [0.863]}}{0.764} X_{5t} + \underset{\substack{(0.114)\\ [0.910]}}{0.082} X_{6t} \\
& - \underset{\substack{(-0.335)\\ [0.738]}}{0.257} X_{7t} - \underset{\substack{(0.173)\\ [0.863]}}{3.408} X_{8t} + \underset{\substack{(0.432)\\ [0.666]}}{0.379} X_{9t} + \underset{\substack{(0.854)\\ [0.393]}}{1.171} X_{10t} + \underset{\substack{(-0.159)\\ [0.874]}}{1.64} X_{11t} \\
& - \underset{\substack{(-0.155)\\ [0.877]}}{0.148} X_{12t} - \underset{\substack{(-1.060)\\ [0.289]}}{0.166} X_{13t} + \underset{\substack{(0.484)\\ [0.629]}}{0.245} X_{14t} + \underset{\substack{(-0.125)\\ [0.900]}}{1.851} X_{15t} \\
& + \underset{\substack{(0.184)\\ [0.854]}}{0.800} X_{1t-1} - \underset{\substack{(-0.376)\\ [0.707]}}{0.527} X_{2t-1} + \underset{\substack{(-0.881)\\ [0.378]}}{1.414} X_{3t-1} + \underset{\substack{(0.391)\\ [0.696]}}{0.048} X_{4t-1} - \underset{\substack{(-0.059)\\ [0.953]}}{0.611} X_{5t-1} \\
& - \underset{\substack{(-0.064)\\ [0.949]}}{0.066} X_{6t-1} + \underset{\substack{(1.295)\\ [0.196]}}{0.206} X_{7t-1} + \underset{\substack{(-0.465)\\ [0.642]}}{2.726} X_{8t-1} - \underset{\substack{(-0.803)\\ [0.421]}}{0.303} X_{9t-1} - \underset{\substack{(-0.424)\\ [0.672]}}{0.937} X_{10t-1} \\
& - \underset{\substack{(0.998)\\ [0.318]}}{1.312} X_{11t-1} + \underset{\substack{(1.532)\\ [0.126]}}{0.119} X_{12t-1} + \underset{\substack{(-0.532)\\ [0.595]}}{0.132} X_{13t-1} - \underset{\substack{(-0.710)\\ [0.478]}}{0.196} X_{14t-1} - \underset{\substack{(-0.562)\\ [0.574]}}{1.481} X_{15t-1} \\
& + 0.0513\epsilon_{1t}
\end{aligned}
$$

where $\sigma_0=\sqrt{0.002634}=0.0513$ and $\epsilon_{1t} \sim \mathsf{N}(0,1)$. Also, $\mathsf{R}^2=1-\frac{\sigma_0^2}{\phi(0)\sigma_{11}}=1-\frac{0.002634}{1.8\times0.072}=0.98$.

The histogram comparison of empirical and theoretical distributions of these coefficients are presented in the appendix A.2.

4.1.2 Forecasting

In this section, we use the Monte Carlo simulation presented above to generate a set of 15 random variables, $\mathbf{X}_t=(X_{1t},...,X_{15t})^\top$, with 250 observations for each variable i.e. $t=1,...,250$. To compare the predictive capacity of GPCA vs PCA, we extract a set of three factors (ei-

[1] Hypothesis testing $H_0: \theta^* - \hat{\theta} = 0$ vs. $H_1: \theta^* - \hat{\theta} \neq 0$

ther GPCs or PCs) from the panel of 15 variables, $\mathbf{X}$=($\mathbf{X}_1, ..., \mathbf{X}_{250}$). The eigenvalue ratio test indicates that three factors account for 97% of the variation in the data. Figure 4.1 and Figure 4.2 are presenting the principal components (PCs) and the generalized principal components (GPCs) extracted from the full sample, respectively.

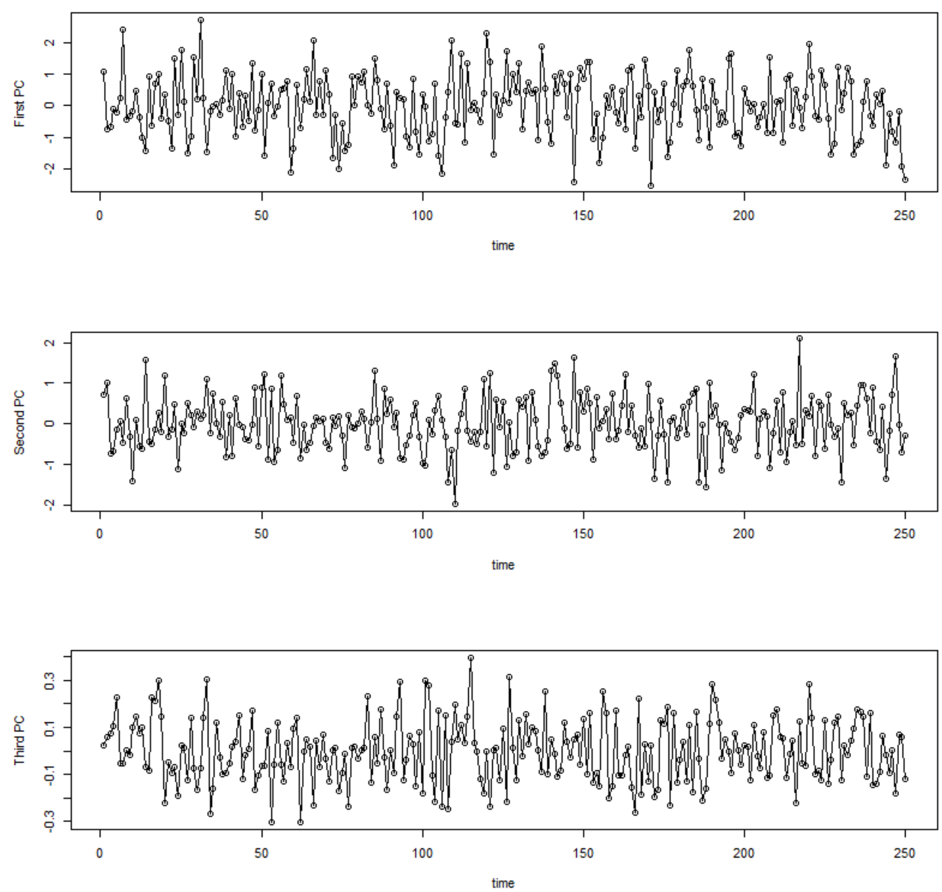


Figure 4.1: Principal Components

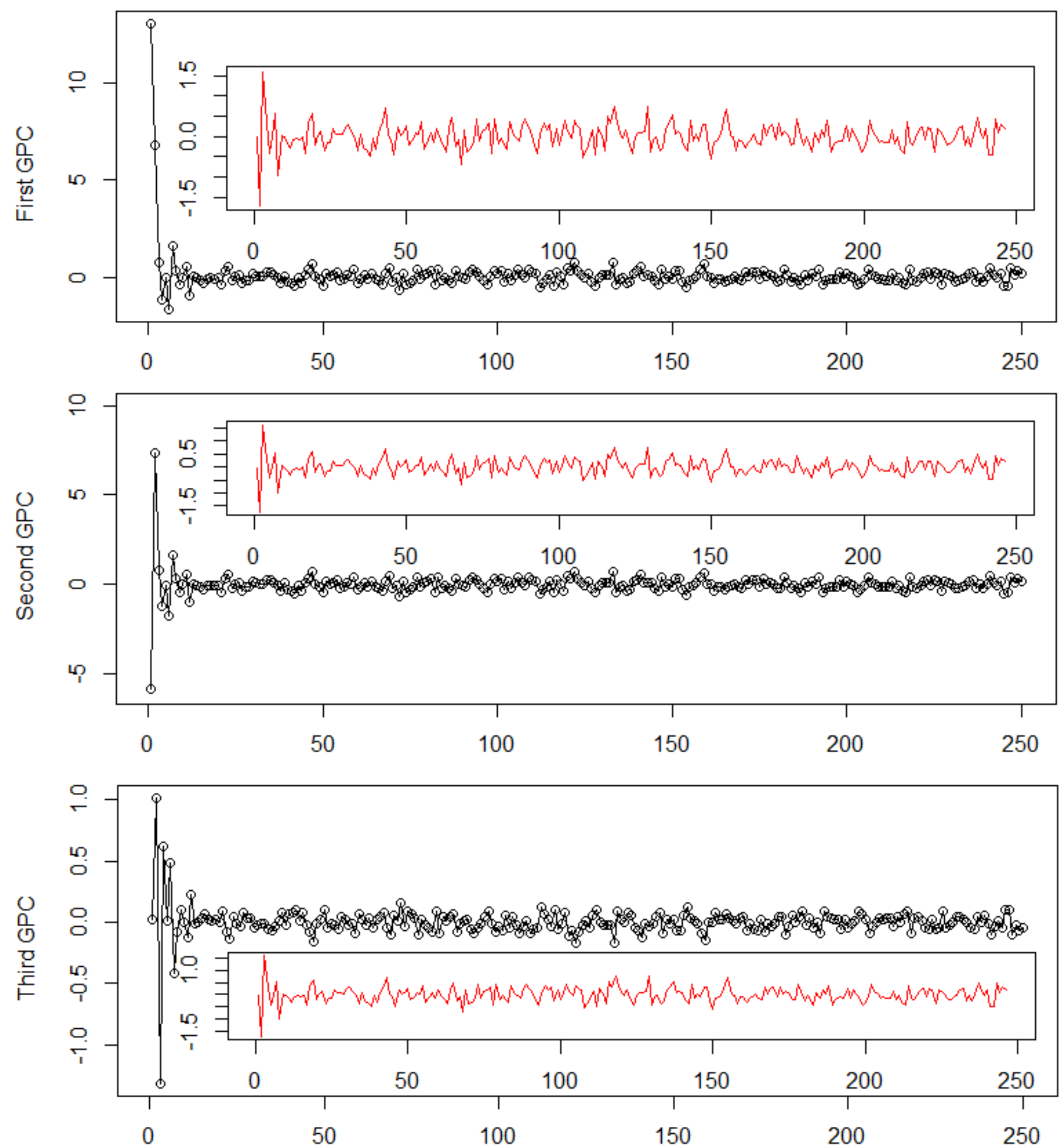


Figure 4.2: Generalized Principal Components

As illustrated in Figure 4.2, GPCs are constructed by decomposing the extended covariance matrix $\mathbf{\Sigma} \otimes \mathbf{\Phi}$; therefore, the most of temporal covariation are captured by a few first points of time.

Our presumption is that the variation of the variables in $\mathbf{X}_t \sim \mathsf{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ can be explained by factor models. Algebraically,

$$X_{it}=\mathbf{F}_{it}+u_{it},\ u_{it}\sim \mathsf{N}(0,\sigma_u^2),\ t=1,...,250,\ i=1,...,15, \tag{4.5}$$

$$\mathbf{F}_{it}=\delta_i f_{1t}+\delta_i f_{2t}+\delta_i f_{3t}, \tag{4.6}$$

where f_{jt} and δ_j, $j=1,2,3$, are factors and factor loadings, respectively. Also, in order to extract factors, we centralize $\mathbf{X}_t$ according to the in-sample data. In addition, we have:

$$0=E(X_{it+h}-X_{it})\quad =E(X_{it+h})-E(X_{it})=E(X_{it+h})-E_{\mathbf{F}_{it}}(E(X_{it}|\mathbf{F}_{it}))=E(X_{it+h})-\mathbf{F}_{it},$$

which implies that:

$$E(X_{it+h})=\mathbf{F}_{it}\Longrightarrow E_{X_{it}}(E(X_{it+h}|X_{it}))=\mathbf{F}_{it}\Longrightarrow E_{X_{it}}(E(X_{it+h}-X_{it}|X_{it}))=\mathbf{F}_{it}-X_{it}\Longrightarrow$$

$$E(X_{it+h}-X_{it})=\mathbf{F}_{it}-X_{it}, \tag{4.7}$$

where h is the forecast horizon. Therefore, we use $\mathbf{F}_{it}-X_{it}$ as a central tendency to forecast $X_{it+h}-X_{it}$:

$$X_{it+h}-X_{it}=\boldsymbol{\alpha}_i+\beta(\mathbf{F}_{it}-X_{it})+\varepsilon_{it+h} \tag{4.8}$$

where $\boldsymbol{\alpha}_i$ is a fixed effect of the i-th variable.

We begin with the first 150 observations to extract factors $\widehat{\mathbf{F}}_{it}$ and estimate the coefficients, i.e.

$$X_{i150}-X_{i(150-h)}=\boldsymbol{\alpha}_i+\beta(\widehat{\mathbf{F}}_{i(150-h)}-X_{i(150-h)})+\varepsilon_{i150}$$

Then we use the estimated coefficients $\hat{\boldsymbol{\alpha}}_i$ and $\hat{\beta}$ to predict the value of $X_{i(150+h)}-X_{i150}$ as follows:

$$X_{i(150+h)} - X_{i150} = \hat{\boldsymbol{\alpha}}_i + \hat{\beta}(\widehat{\mathbf{F}}_{i150} - X_{i150})$$

We will follow the same recursive procedure by adding another observation to the end of the in-sample data set to generate forecasts.

Figure 4.3 illustrates the above procedure for horizon h=4:

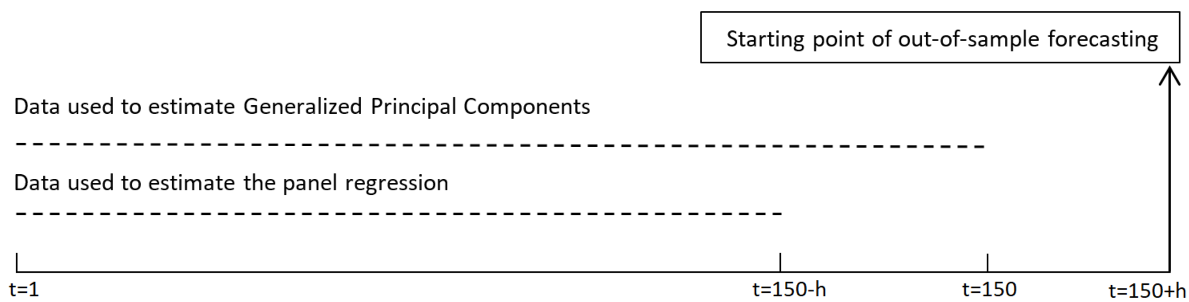


Figure 4.3: Forecasting Procedure for horizon h

The forecast evaluation is based on comparing the root mean squared prediction errors (RMSPE). We compare RMSPE of the factor model (either GPCA or PCA) with random walk model to examine the predictive capacity of the factor model using Theil's U-statistic (Theil [1971]). The U-statistic is defined as follows:

$$U-\text{statistics}=\frac{RMSPE_{\text{factor model}}}{RMSPE_{\text{random walk}}}$$

The U-statistic less than one means that the factor model has a better performance than the random walk model. Also, we use the t-test proposed by Clark and West [2006] to test the hypothesis that H_0: U=1 vs H_1: $U < 1$, based on a .025 significance level with rejection region defined by $(\tau(\mathbf{X}) > 1.96)$.

Table 4.1 presents the median U-statistics in each forecast horizon for both models (GPCA and PCA) and the number of individual variables (out of 15) with U-statistic less than one and Clark and West t-test greater than 1.960. The detailed table of individual variables for both models are in the appendix (see Table A.1 and Table A.2).

Table 4.1: Forecast evaluation: GPCA vs PCA

		Horizon h			
Model	Measurement	h=1	h=4	h=8	h=12
$\widehat{\mathbf{GPC}}_{it} - X_{it}$	Median U-statistic	0.697	0.685	0.713	0.735
	(#$U < 1$ out of 15)	(15)	(15)	(15)	(15)
	[#$t > 1.960$ out of 15]	[15]	[7]	[5]	[1]
$\widehat{\mathbf{PC}}_{it} - X_{it}$	Median U-statistic	0.995	0.998	0.997	0.997
	(#$U < 1$ out of 15)	(10)	(10)	(9)	(8)
	[#$t > 1.960$ out of 15]	[2]	[0]	[0]	[0]

Note: $\widehat{\mathbf{GPC}}_{it} - s_{it}$ and $\widehat{\mathbf{PC}}_{it} - s_{it}$ represent deviations from factors produced by the GPCA and the classical PCA, respectively. The number of variables (out of 15) with U-statistic (Theil [1971]) less than one and the number of variables (out of 15) with Clark-West t-statistic (Clark and West [2006]) more than 1.960 are reported in parenthesis and brackets, respectively.

To illustrate the predictive capacity of the GPCA method and compare it to that of the PCA method, Figures 4.4 to 4.7 compare the actual observation $(X_{it+h} - X_i)$ with predicted values using both GPCA method $(\hat{\boldsymbol{\alpha}}_i + \hat{\beta}(\widehat{\mathbf{GPC}}_{it} - X_{it}))$ and PCA method $(\hat{\boldsymbol{\alpha}}_i + \hat{\beta}(\widehat{\mathbf{PC}}_{it} - X_{it}))$ for horizon h=1 and i=1, ..., 4.[2]

[2]Plots of the all variables in all horizons are presented in the appendix A.3

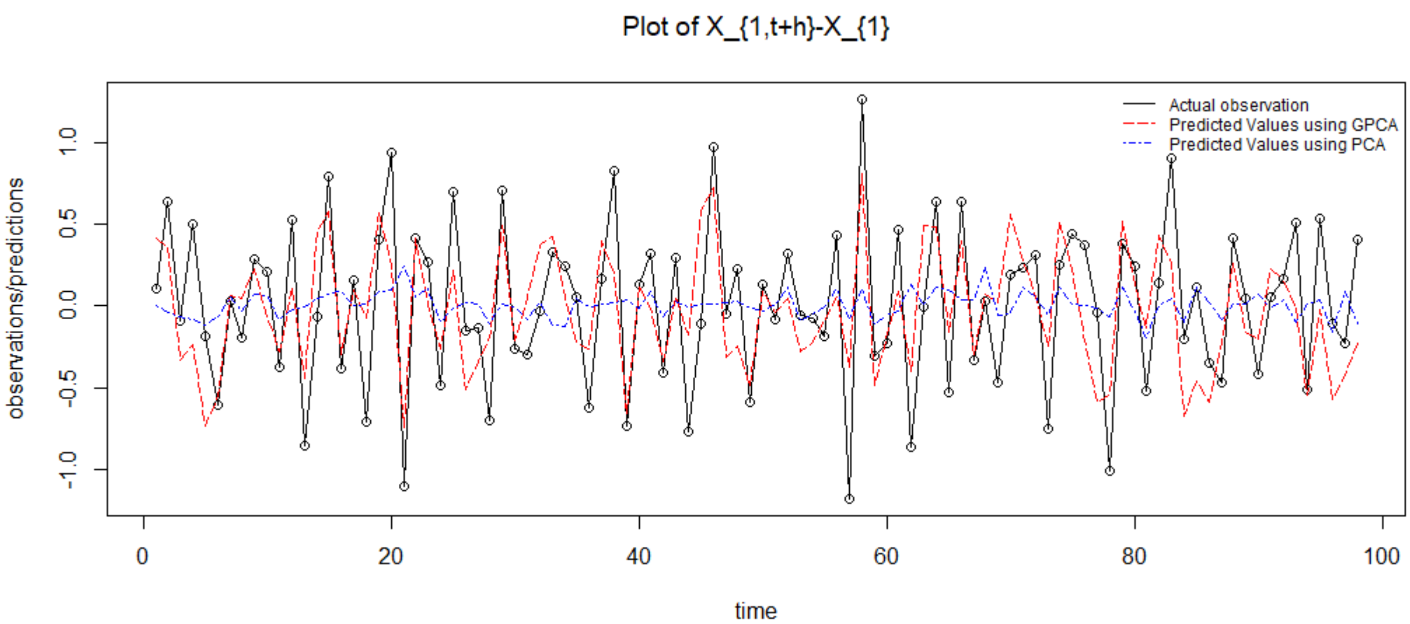


Figure 4.4: Predicted values vs Actual observation ($h = 1$)

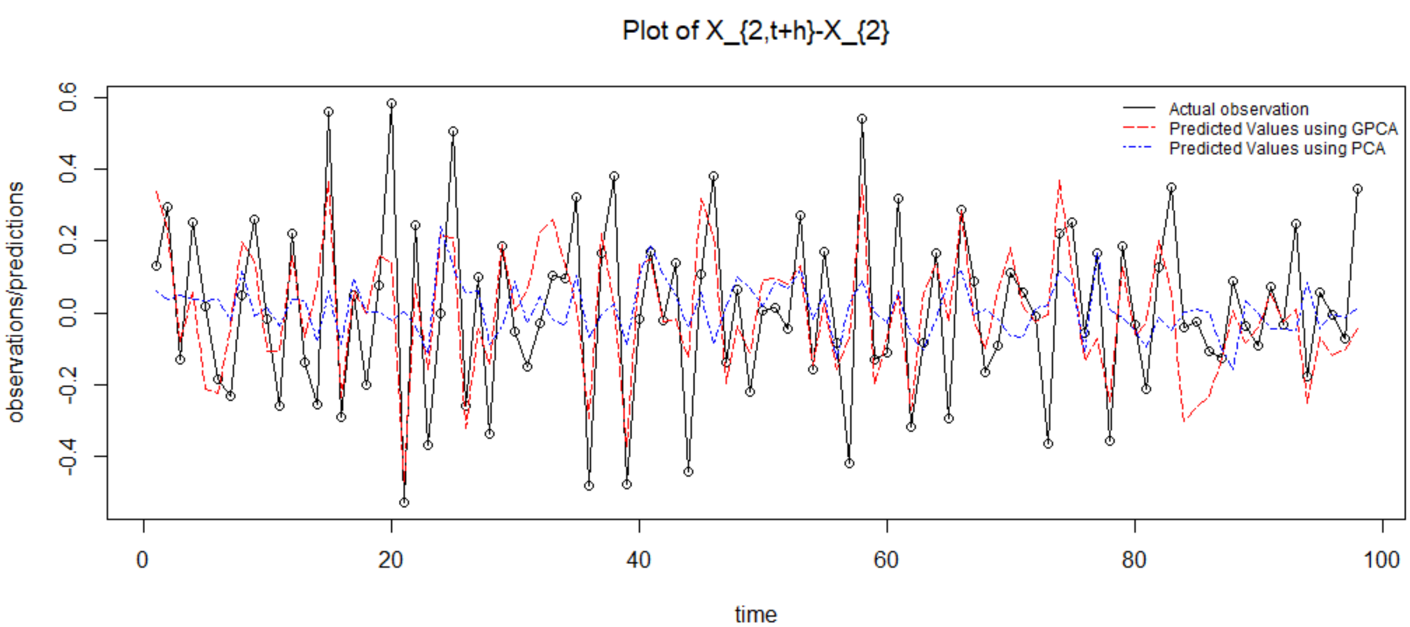


Figure 4.5: Predicted values vs Actual observation ($h = 1$)

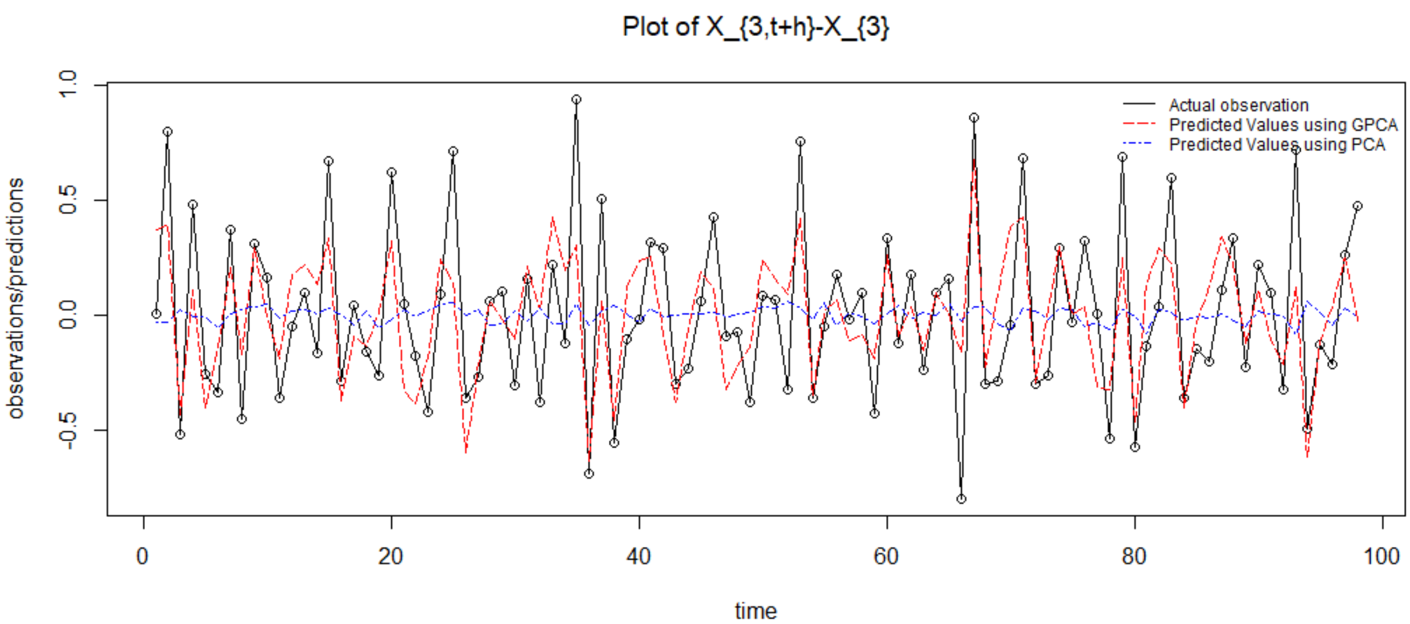


Figure 4.6: Predicted values vs Actual observation ($h = 1$)

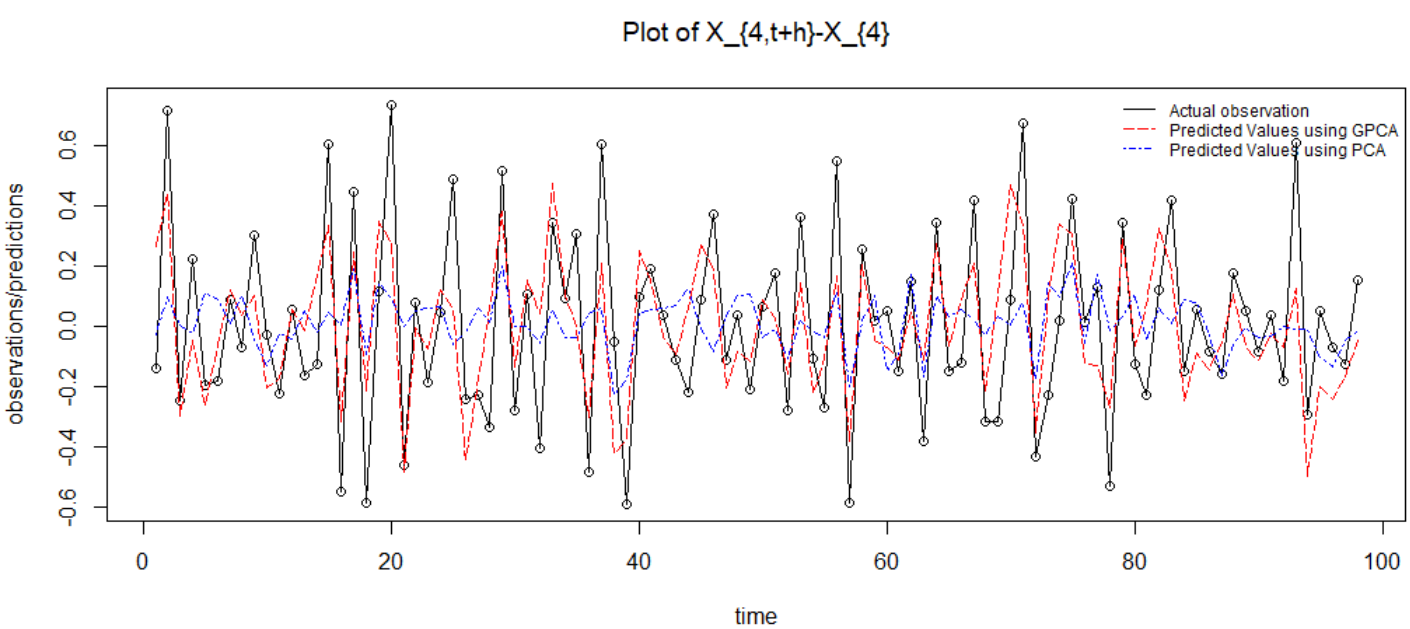


Figure 4.7: Predicted values vs Actual observation ($h = 1$)

4.2 The Student's t VAR (StVAR) Simulation

The Student's t VAR model presented in table 3.3 can be re-parameterized as a StDLR model by introducing a different partition of the bivariate joint distribution presented in 3.27. Let $\mathbf{X}_t=(X_{1t},...,X_{mt})^\top$ and $\boldsymbol{\mu}=E(\mathbf{X}_t)$:

$$\begin{aligned}\begin{pmatrix}\mathbf{X}_t\\ \mathbf{X}_{t-1}\end{pmatrix} &\sim \mathsf{St}_{2m}\left(\begin{pmatrix}\boldsymbol{\mu}\\ \boldsymbol{\mu}\end{pmatrix},\begin{pmatrix}\phi(0) & \phi(1)\\ \phi(1) & \phi(0)\end{pmatrix}\otimes\boldsymbol{\Sigma};\nu\right) &(4.9)\\ &\sim \mathsf{St}_{2m}\left(\begin{pmatrix}\boldsymbol{\mu}\\ \boldsymbol{\mu}\end{pmatrix},\begin{pmatrix}\phi(0)\boldsymbol{\Sigma} & \phi(1)\boldsymbol{\Sigma}\\ \phi(1)\boldsymbol{\Sigma} & \phi(0)\boldsymbol{\Sigma}\end{pmatrix};\nu\right) &(4.10)\end{aligned}$$

Let define $\mathbf{X}_t^j=(X_{1t},...,X_{(j-1)t},X_{(j+1)t},...,X_{mt})$, $\mathbf{W}_t^j=\begin{pmatrix}\mathbf{X}_t^j\\ \mathbf{X}_{t-1}\end{pmatrix}$ and $E(\mathbf{W}_t^j)=\boldsymbol{\mu}_{W_t^j}$ where $j=1,...,m$. For simplicity, assuming $j=1$, the joint distribution in 4.9 can be written as follows:

$$\begin{pmatrix}X_{1t}\\ \mathbf{W}_t^1\end{pmatrix}\sim \mathsf{St}_{2m}\left(\begin{pmatrix}\mu_1\\ \boldsymbol{\mu}_{W_t^1}\end{pmatrix},\begin{pmatrix}\sigma_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix};\nu\right) \tag{4.11}$$

where $\mu_1=E(X_{1t})$, $\frac{1}{\nu-2}\sigma_{11}=Var(X_{1t})$, $\frac{1}{\nu-2}\boldsymbol{\Sigma}_{12}=Cov(X_{1t},\mathbf{W}_t^1)=\frac{1}{\nu-2}\boldsymbol{\Sigma}_{21}^\top$, and $\frac{1}{\nu-2}\boldsymbol{\Sigma}_{22}=Cov(\mathbf{W}_t^1)$.

As we have explained in section 3.2.2, the Student's t, Markov and Stationary process cannot be modeled only in terms of conditional distribution because the weak exogeneity property does not hold. In order to model, Spanos [1994] argues that a estimation of GLS-type estimators can be used to estimate the parameters.[3] The conditional and marginal distributions obtained from decomposing the joint distribution 4.11 is as follows:

[3] Poudyal [2017] provides an R package (StVAR) that is based on the derivations presented in Spanos [1994].

$$(X_{1t}|\mathbf{W}_t^1) \sim \mathsf{St}\big(\alpha + \boldsymbol{\beta}\mathbf{W}_t^1, \sigma; \nu + (2m-1)\big) \tag{4.12}$$

$$\alpha = \mu_1 - \boldsymbol{\beta}\boldsymbol{\mu}_{W_t^1}, \boldsymbol{\beta} = \tfrac{\boldsymbol{\Sigma}_{12}}{\sigma_{11}}, \sigma = \big(\sigma_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\big).q(\mathbf{W}_t^1),$$
$$q(\mathbf{W}_t^1) := [1 + \tfrac{1}{\nu}\mathbf{W}_t^{1\top}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{W}_t^1]$$

$$\mathbf{W}_t^1 \sim \mathsf{St}_{2m-1}(\boldsymbol{\mu}_{W_t^1}, \boldsymbol{\Sigma}_{22}) \tag{4.13}$$

4.2.1 Simulation Design And Forecasting

We use the same $\boldsymbol{\Sigma}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\mu}$ as used in section 4.1 to generate a set of 15 variables with 250 observations based on the joint Student's t distribution with degree of freedom $\nu = 30$.

Also, the forecasting method is similar to that presented in section 4.1 with a different distributional assumption. Again, we use $\mathbf{F}_{it} - X_{it}$ as a central tendency to forecast $X_{it+h} - X_{it}$:

$$X_{it+h} - X_{it} = \boldsymbol{\alpha}_i + \beta(\mathbf{F}_{it} - X_{it}) + \varepsilon_{it+h}, \varepsilon_{it+h} \sim \mathsf{St}(0, \mathbf{V}_t; \nu = 30+1),$$
$$\mathbf{V}_t = \tfrac{\nu}{\nu-1}\big(\sigma_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\big).q(\mathbf{F}_{it} - X_{it}), \tag{4.14}$$
$$q(\mathbf{F}_{it} - X_{it}) := [1 + \tfrac{1}{\nu}(\mathbf{F}_{it} - X_{it})^\top \boldsymbol{\Sigma}_{22}^{-1}(\mathbf{F}_{it} - X_{it})]$$

Table 4.1 presents the median U-statistics in each forecast horizon for both models (Student's t GPCA (StGPCA) vs. Classical PCA) and the number of individual variables (out of 15) with U-statistic less than one and Clark and West t-test greater than 1.960. The detailed table of individual variables for both models are in the appendix (see Table A.3 and Table A.4).

Table 4.2: Forecast evaluation: StGPCA vs. PCA

Model	Measurement	Horizon h h=1	h=4	h=8	h=12
$\widehat{\mathbf{StGPC}}_{it} - X_{it}$	Median U-statistic	0.712	0.743	0.752	0.710
	(#$U < 1$ out of 15)	(15)	(15)	(15)	(15)
	[#$t > 1.960$ out of 15]	[15]	[4]	[0]	[0]
$\widehat{\mathbf{PC}}_{it} - X_{it}$	Median U-statistic	0.999	1.002	0.996	1.000
	(#$U < 1$ out of 15)	(9)	(7)	(12)	(7)
	[#$t > 1.960$ out of 15]	[1]	[0]	[0]	[0]

Note: $\widehat{\mathbf{StGPC}}_{it} - s_{it}$ and $\widehat{\mathbf{PC}}_{it} - s_{it}$ represent deviations from factors produced by the StGPCA and the classical PCA, respectively. The number of variables (out of 15) with U-statistic (Theil [1971]) less than one and the number of variables (out of 15) with Clark-West t-statistic (Clark and West [2006]) more than 1.960 are reported in parenthesis and brackets, respectively.

Also, Figure 4.8 to 4.11 presents the comparison between actual observations, the StGPCA predictions, and the classical PCA predictions for horizon $h = 1$.[4]

[4]Plots of the all variables in all horizons are presented in the appendix A.5

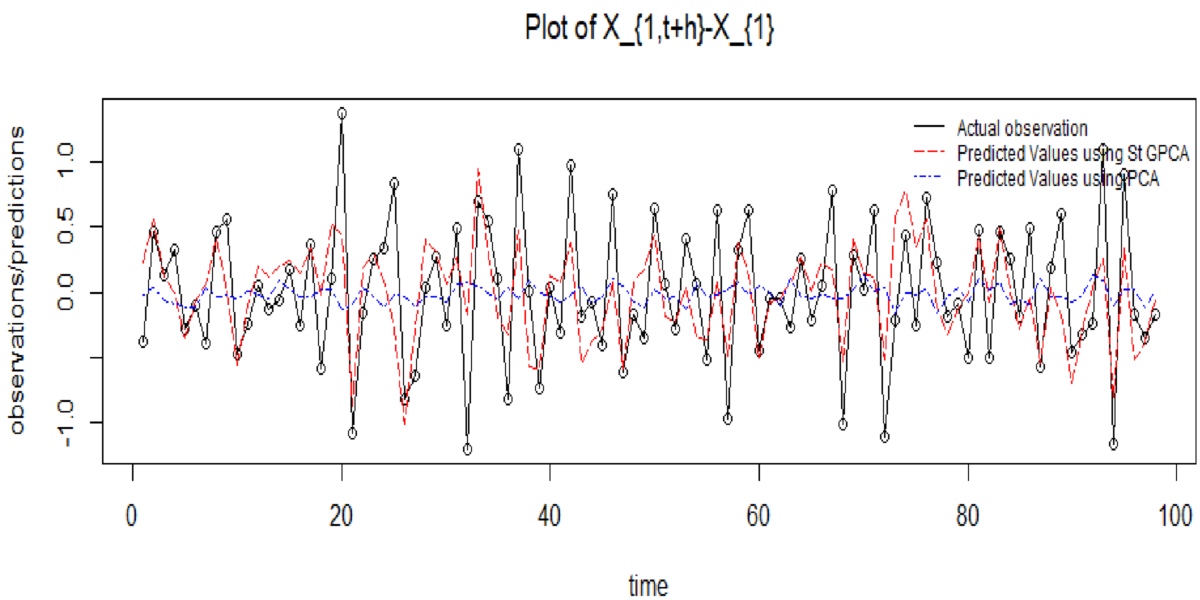


Figure 4.8: Predicted values vs Actual observation ($h = 1$)

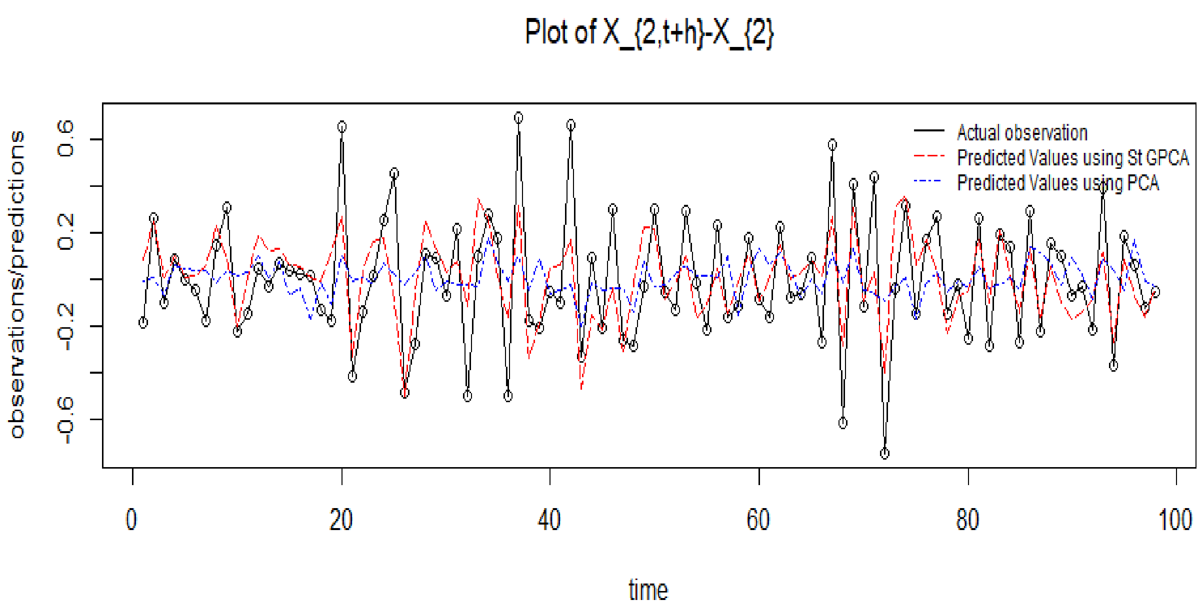


Figure 4.9: Predicted values vs Actual observation ($h = 1$)

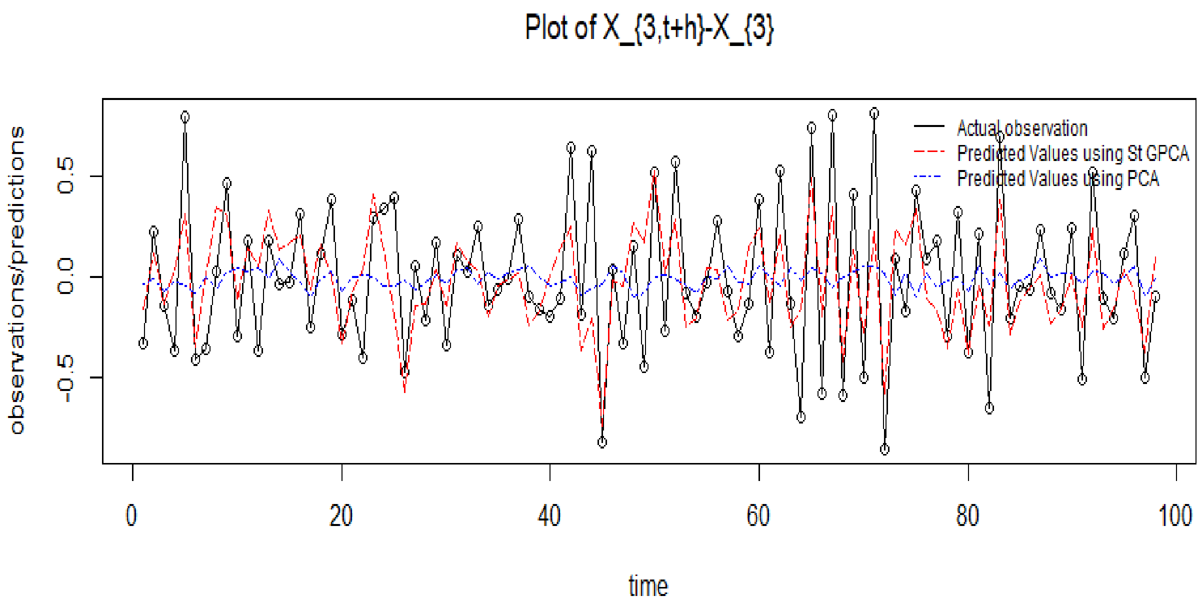


Figure 4.10: Predicted values vs Actual observation ($h = 1$)

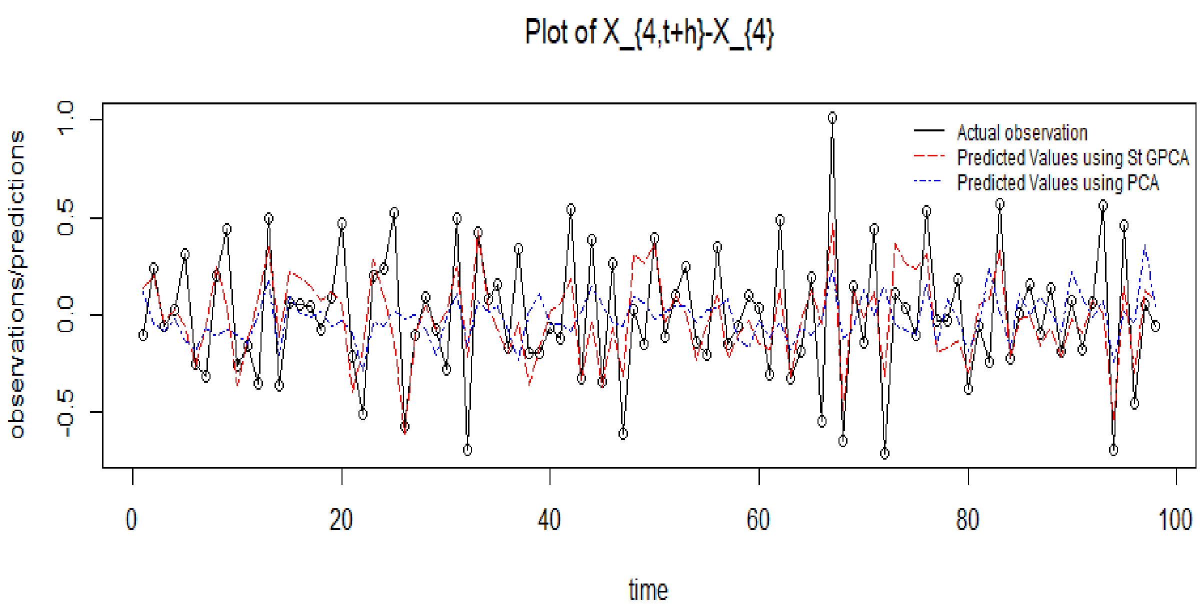


Figure 4.11: Predicted values vs Actual observation ($h = 1$)

Chapter 5

Empirical Study

5.1 Introduction

The random walk model is hard to beat in forecasting exchange rates, and this finding has more or less survived the numerous studies since Meese and Rogoff [1983a] and Meese and Rogoff [1983b]. The model essentially forecasts that log level of exchange rate remains the same in the future, and this seemingly simple model beats well-founded, sophisticated models of exchange rates that make use of economic fundamentals like output, interest rates, or inflation rates. It is a well-established finding for horizons from 1 quarter to 3 years, while the results are more ambiguous for longer horizons.[1]

Instead of looking for new fundamentals or econometric methods to beat the random walk, some recent papers look for predictability of the exchange rates. In particular, factors are extracted from a panel of exchange rates, and the deviations of the exchange rates from the factors are used to forecast their future changes.[2] Engel et al. [2015] first propose this new direction and find mixed results. They extract three principal components from a panel of 17 exchange rates (with the US dollar as the base currency), and they find that the factors

[1]Engel and West [2005] shows that random walk dominates when the discount factor is near one and the fundamentals are persistent. For a recent survey on the empirical findings, see Rossi [2013] and Maasoumi and Bulut [2013].

[2]For simplicity, we abuse the usage and refer to principal component as "factor" in this chapter.

improve the random walk only for long horizons during the more recent period (1999 to 2007). Wang and Wu [2015] adopt the method of independent component factors that is robust to fat tails, and, using a longer sample, they are able to beat the random walk at medium and long horizons regardless of the sample periods.

Our empirical study here follows this line of research and extracts factors in a simple and intuitive way. We adopt a more general approach and make use of temporal covariations as well as contemporaneous covariations as we have explained in Chapter 3. Though we are agnostic on what the factors represent, we believe that we are better at capturing unobserved fundamentals that make exchange rates persistent and correlated through time. Indeed, we find that relaxing the assumptions imposed to the classical PCA substantially improves the forecasting performance of the factors in Engel et al. [2015] by beating the random walk in all horizons and sample periods. We also show that our more transparent method improves upon that proposed by Wang and Wu [2015].

We use the method explained in Chapter 3 to extract the GPCs and compare our forecasting performance with that in Engel et al. [2015] and Wang and Wu [2015], using the same data analyzed in each paper.

5.2 Empirical Results

5.2.1 Data

We use end of quarter data on nominal bilateral US dollar exchange rates of 17 *Organization for Economics Co-operation and Development* (OECD) countries from 1973:1 to

2007:4.[3] The countries are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Japan, Italy, Korea, Netherlands, Norway, Spain, Sweden, Switzerland, and the United Kingdom. Table 5.1 presents the descriptive statistic summary of the data.

Table 5.1: Summary Statistics

Country	N	Mean	SD	Min	Max	Skew	Kurtosis
Australia	140	0.189	0.269	−0.399	0.715	−0.44	2.39
Canada	140	0.220	0.130	−0.032	0.466	−0.13	2.28
Denmark	140	1.897	0.180	1.624	2.421	0.95	3.34
United Kingdom	140	−0.549	0.157	−0.949	−0.145	−0.45	2.94
Japan	140	5.061	0.372	4.438	5.721	0.42	1.63
Korea	140	6.659	0.334	5.985	7.435	−0.23	2.42
Norway	140	1.880	0.163	1.577	2.230	0.09	2.34
Sweden	140	1.866	0.264	1.371	2.384	−0.33	2.09
Switzerland	140	0.512	0.268	0.118	1.177	0.70	2.54
Austria	140	2.614	0.212	2.235	3.093	0.36	2.06
Belgium	140	3.609	0.184	3.311	4.144	0.86	3.45
France	140	1.731	0.196	1.391	2.261	0.54	2.86
Germany	140	0.657	0.209	0.284	1.147	0.37	2.11
Spain	140	4.732	0.342	4.025	5.279	−0.66	2.36
Italy	140	7.185	0.344	6.335	7.733	−0.80	2.73
Finland	140	1.546	0.177	1.263	1.948	0.37	2.11
Netherlands	140	0.762	0.197	0.404	1.267	0.43	2.40

Note: Quarterly log-exchange rates based on the US dollar 1973:1-2007:4

5.2.2 Models Of Exchange Rates

In this section we show that the GPCA can perform better than other methods of factor modeling in the context of exchange rate forecasting, and we will focus the discussion on certain arguments that have been presented by Engel et al. [2015]. The reason is that Engel et al. [2015] includes a comprehensive analysis of factor model specifications and auxiliary macro-variables along with the results from different factor models adopted to conduct out-of-sample forecasting of exchange rates. We want to examine if there is any improvement

[3] The data source is International Financial Statistics.

in the context of out-of-sample forecasting by replacing their factorization method with the GPCA method. Although in some cases the PCA method for British Pound, as a base currency, shows improvement when compare to the factor analysis (FA) method, Engel et al. [2015] conclude that overall results for the FA method are better than the PCA method. We compare the out-of-sample forecasting capacity of the GPCA method to the FA method adopted by Engel et al. [2015].

We construct three sets of out-of-sample forecasting. First, for the 9 non-Euro currencies (Australia, Canada, Denmark, Japan, Korea, Norway, Sweden, Switzerland, and the United Kingdom) called "long sample" forecasting for the time period 1987:1 to 2007:4. Second, for the 17 currencies (Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Japan, Italy, Korea, Netherlands, Norway, Spain, Sweden, Switzerland, and the United Kingdom) called "early sample" forecasting for the time period 1987:1 to 1998:4 (before Euro). Finally, for the 10 currencies (countries included in long sample plus Euro) called "late sample" for the time period 1999:1 to 2007:4.

To determine the number of GPCs we use the eigenvalue test which gives the percentage of variation explained through the retained GPCs. Three components, will explain 96% of the variation in the data (similar to the PCA). By the method explained in section 3.1 we derive GPCs and estimate the coefficients based on the following model:

$$\begin{aligned} s_{it} &= const. + \delta_{1i} gpc_{1t} + \delta_{2i} gpc_{2t} + \delta_{3i} gpc_{3t} + u_{it}, \\ &= const. + GPC_{it} + u_{it}, u_{it} \sim \mathsf{NIID}(0, \sigma_u^2), \end{aligned} \tag{5.1}$$

where s_{it}, (i=1, ..., 17), is the log of nominal exchange rates of currency i based on the US dollar, the derived GPCs are gpc_{1t}, gpc_{2t} & gpc_{3t}. We aim to use $GPC_{it} = \delta_{1i} gpc_{1t} + \delta_{2i} gpc_{2t} + \delta_{3i} gpc_{3t}$ to forecast s_{it}.

The rest of the model specifications that we take into account is similar to what has been proposed by Engel et al. [2015]. First we assume that $GPC_{it} - s_{it}$ is stationary and can be useful to capture the stationary regularity of future values of s_{it} through $s_{it+h} - s_{it}$ where h=1, 4, 8, 12 is quarterly horizons of forecasting.

Let $\widehat{GPC_{it}}$=$\widehat{\delta}_{1i}\widehat{gpc}_{1t} + \widehat{\delta}_{2i}\widehat{gpc}_{2t} + \widehat{\delta}_{3i}\widehat{gpc}_{3t}$ for currencies i=1, ..., 17. We use it as a central tendency to estimate the coefficients of the following regression:

$$s_{it+h} - s_{it} = \alpha_i + \beta(\widehat{GPC_{it}} - s_{it}) + \epsilon_{it+h}, \epsilon_{it+h} \sim \mathsf{NIID}(0, \sigma^2_{\epsilon_i}) \tag{5.2}$$

where α_i is the individual effect of currency i. The estimated coefficients $\widehat{\alpha}_i$ and $\widehat{\beta}$ can be used to predict the future value of the nominal exchange rates.

As a typical example, figure 5.1 illustrates the procedure for quarterly horizon h=4 in the "long sample" forecasting:

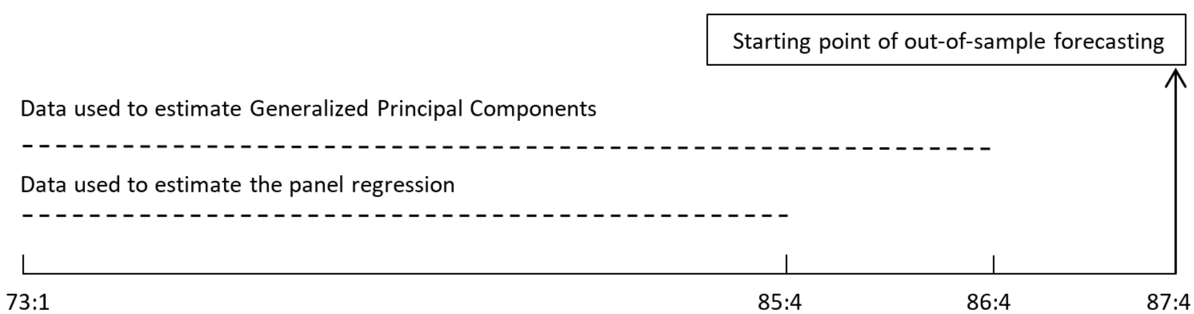


Figure 5.1: Forecasting Procedure ($h = 4$, long sample)

We use data from 1973:1 to 1986:4 to estimate $\widehat{GPC_{it}}$ and then estimate the panel regression

$$s_{it+4} - s_{it} = \alpha_i + \beta(\widehat{GPC_{it}} - s_{it}) + \epsilon_{it+4},\ t \in \{1973{:}1, ..., 1985{:}4\}. \tag{5.3}$$

Using the estimated coefficients $\widehat{\alpha}_i$ and $\widehat{\beta}$ from the regression (5.3), we evaluate the predicted value of $s_{i,1987:4} - s_{i,1986:4}$ using the following equation

$$s_{i,1987:4} - s_{i,1986:4} = \widehat{\alpha}_i + \widehat{\beta}(\widehat{GPC}_{i,1986:4} - s_{i,1986:4}). \tag{5.4}$$

We repeat this procedure by adding another observation to the end of the sample to produce predictions by a recursive method.[4] Also, the forecast evaluation is based on the method and measurement presented in Engel et al. [2015] which is root mean squared prediction error (RMSPE). We compute Theil's U-statistic (Theil [1971]) that is equal to a ratio by dividing RMSPE of factor model (GPCA or FA) to the RMSPE of the random walk model. The U-statistic less than one means that the factor model has a better performance than the assumed random walk model.

5.2.3 Discussion Of Results

In the PCA, most of the variation among individual variables has been explained by the first three factors. In addition to what has been captured by the PCA, the GPCA also captures the variation and co-variation between different points of time across all variables. That is the reason why factors are converging to the same pattern despite some differences at the beginning. Figure 5.2[5] depicts the time plot of three GPCs for the whole sample 1973:1 to 2007:4.

[4] We need to centralize data to extract the factors, and, to make sure that the forecasts are truly out-of-sample, data are centralized only using in-sample data.

[5] Note that the boxes are the plots of GPCs from 10^{th} observation to the end of data set

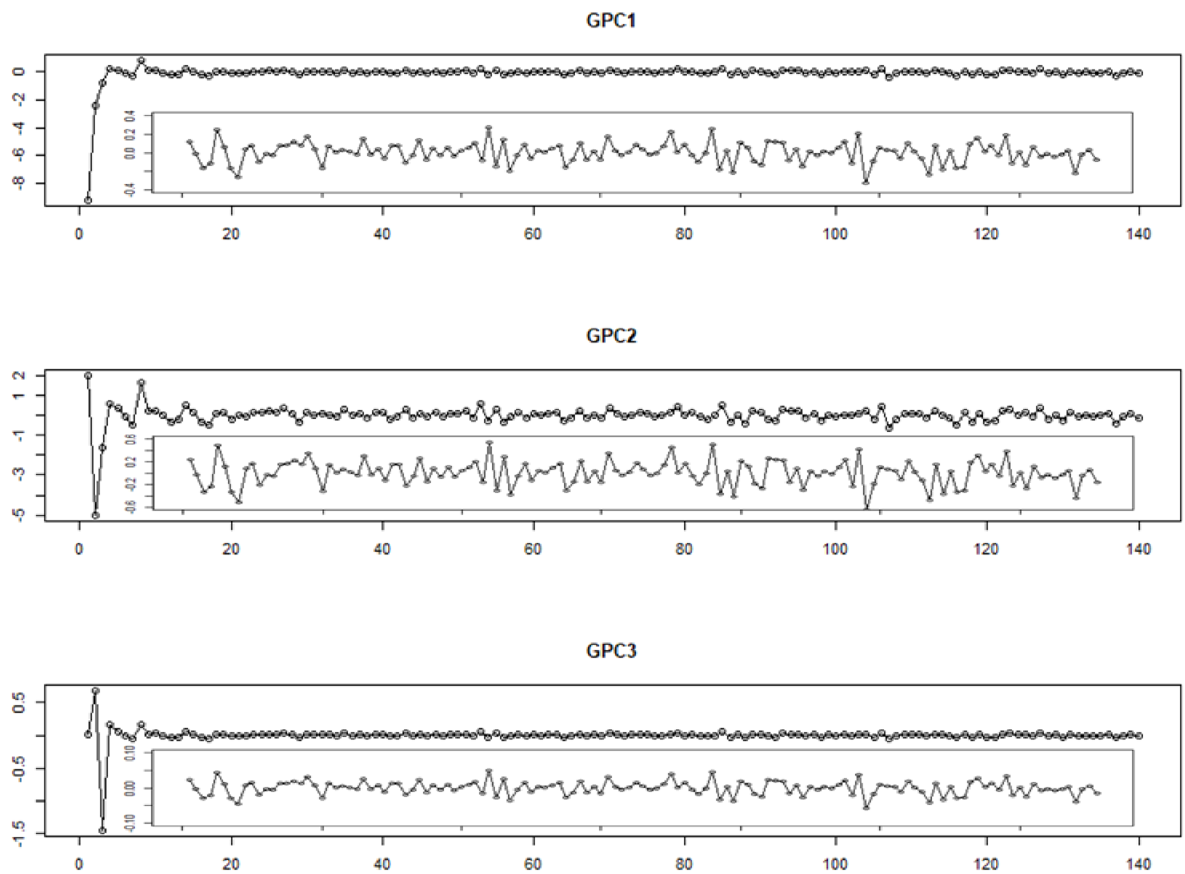


Figure 5.2: Generalized Principal Components t-plot

Table 5.2 presents the median Theil's U-statistics for early, late and long samples regarding to the following model: [6]

- The model that uses the GPCA to extract factors for $(\widehat{GPC}_{it} - s_{it})$, and
- The model that uses FA to extract factors for $(\widehat{F}_{it} - s_{it})$.[7]

[6] The U-statistic is defined as the ratio of $RMSPE_{Model}$ to $RMSPE_{RandomWalk}$. Results for individual countries are available upon request.

[7] Although the results for three factors model have not been reported in Engel et al. [2015], fortunately, they have made their codes available on their website (http://www.ssc.wisc.edu/~cengel/Data/Factor/FactorData.htm) for replication.

Table 5.2: Forecast evaluation: GPCA vs FA (Engel et al. [2015])

Model	Sample(# Currencies)	Measurement	Horizon h h=1	h=4	h=8	h=12
$\widehat{GPC}_{it} - s_{it}$	Long sample (9)	Median U-statistic ($\#U < 1$)	**0.996**(7)	**0.963**(7)	**0.926**(8)	**0.905**(8)
$\widehat{F}_{it} - s_{it}$	Long sample (9)	Median U-statistic ($\#U < 1$)	1.003(3)	**0.996**(5)	**0.996**(5)	1.038(4)
$\widehat{GPC}_{it} - s_{it}$	Early sample (17)	Median U-statistic ($\#U < 1$)	**0.993**(15)	**0.957**(14)	**0.919**(16)	**0.973**(9)
$\widehat{F}_{it} - s_{it}$	Early sample (17)	Median U-statistic ($\#U < 1$)	1.000(10)	**0.995**(9)	1.000(9)	1.130(3)
$\widehat{GPC}_{it} - s_{it}$	Late sample (10)	Median U-statistic ($\#U < 1$)	**0.993**(7)	**0.970**(8)	**0.888**(9)	**0.788**(10)
$\widehat{F}_{it} - s_{it}$	Late sample (10)	Median U-statistic ($\#U < 1$)	1.008(3)	1.020(3)	**0.953**(8)	**0.822**(8)

Note: $\widehat{GPC}_{it} - s_{it}$ and $\widehat{F}_{it} - s_{it}$ represent deviations from factors produced by the GPCA and the FA, respectively.

The first column indicates the factor model that has been used in the forecasting evaluations. The second column lists the type of sample and number of currencies in that sample. The third column presents the measurement method that has been used to evaluate the forecastability power of the model which is the median U-statistic. Also, it reports the number of currencies in the sample that have the U-statistic value less than one[8]. The last four columns are reporting the median U-statistic for different horizons (h=1, 4, 8, 12) and samples.

The results presented in Table 5.2 show that the GPCA outperforms both the FA and the driftless random walk models in all cases. The FA model used by Engel et al. [2015] has better predictive performance than the random walk model only in 5 cases, and in all the 5 cases the FA is dominated by the GPCA.

Engel et al. [2015] use three sets of auxiliary macro variables as a measure of central tendency

[8] U-statistic less than one means that the model has smaller RMSPE compare to the random walk.

to improve the factor model in a way that captures more regularity pattern in the exchange rates to forecast more accurately. These auxiliary macro variables are "monetary model" (Mark [1995]), "Taylor rule" (Molodtsova and Papell [2009]) and PPP (Engel et al. [2007]). Table 5.3 compares the results that has been obtained by the GPCA method without any auxiliary macro variables with the FA method presented in Engel et al. [2015] using auxiliary macro variables.

Table 5.3: Forecast evaluation: GPCA vs FA+Macro variables (Engel et al. [2015])

			Horizon h			
Model	Sample(# Currencies)	Measurement	h=1	h=4	h=8	h=12
$\widehat{GPC}_{it} - s_{it}$	Long sample (9)	Median U-statistic ($\#U<1$)	**0.996**(7)	**0.963**(7)	**0.926**(8)	**0.905**(8)
$\widehat{F}_{it} - s_{it} + Taylor$	Long sample (9)	Median U-statistic ($\#U<1$)	1.008(1)	1.035(0)	1.068(1)	1.052(3)
$\widehat{F}_{it} - s_{it} + Monetary$	Long sample (9)	Median U-statistic ($\#U<1$)	1.008(3)	1.064(3)	1.200(4)	1.456(4)
$\widehat{F}_{it} - s_{it} + PPP$	Long sample (9)	Median U-statistic ($\#U<1$)	1.002(3)	**0.993**(6)	**0.942**(7)	**0.903**(5)
$\widehat{GPC}_{it} - s_{it}$	Early sample (17)	Median U-statistic ($\#U<1$)	**0.993**(15)	**0.957**(14)	**0.919**(16)	**0.973**(9)
$\widehat{F}_{it} - s_{it} + Taylor$	Early sample (17)	Median U-statistic ($\#U<1$)	1.010(3)	1.041(2)	1.103(4)	1.156(3)
$\widehat{F}_{it} - s_{it} + Monetary$	Early sample (17)	Median U-statistic ($\#U<1$)	**0.995**(10)	**0.997**(9)	1.115(7)	1.190(7)
$\widehat{F}_{it} - s_{it} + PPP$	Early sample (17)	Median U-statistic ($\#U<1$)	**0.998**(7)	**0.972**(14)	1.015(8)	1.098(3)
$\widehat{GPC}_{it} - s_{it}$	Late sample (10)	Median U-statistic ($\#U<1$)	**0.993**(7)	**0.970**(8)	**0.888**(9)	**0.788**(10)
$\widehat{F}_{it} - s_{it} + Taylor$	Late sample (10)	Median U-statistic ($\#U<1$)	1.009(2)	1.036(2)	1.004(4)	**0.828**(8)
$\widehat{F}_{it} - s_{it} + Monetary$	Late sample (10)	Median U-statistic ($\#U<1$)	1.013(3)	1.033(4)	**0.977**(6)	1.126(5)
$\widehat{F}_{it} - s_{it} + PPP$	Late sample (10)	Median U-statistic ($\#U<1$)	1.005(4)	**0.999**(5)	**0.900**(8)	**0.727**(9)

Note: $\widehat{GPC}_{it} - s_{it}$ and $\widehat{F}_{it} - s_{it}$ represent deviations from factors produced by the GPCA and the FA, respectively.

Based on the results presented in Table 5.3, the GPCA by itself is outperforming the FA method with auxiliary macro variables on forecasting grounds.

5.2.4 Comparing With An Alternative Method

Another study related to the predictability of exchange rates that has been published recently is a paper by Wang and Wu [2015]. In this paper it is argued that the information relating to the third moment can be useful to improve the forecasting capacity of the exchange rates forecasting models. They apply the denoising source separation (DSS) algorithm (Särelä and Valpola [2005]) on the normalized nominal exchange rates $s_{it}^n = \frac{s_{it} - \mu_{s_i}}{\sigma_{s_i}}$ to extract independent components (IC) and mixing coefficients that can be used to construct an IC-based fundamental exchange rate ($\widehat{E}_{it}$). $\widehat{E}_{it} - s_{it}^n$ can be used to predict $s_{it+h} - s_{it}$. The rest of the model is similar to the model presented by Engel et al. [2015]. The number of factors has been determined by three different criteria, the cumulative percentage of total variance (CPV) (Jackson [1993]), Bayesian information criterion (BIC_3) and IC_{p2} (Bai and Ng [2004]). They conclude that the IC-based model can perform better than the PCA in the context of out-of-sample forecasting of exchange rates.

Wang and Wu [2015] use the quarterly log-exchange rates based on US dollar for the same 17 OECD countries that have been used in Engel et al. [2015]. Although, they have used the data from 1973:1 to 2011:2. Table 5.4 presents the descriptive statistics of this data.

Table 5.4: Summary Statistics

Country	N	Mean	SD	Min	Max	Skew	Kurtosis
Australia	154	0.183	0.260	−0.399	0.715	−0.39	2.46
Canada	154	0.205	0.134	−0.036	0.466	−0.03	2.13
Denmark	154	1.876	0.185	1.551	2.421	0.92	3.38
United Kingdom	154	−0.544	0.154	−0.949	−0.145	−0.51	3.03
Japan	154	5.012	0.388	4.391	5.721	0.45	1.78
Korea	154	6.695	0.339	5.985	7.435	−0.35	2.37
Norway	154	1.870	0.161	1.577	2.230	0.18	2.38
Sweden	154	1.873	0.254	1.371	2.384	−0.40	2.26
Switzerland	154	0.467	0.294	−0.181	1.177	0.47	2.62
Austria	154	2.583	0.225	2.164	3.093	0.35	2.11
Belgium	154	3.586	0.191	3.239	4.144	0.80	3.39
France	154	1.713	0.196	1.391	2.261	0.64	2.94
Germany	154	0.627	0.222	0.213	1.147	0.36	2.15
Spain	154	4.736	0.327	4.025	5.279	−0.72	2.60
Italy	154	7.188	0.329	6.335	7.733	−0.87	3.02
Finland	154	1.537	0.173	1.263	1.948	0.49	2.26
Netherlands	154	0.733	0.210	0.332	1.267	0.39	2.38

Note: Quarterly log-exchange rates based on the US dollar 1973:1-2011:2

Table 5.5 compares the results based on the GPCA method and the IC-based model, using the data provided by Wang and Wu [2015]. For most horizons the GPCA forecasts better than the IC-based model, the only exception being horizon $h = 12$.

Table 5.5: Forecast evaluation: GPCA vs ICA (Wang and Wu [2015])

Model	Sample(# Currencies)	Measurement	Horizon h h=1	h=4	h=8	h=12
$\widehat{GPC}_{it} - s_{it}$	Long sample (9)	Median U-statistic ($\#U<1$)	**0.995**(7)	**0.969**(8)	**0.938**(7)	**0.961**(7)
$\widehat{E}_{it} - s_{it}^n$(Criterion: CPV)	Long sample (9)	Median U-statistic ($\#U<1$)	1.000(4)	**0.986**(7)	**0.956**(7)	**0.946**(9)
$\widehat{E}_{it} - s_{it}^n$(Criterion: BIC$_3$)	Long sample (9)	Median U-statistic ($\#U<1$)	1.000(4)	**0.991**(8)	**0.955**(7)	**0.941**(9)
$\widehat{E}_{it} - s_{it}^n$(Criterion: IC$_{p2}$)	Long sample (9)	Median U-statistic ($\#U<1$)	1.001(3)	1.002(3)	**0.974**(7)	**0.951**(9)
$\widehat{GPC}_{it} - s_{it}$	Early sample (17)	Median U-statistic ($\#U<1$)	**0.993**(15)	**0.957**(14)	**0.919**(16)	**0.973**(9)
$\widehat{E}_{it} - s_{it}^n$(Criterion: CPV)	Early sample (17)	Median U-statistic ($\#U<1$)	**0.999**(11)	**0.991**(13)	**0.950**(13)	**0.965**(10)
$\widehat{E}_{it} - s_{it}^n$(Criterion: BIC$_3$)	Early sample (17)	Median U-statistic ($\#U<1$)	**0.999**(10)	**0.994**(13)	**0.956**(14)	**0.964**(10)
$\widehat{E}_{it} - s_{it}^n$(Criterion: IC$_{p2}$)	Early sample (17)	Median U-statistic ($\#U<1$)	1.000(7)	**0.999**(9)	**0.976**(14)	**0.976**(10)

Note: $\widehat{GPC}_{it} - s_{it}$ and $\widehat{E}_{it} - s_{it}^n$ represent deviations from factors produced by the GPCA and the ICA, respectively.

5.2.5 Forecasting Using The Updated Data

To evaluate the reliability of GPCA method further, it is important to investigate the consistency of the results when we increase the sample size. Therefore, we report the results by updating the data to 2017:4. Table 5.6 presents the descriptive statistics for the period 1973:1 to 2017:4; and Table 5.7 presents the median Theil's U-statistic obtained from using the updated data.

Table 5.6: Summary Statistics

Country	N	Mean	SD	Min	Max	Skew	Kurtosis
Australia	180	0.179	0.246	−0.399	0.715	−0.36	2.63
Canada	180	0.197	0.133	−0.036	0.466	0.01	2.09
Denmark	180	1.867	0.176	1.551	2.421	1.03	3.80
United Kingdom	180	−0.523	0.155	−0.949	−0.145	−0.55	3.20
Japan	180	4.955	0.389	4.339	5.721	0.60	2.05
Korea	180	6.742	0.334	5.985	7.435	−0.60	2.51
Norway	180	1.881	0.164	1.577	2.230	0.18	2.20
Sweden	180	1.893	0.245	1.371	2.384	−0.55	2.49
Switzerland	180	0.392	0.328	−0.181	1.177	0.40	2.43
Austria	180	2.560	0.218	2.164	3.093	0.53	2.32
Belgium	180	3.573	0.182	3.239	4.144	0.93	3.78
France	180	1.709	0.185	1.391	2.261	0.71	3.27
Germany	180	0.605	0.215	0.213	1.147	0.54	2.37
Spain	180	4.762	0.310	4.025	5.279	−0.92	3.04
Italy	180	7.215	0.313	6.335	7.733	−1.06	3.52
Finland	180	1.544	0.164	1.263	1.948	0.40	2.34
Netherlands	180	0.713	0.203	0.332	1.267	0.56	2.61

Note: Quarterly log-exchange rates based on the US dollar 1973:1-2017:4

Table 5.7: Forecast evaluation: GPCA using data from 1973:1 to 2017:4

			Horizon h			
Model	Sample(# Currencies)	Measurement	h=1	h=4	h=8	h=12
$\widehat{GPC_{it}} - s_{it}$	Long sample (9)	Median U-statistic ($\#U < 1$)	**0.992**(9)	**0.961**(8)	**0.923**(7)	**0.915**(7)
$\widehat{GPC_{it}} - s_{it}$	Early sample (17)	Median U-statistic ($\#U < 1$)	**0.993**(15)	**0.957**(14)	**0.919**(16)	**0.973**(9)
$\widehat{GPC_{it}} - s_{it}$	Late sample (10)	Median U-statistic ($\#U < 1$)	**0.992**(9)	**0.964**(9)	**0.907**(9)	**0.854**(9)

Note: $\widehat{GPC_{it}} - s_{it}$ represents deviations from factors produced by the GPCA.

Chapter 6

Conclusion

The discussion in this dissertation extends the traditional PCA to include temporal dependence as well as non-Gaussian distributions. The proposed generalized principal components analysis (GPCA) method substantially improves the out-of-sample predictability of factors. Using two Monte Carlo simulation designs, we show that the GPCA method can capture most of the volatility in the data while the classical PCA method performs poorly due to ignoring the temporal dependence and distributional nature of the data.

In addition, the empirical study using exchange rate data shows that employing factors that incorporate both contemporaneous and temporal covariation in the data, substantially improves the out-of-sample forecasting performance. In addition, exchange rates are found to converge to the GPCA factors, while the convergence is not as clear when traditional methods of extracting factors are used (with or without including macroeconomic fundamentals).

As with the traditional PCA, the retained factors in the GPCA represent a linear combinations of the original observable variables and thus any attempt to interpret them, or use them for policy analysis will often be difficult. The PCA and GPCA should be viewed as data-based statistical models whose substantive interpretation is not clear cut.

The results of this dissertation can be extended in a number of different directions, including:

- Replacing the Student's t with other distributions within the Elliptically symmetric family.

- Explore different types of temporal dependence.

Bibliography

Theodore W Anderson and Herman Rubin. Statistical inference in factor analysis. 5:1, 1956.

Theodore Wilbur Anderson. Nonnormal multivariate distributions: Inference based on elliptically contoured distributions. 1992.

Theodore Wilbur Anderson. *An introduction to multivariate statistical analysis*. Wiley-Interscience, 2003a.

Theodore Wilbur Anderson. An introduction to multivariate statistical analysis. 2003b.

Theodore Wilbur Anderson and Kai-Tai Fang. On the theory of multivariate elliptically contoured distributions and their applications. 1982.

Jushan Bai. Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics*, 122(1):137–183, 2004. ISSN 0304-4076.

Jushan Bai and Serena Ng. A panic attack on unit roots and cointegration. *Econometrica*, 72(4):1127–1177, 2004. ISSN 1468-0262.

Eugenio Beltrami. Sulle funzioni bilineari. *Giornale di Matematiche ad Uso degli Studenti Delle Universita*, 11(2):98–106, 1873.

Stamatis Cambanis, Steel Huang, and Gordon Simons. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3):368–385, 1981. ISSN 0047-259X.

Todd E Clark and Kenneth D West. Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *Journal of Econometrics*, 135(1):155–186, 2006.

Charles Engel and Kenneth D West. Exchange rates and fundamentals. *Journal of political Economy*, 113(3):485–517, 2005. ISSN 0022-3808.

Charles Engel, Nelson C Mark, Kenneth D West, Kenneth Rogoff, and Barbara Rossi. Exchange rate models are not as bad as you think [with comments and discussion]. *NBER Macroeconomics annual*, 22:381–473, 2007. ISSN 0889-3365.

Charles Engel, Nelson C Mark, and Kenneth D West. Factor model forecasts of exchange rates. *Econometric Reviews*, 34(1-2):32–55, 2015. ISSN 0747-4938.

MA Fisher. On the mathematical foundations of theoretical statistics. *Phil. Trans. R. Soc. Lond. A*, 222(594-604):309–368, 1922.

Francis Galton. Natural inheritance (1889). *SERIES E: PHYSIOLOGICAL PSYCHOLOGY*, 1889.

AK Gupta and T Varga. A new class of matrix variate elliptically contoured distributions. *Statistical Methods Applications*, 3(2):255–270, 1994a. ISSN 1618-2510.

AK Gupta and T Varga. A new class of matrix variate elliptically contoured distributions. *Journal of the Italian Statistical Society*, 3(2):255–270, 1994b. ISSN 1121-9130.

Arjun K Gupta and Daya K Nagar. *Matrix variate distributions*, volume 104. CRC Press, 1999. ISBN 1584880465.

Arjun K Gupta and T Varga. Some applications of the stochastic representation of elliptically contoured distribution. *Random Operators and Stochastic Equations*, 2(1):1–12, 1994c. ISSN 1569-397X.

Arjun K Gupta, Tamas Varga, and Taras Bodnar. *Elliptically contoured models in statistics and portfolio theory*. Springer, 2013. ISBN 1461481538.

S Das Gupta, Morris L Eaton, Ingram Olkin, M Perlman, Leonard J Savage, and Milton Sobel. Inequalities on the probability content of convex regions for elliptically contoured distributions. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971)*, volume 2, pages 241–265, 1972.

Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933. ISSN 1939-2176.

Donald A Jackson. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology*, 74(8):2204–2214, 1993. ISSN 1939-9170.

Ian T Jolliffe. *Principal Component Analysis*. Springer, 1986.

Camille Jordan. Mémoire sur les formes bilinéaires. *Journal de mathématiques pures et appliquées*, 19:35–54, 1874.

Douglas Kelker. Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 419–430, 1970.

Esfandiar Maasoumi and Levent Bulut. Predictability and specification in models of exchange rate determination. In *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis: Essays in Honor of Halbert L. White Jr, edited by Xiaohong Chen and Norman R. Swanson*, pages 411–436. Springer, 2013.

Nelson C Mark. Exchange rates and fundamentals: Evidence on long-horizon predictability. *The American Economic Review*, pages 201–218, 1995. ISSN 0002-8282.

Richard Meese and Kenneth Rogoff. *The out-of-sample failure of empirical exchange rate models: sampling error or misspecification?*, pages 67–112. University of Chicago Press, 1983a.

Richard A Meese and Kenneth Rogoff. Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of international economics*, 14(1-2):3–24, 1983b. ISSN 0022-1996.

Tanya Molodtsova and David H Papell. Out-of-sample exchange rate predictability with taylor rule fundamentals. *Journal of international economics*, 77(2):167–180, 2009. ISSN 0022-1996.

Bishwa Nath Mukherjee and Sadhan Samar Maiti. On some properties of positive definite toeplitz matrices and their possible applications. *Linear algebra and its applications*, 102: 211–240, 1988. ISSN 0024-3795.

Karl Pearson. Principal components analysis. *The London, Edinburgh and Dublin Philosophical Magazine and Journal*, 6(2):566, 1901.

Niraj Poudyal. *StVAR: Student's t Vector Autoregression (StVAR)*, 2017. R package version 1.1.

Barbara Rossi. Exchange rate predictability. *Journal of economic literature*, 51(4):1063–1119, 2013. ISSN 0022-0515.

Minoru Siotani. Modern multivariate statistical analysis; a graduate course and handbook. 1985.

Aris Spanos. On modeling heteroskedasticity: the student's t and elliptical linear regression models. *Econometric Theory*, 10(02):286–315, 1994. ISSN 1469-4360.

Aris Spanos. *Probability theory and statistical inference: econometric modeling with observational data*. Cambridge University Press, 1999. ISBN 0521424089.

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

James H Stock and Mark W Watson. Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179, 2002. ISSN 0162-1459.

Jaakko Särelä and Harri Valpola. Denoising source separation. *Journal of machine learning research*, 6(Mar):233–272, 2005.

Henri Theil. Applied economic forecasting. 1971.

Yi-chiuan Wang and Jyh-lin Wu. Fundamentals and exchange rate prediction revisited. *Journal of Money, Credit and Banking*, 47(8):1651–1671, 2015. ISSN 1538-4616.

Yao-Ting Zhang, Kai-Tai Fang, and Hanfeng Chen. Matrix elliptically contoured distributions. *Acta Math. Sinica*, 5:341–353, 1985.

Appendices

Appendix A

Monte Carlo Simulation

A.1 The Normal VAR Detailed Forecasting Results

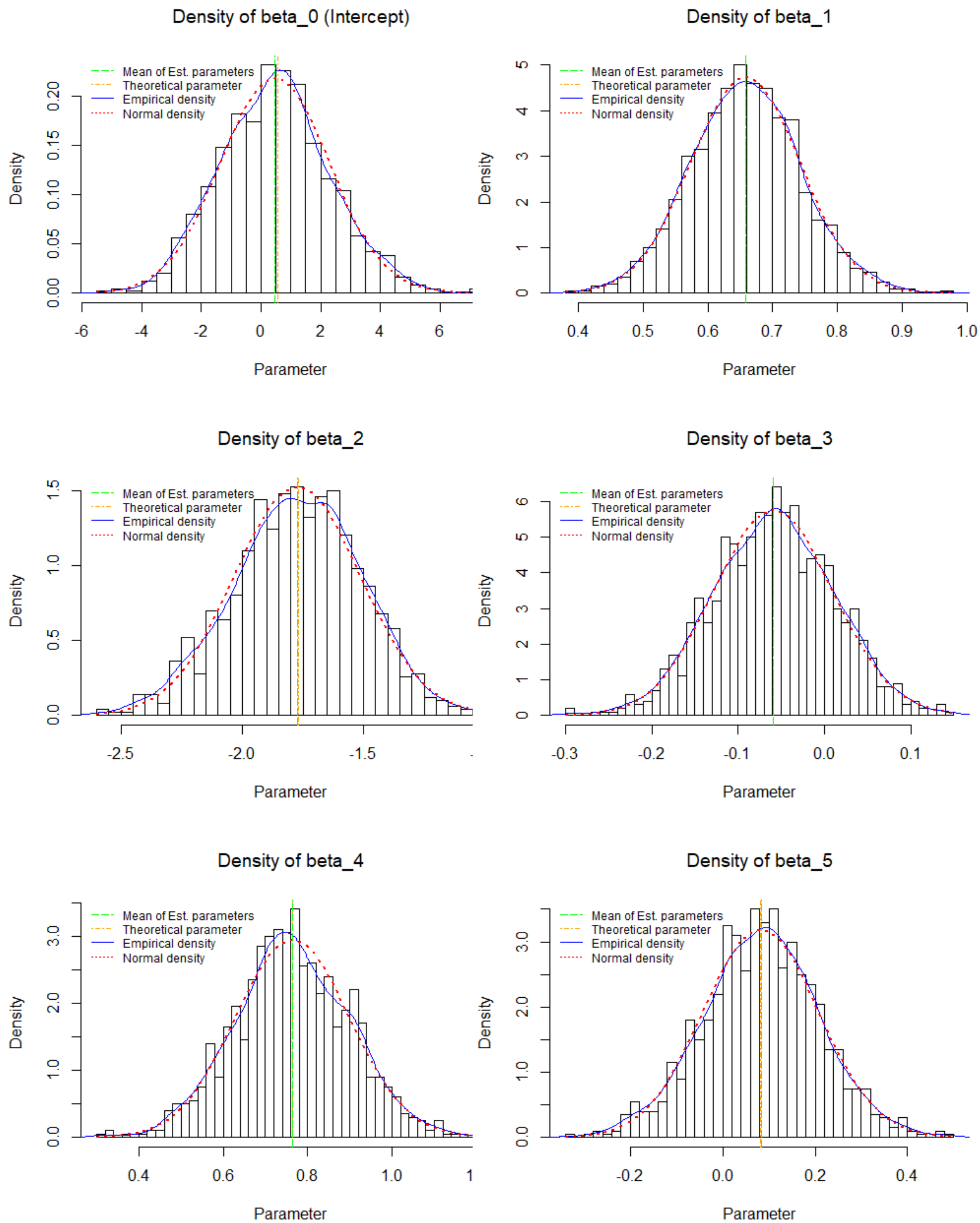
Table A.1: Individual Forecast Evaluation for GPCA

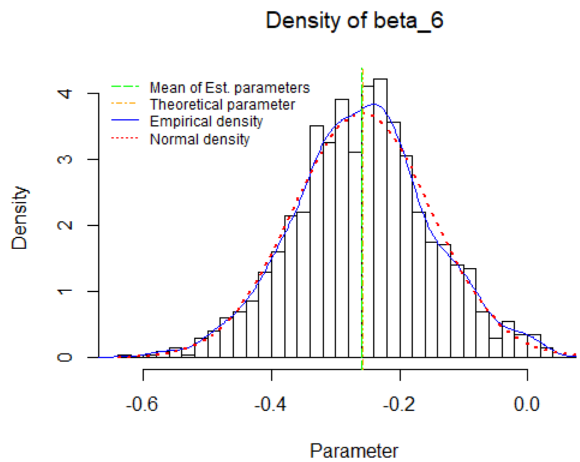
	Theil's U-statistic				Clack and West t-test			
Variables	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
$\mathbf{X}_1$	0.731	0.647	0.768	0.720	3.014	2.383	1.291	1.310
$\mathbf{X}_2$	0.710	0.633	0.764	0.679	2.972	2.684	1.189	2.033
$\mathbf{X}_3$	0.683	0.693	0.678	0.731	3.442	1.646	1.893	1.377
$\mathbf{X}_4$	0.693	0.733	0.714	0.728	3.456	1.488	1.547	1.274
$\mathbf{X}_5$	0.735	0.668	0.713	0.736	3.035	2.097	1.735	1.314
$\mathbf{X}_6$	0.752	0.659	0.719	0.742	2.648	2.392	2.013	1.359
$\mathbf{X}_7$	0.671	0.685	0.720	0.746	3.191	1.826	1.571	1.351
$\mathbf{X}_8$	0.661	0.691	0.695	0.736	3.163	1.676	1.949	1.431
$\mathbf{X}_9$	0.671	0.691	0.669	0.743	3.298	1.661	2.019	1.240
$\mathbf{X}_{10}$	0.697	0.685	0.685	0.741	3.716	1.906	1.904	1.406
$\mathbf{X}_{11}$	0.662	0.693	0.696	0.736	3.133	1.634	1.919	1.418
$\mathbf{X}_{12}$	0.731	0.655	0.720	0.734	2.994	2.675	2.107	1.438
$\mathbf{X}_{13}$	0.726	0.655	0.736	0.726	2.847	2.621	1.794	1.284
$\mathbf{X}_{14}$	0.726	0.686	0.693	0.731	2.615	2.085	2.002	1.290
$\mathbf{X}_{15}$	0.660	0.694	0.687	0.735	3.201	1.598	2.024	1.440

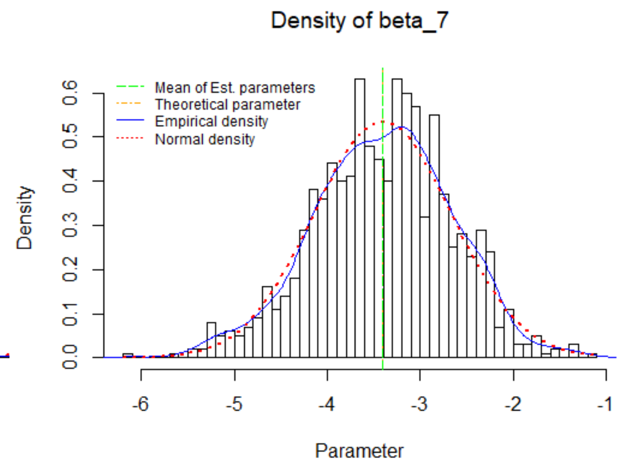
Table A.2: Individual Forecast Evaluation for PCA

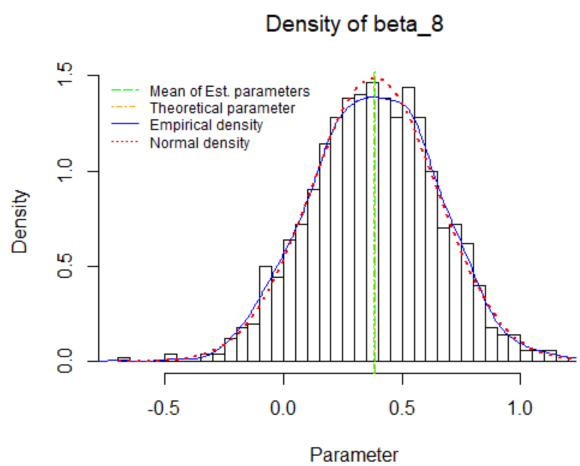
	Theil's U-statistic				Clack and West t-test			
Variables	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
$\mathbf{X}_1$	0.979	0.978	1.003	0.979	1.162	0.689	-0.045	0.246
$\mathbf{X}_2$	0.952	0.958	0.997	0.924	1.686	0.797	0.030	0.645
$\mathbf{X}_3$	0.990	0.996	0.998	0.992	0.985	0.199	0.078	0.234
$\mathbf{X}_4$	0.930	0.963	0.962	0.946	2.152	0.575	0.375	0.461
$\mathbf{X}_5$	0.987	0.993	1.001	1.002	0.846	0.239	-0.014	-0.025
$\mathbf{X}_6$	1.002	1.004	0.981	1.021	-0.172	-0.161	0.507	-0.382
$\mathbf{X}_7$	0.983	0.984	0.989	0.980	2.118	0.803	0.349	0.393
$\mathbf{X}_8$	1.012	1.010	1.009	1.010	-2.362	-0.728	-0.547	-0.345
$\mathbf{X}_9$	0.998	0.999	0.994	1.008	0.326	0.048	0.274	-0.274
$\mathbf{X}_{10}$	0.995	0.998	1.008	1.005	0.546	0.091	-0.196	-0.080
$\mathbf{X}_{11}$	1.013	1.009	1.012	1.011	-2.546	-0.605	-0.645	-0.366
$\mathbf{X}_{12}$	1.003	1.000	0.985	0.997	-0.345	-0.010	0.585	0.082
$\mathbf{X}_{13}$	0.999	0.998	0.990	0.985	0.086	0.130	0.355	0.404
$\mathbf{X}_{14}$	0.990	0.990	0.972	0.978	0.559	0.256	0.444	0.215
$\mathbf{X}_{15}$	1.011	1.010	1.008	1.009	-2.006	-0.734	-0.471	-0.350

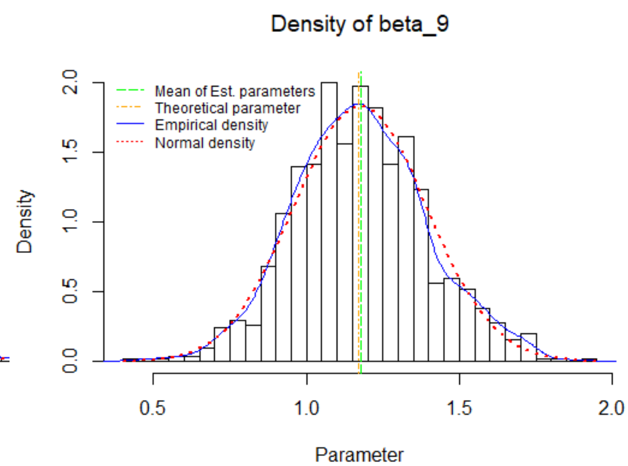
A.2 Histograms Of Estimated Coefficients (Normal VAR)

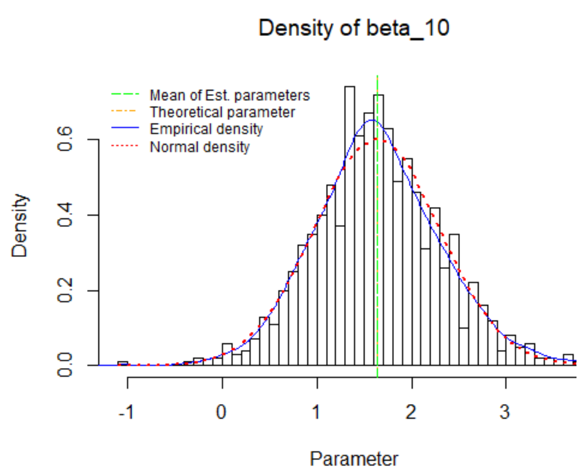


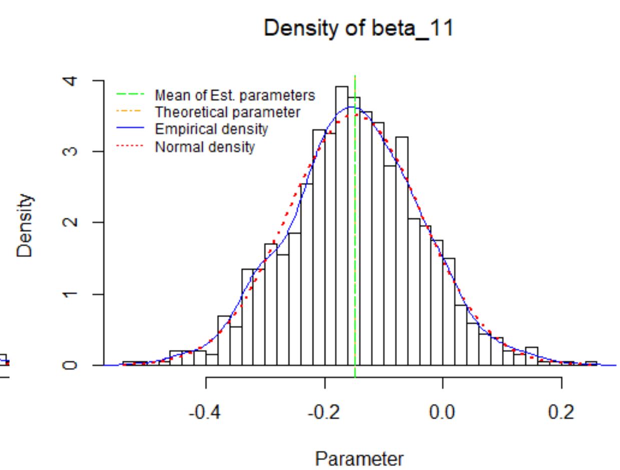

Density of beta_6
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
0
1
2
3
4
-0.6
-0.4
-0.2
0.0
Parameter

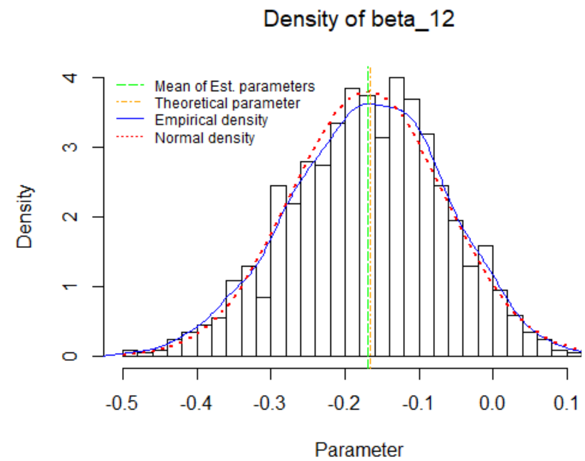

Density of beta_7
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
-6
-5
-4
-3
-2
-1
Parameter

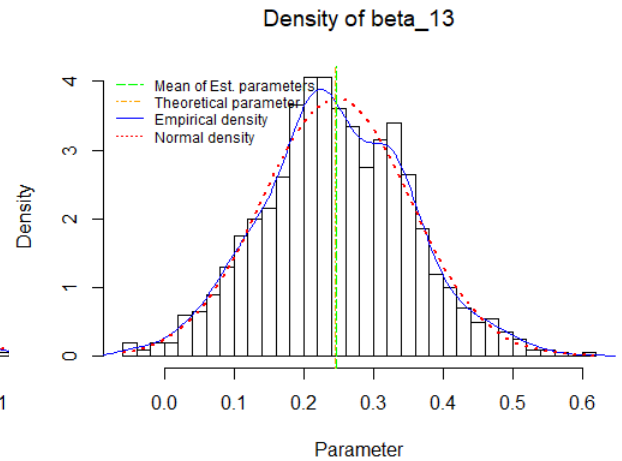

Density of beta_8
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
0.0
0.5
1.0
1.5
-0.5
0.0
0.5
1.0
Parameter

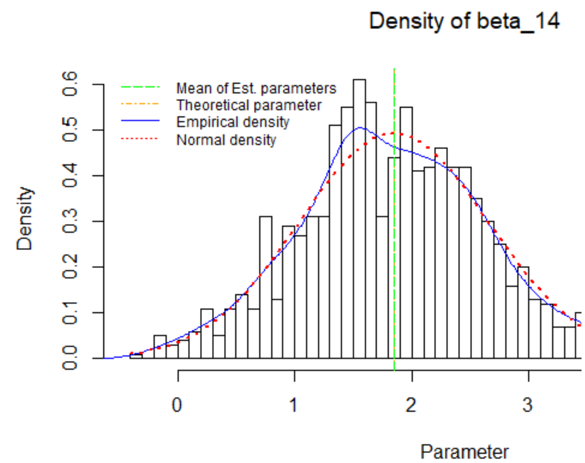

Density of beta_9
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
0.0
0.5
1.0
1.5
2.0
0.5
1.0
1.5
2.0
Parameter

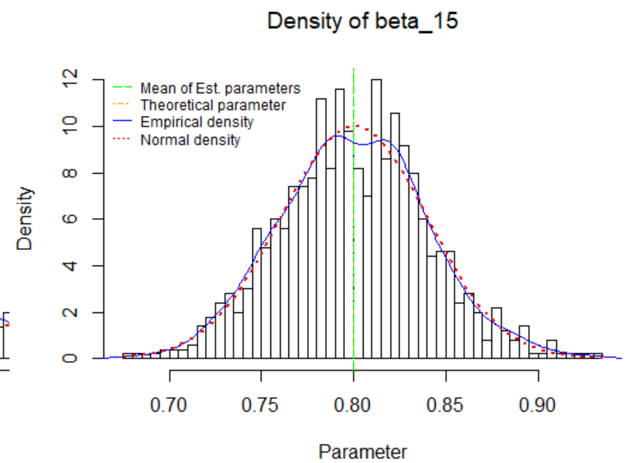

Density of beta_10
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
0.0
0.2
0.4
0.6
-1
0
1
2
3
Parameter

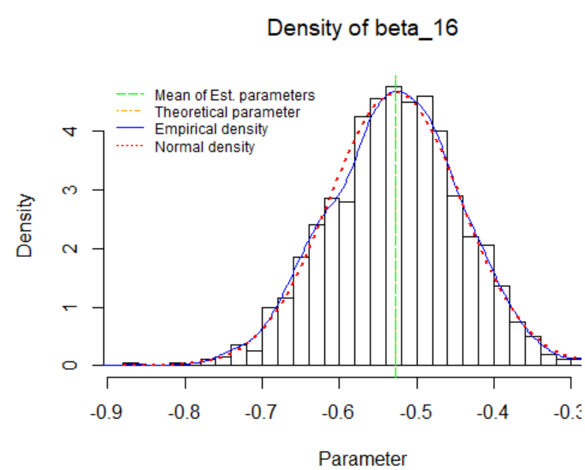

Density of beta_11
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
0
1
2
3
4
-0.4
-0.2
0.0
0.2
Parameter

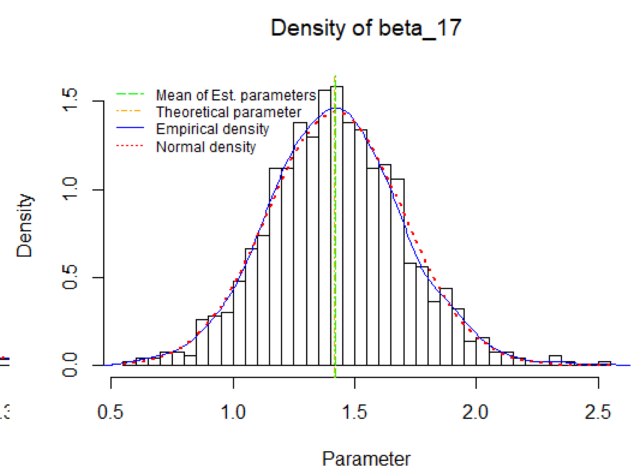

Density of beta_12
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
0
1
2
3
4
-0.5
-0.4
-0.3
-0.2
-0.1
0.0
0.1
Parameter

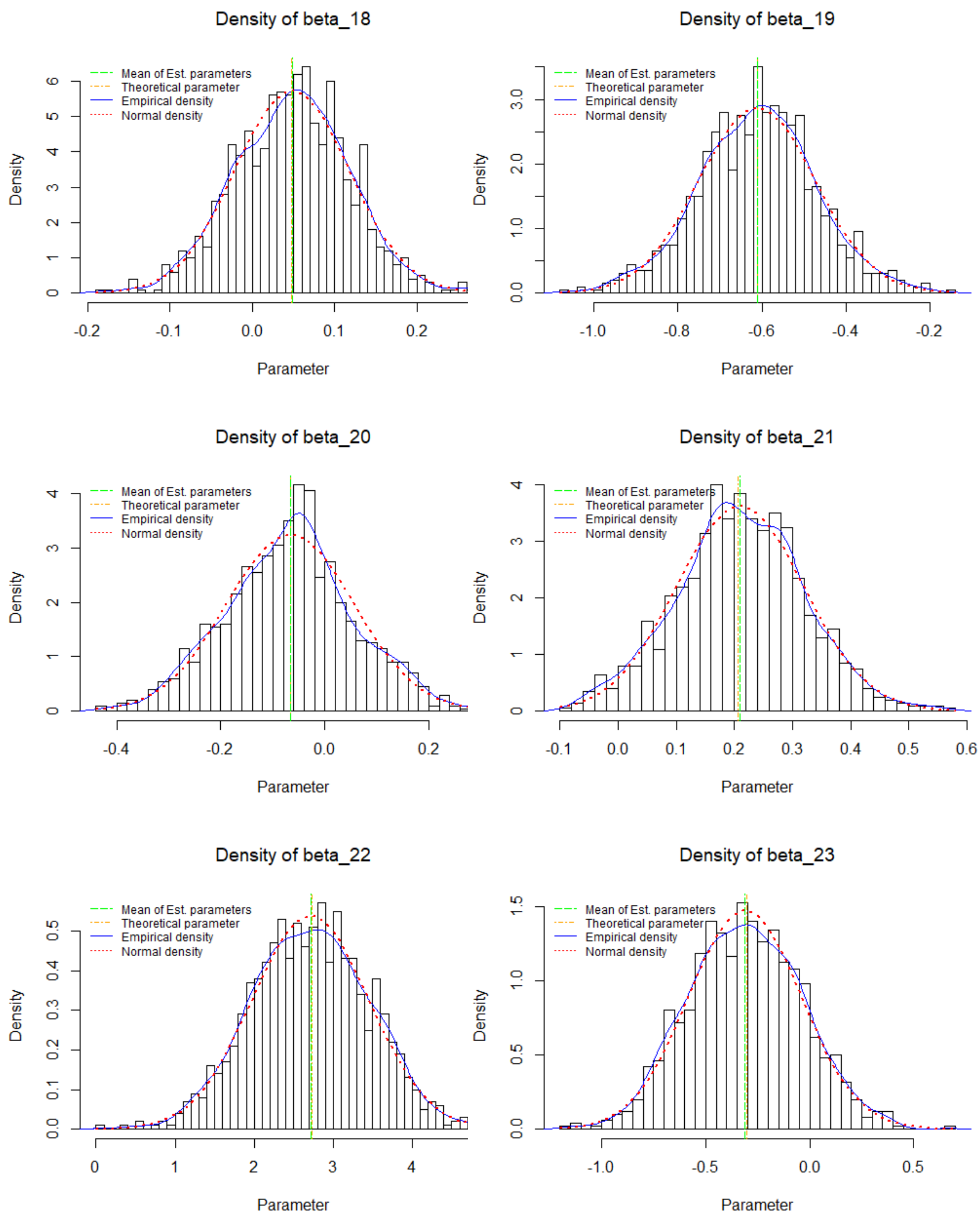

Density of beta_13
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
0
1
2
3
4
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Parameter

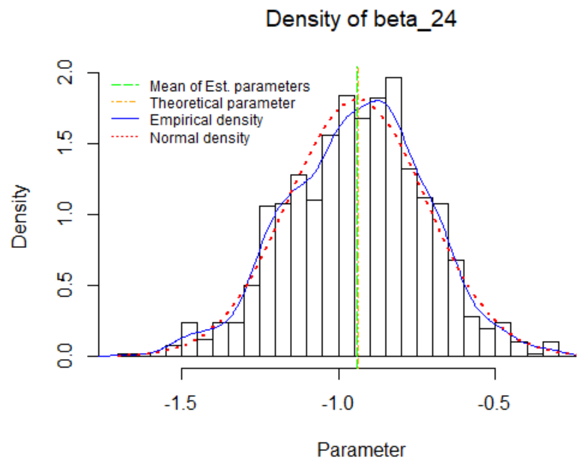

Density of beta_14
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
1
2
3
Parameter

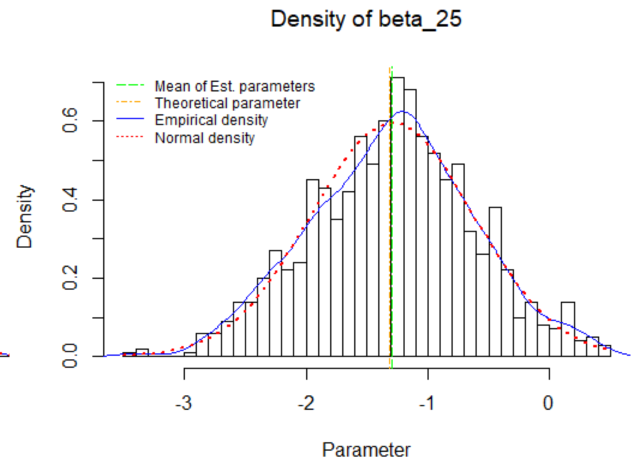

Density of beta_15
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
0
2
4
6
8
10
12
0.70
0.75
0.80
0.85
0.90
Parameter

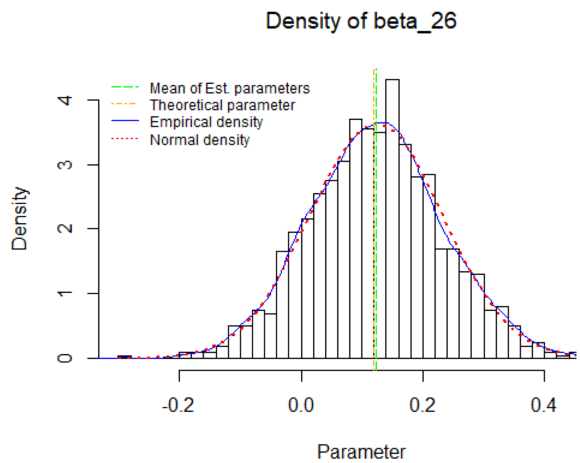

Density of beta_16
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
0
1
2
3
4
-0.9
-0.8
-0.7
-0.6
-0.5
-0.4
Parameter

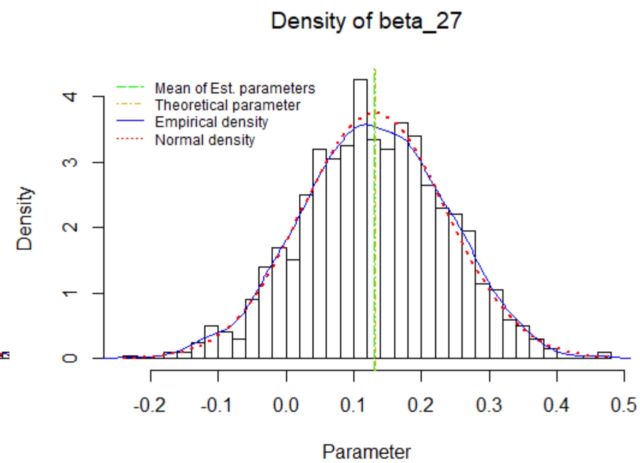

Density of beta_17
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
0.0
0.5
1.0
1.5
0.5
1.0
1.5
2.0
2.5
Parameter

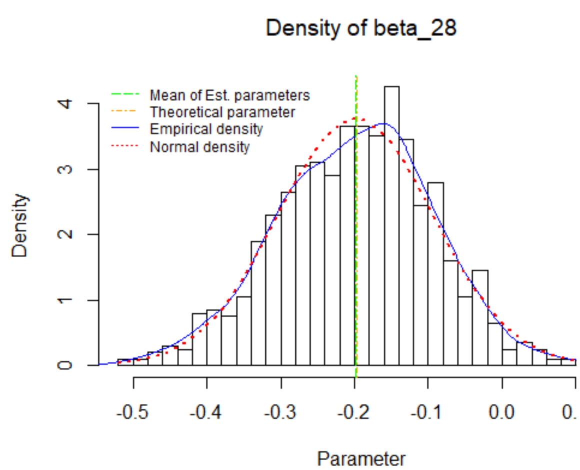

Density of beta_18
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
Parameter
Density of beta_19
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
Parameter
Density of beta_20
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
Parameter
Density of beta_21
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
Parameter
Density of beta_22
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
Parameter
Density of beta_23
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
Parameter

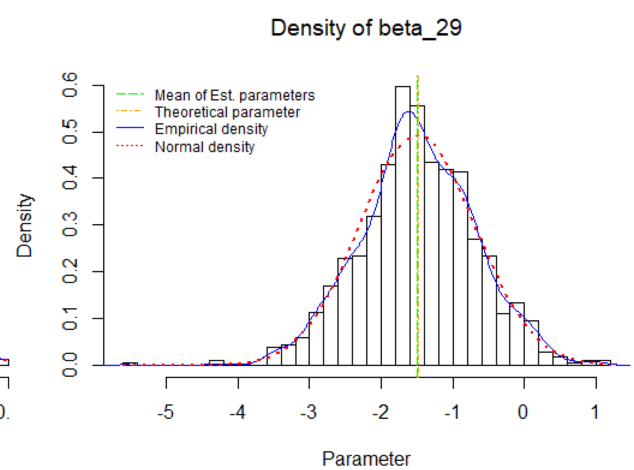

Density of beta_24
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
Parameter


Density of beta_25
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
Parameter


Density of beta_26
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
Parameter

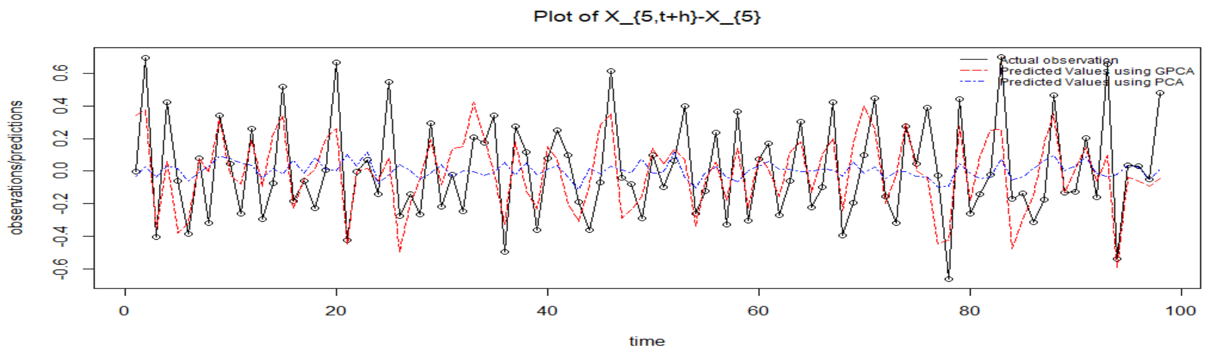

Density of beta_27
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
Parameter

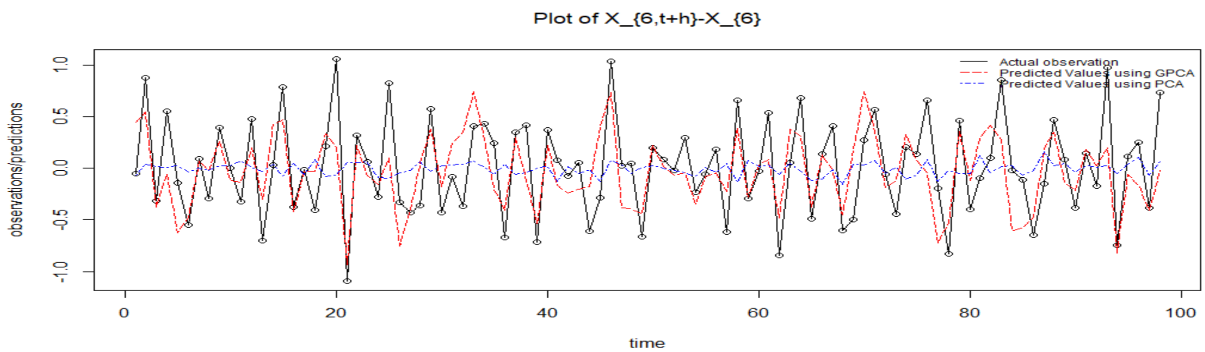

Density of beta_28
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
Parameter

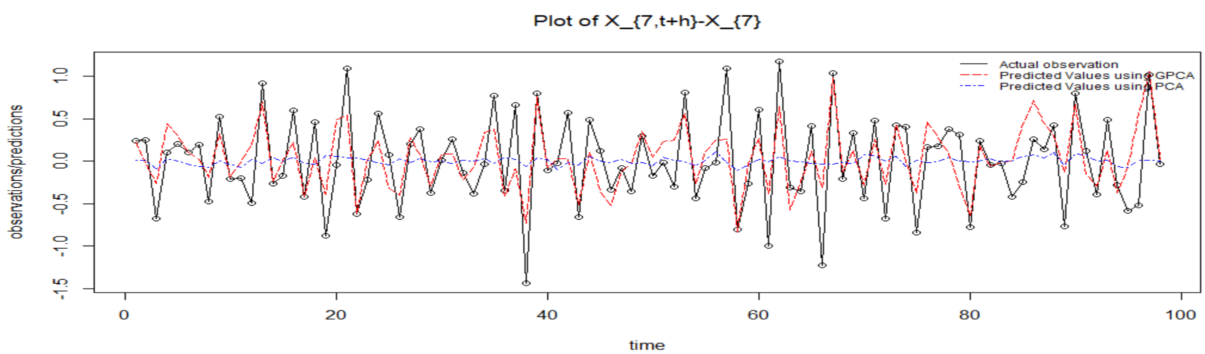

Density of beta_29
Mean of Est. parameters
Theoretical parameter
Empirical density
Normal density
Density
Parameter

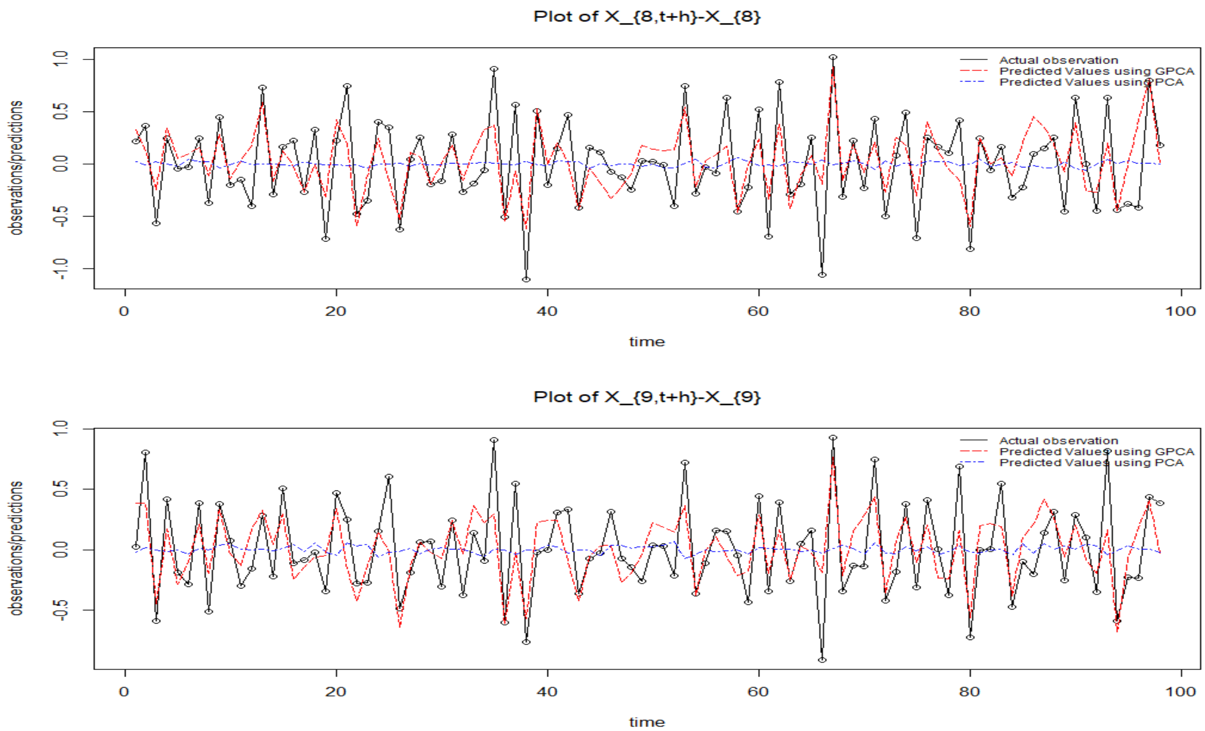
A.3 The Normal VAR: Predictions vs. Actual Observations Plots

A.3.1 Horizon $h = 1$

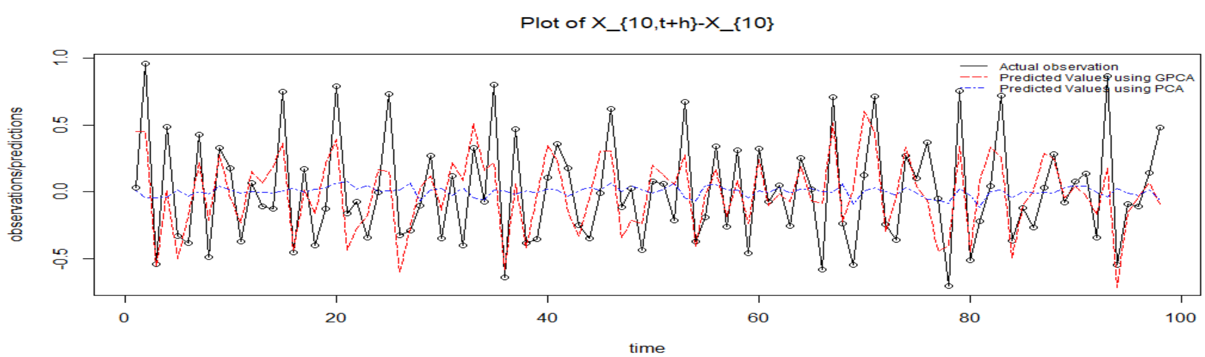




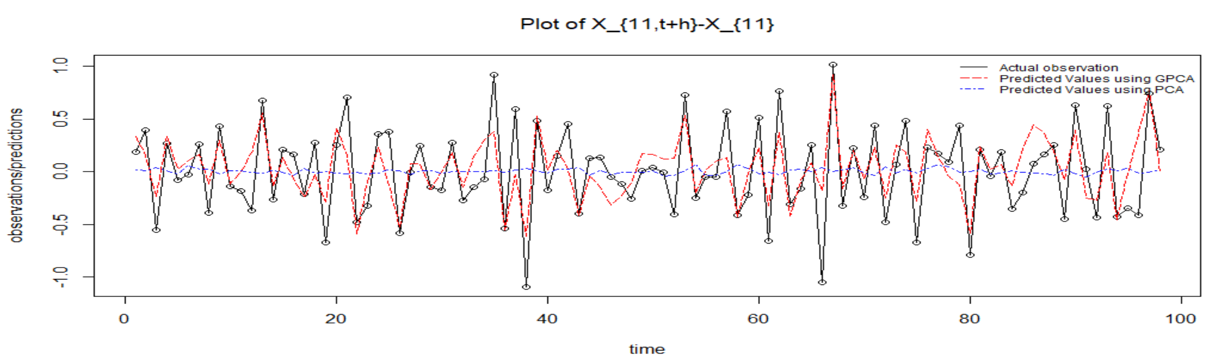




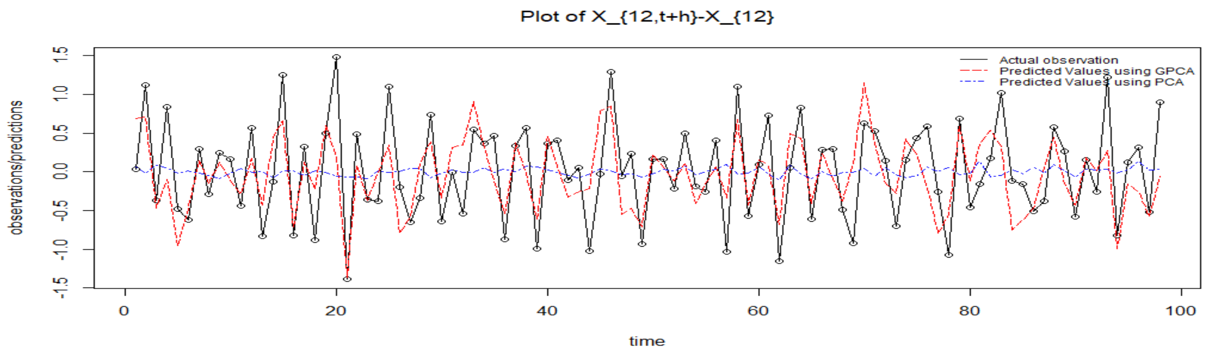
(a) Predicted values vs Actual observation ($h = 1$)



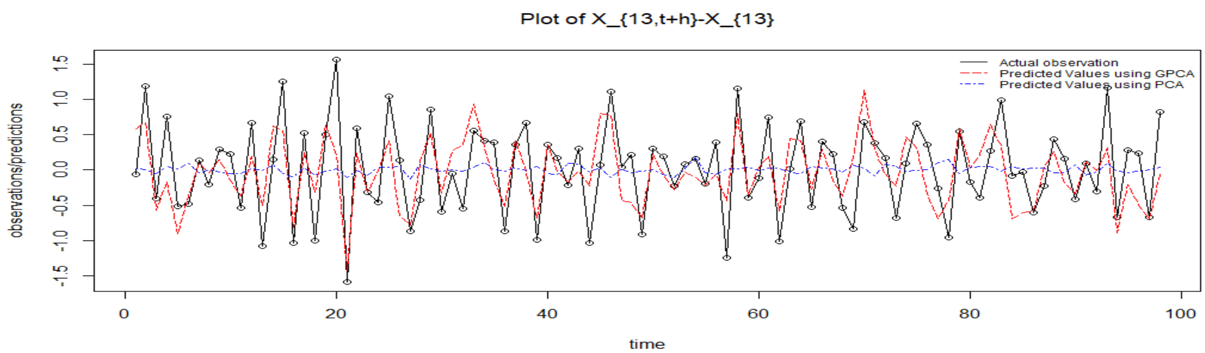
(b) Predicted values vs Actual observation ($h = 1$)



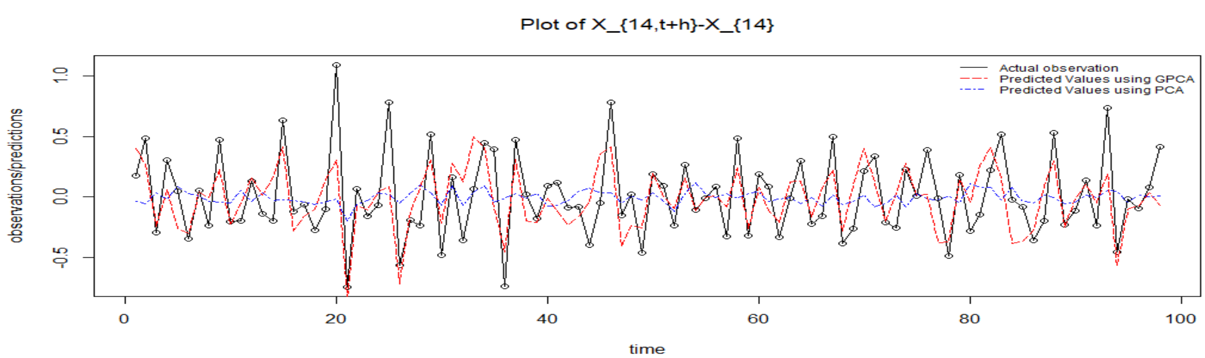
(c) Predicted values vs Actual observation ($h = 1$)



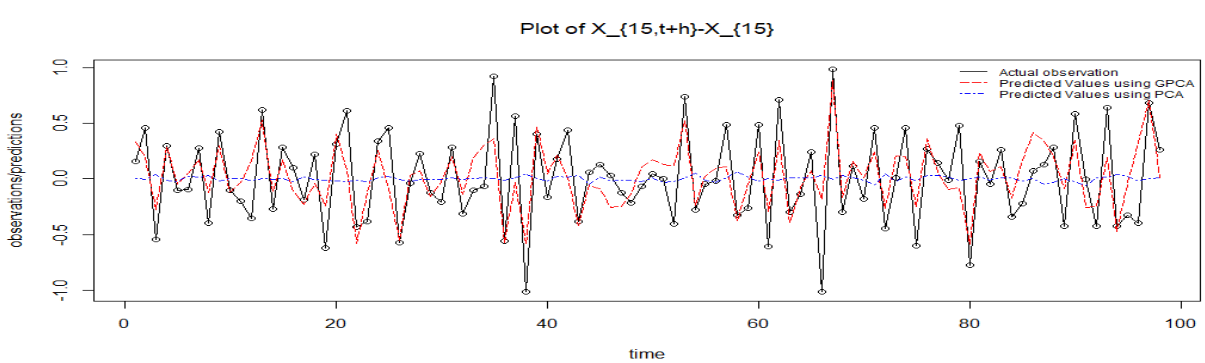
(a) Predicted values vs Actual observation ($h = 1$)



(b) Predicted values vs Actual observation ($h = 1$)

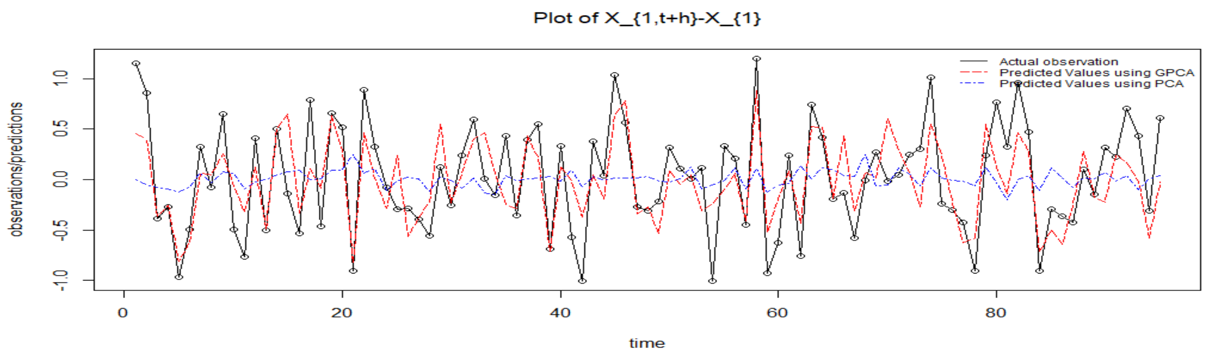


(c) Predicted values vs Actual observation ($h = 1$)

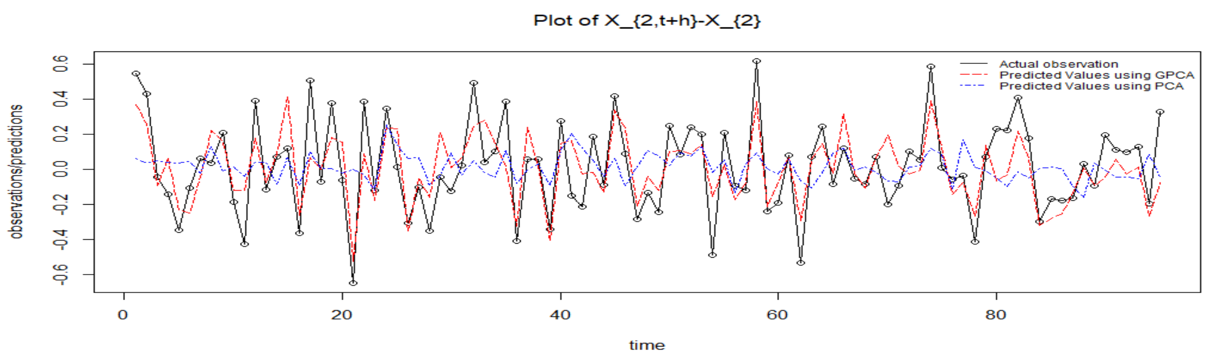


(d) Predicted values vs Actual observation ($h = 1$)

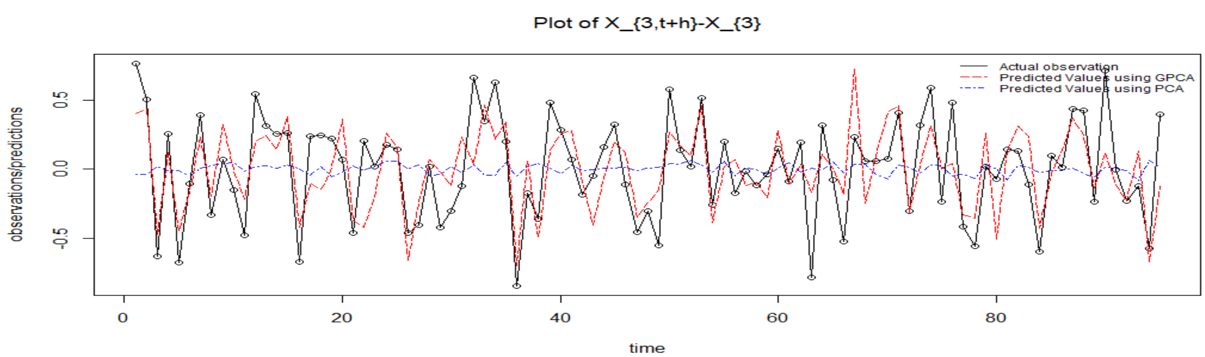
A.3.2 Horizon $h = 4$



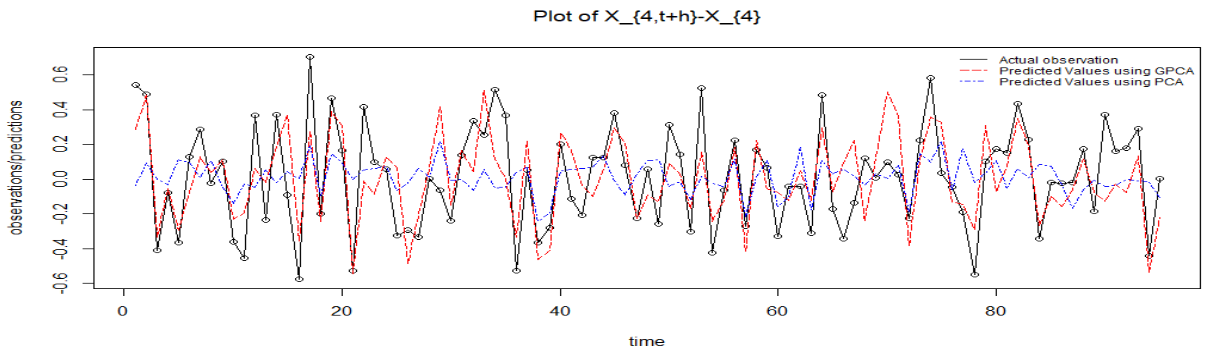
(a) Predicted values vs Actual observation ($h = 4$)



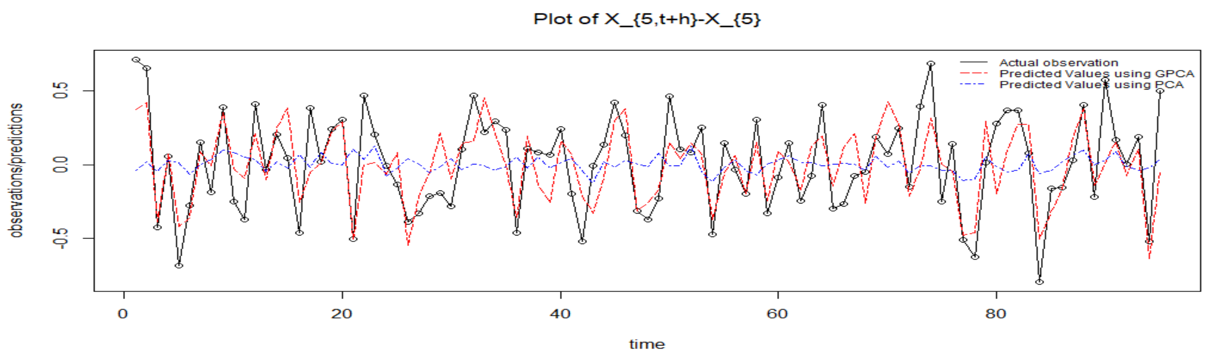
(b) Predicted values vs Actual observation ($h = 4$)



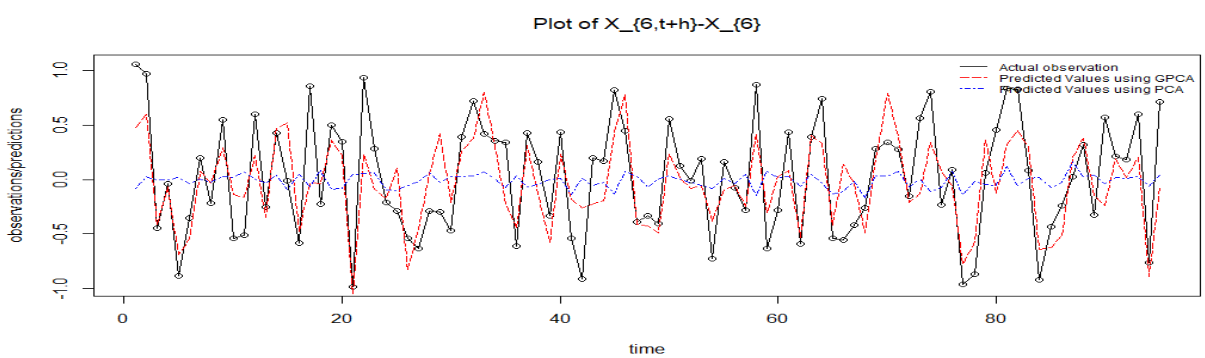
(c) Predicted values vs Actual observation ($h = 4$)



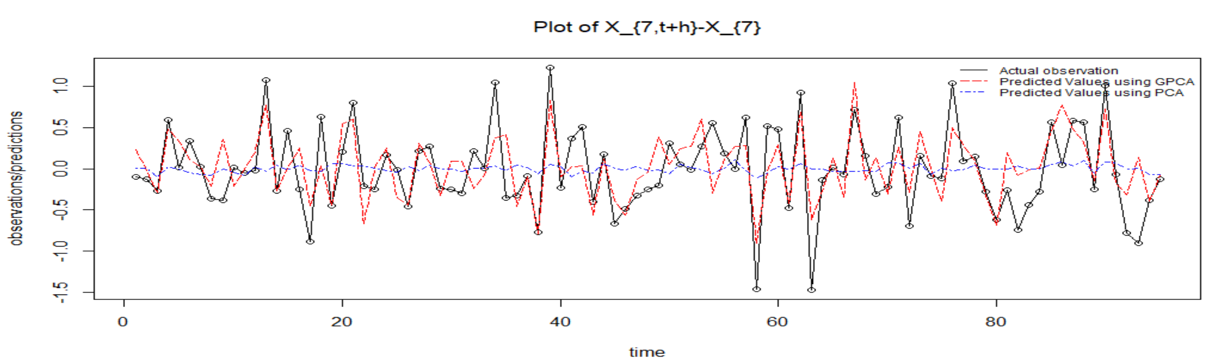
(a) Predicted values vs Actual observation ($h = 4$)



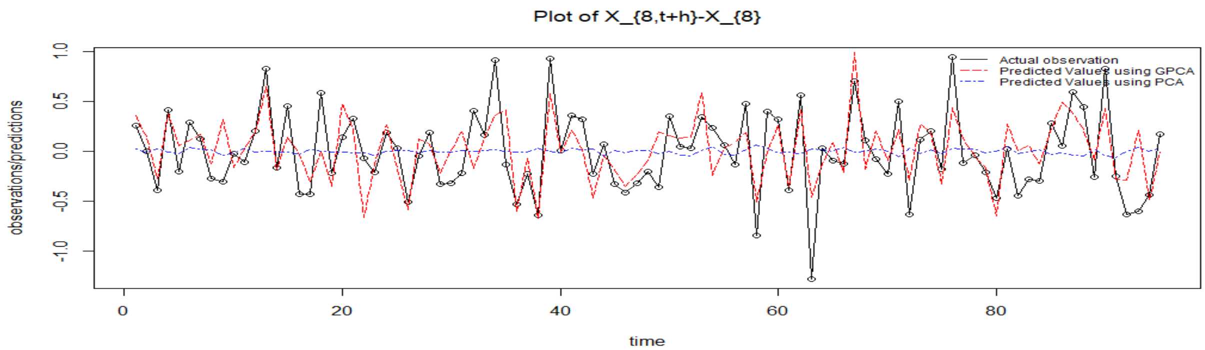
(b) Predicted values vs Actual observation ($h = 4$)



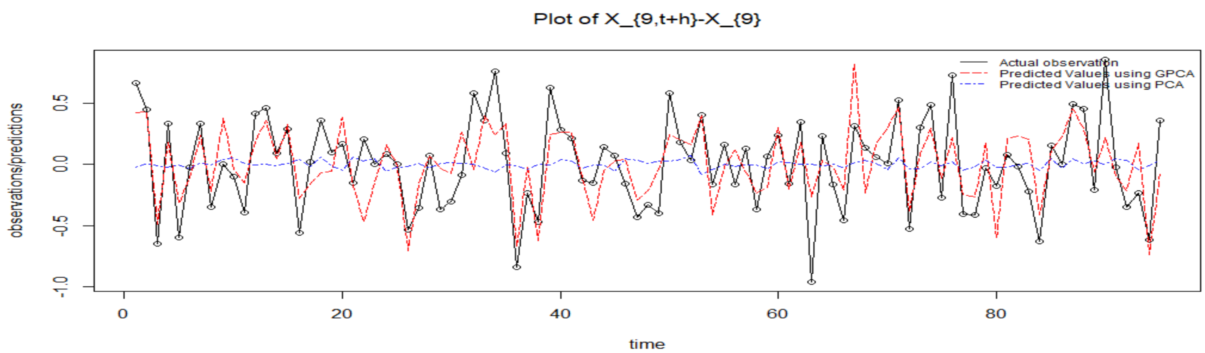
(c) Predicted values vs Actual observation ($h = 4$)



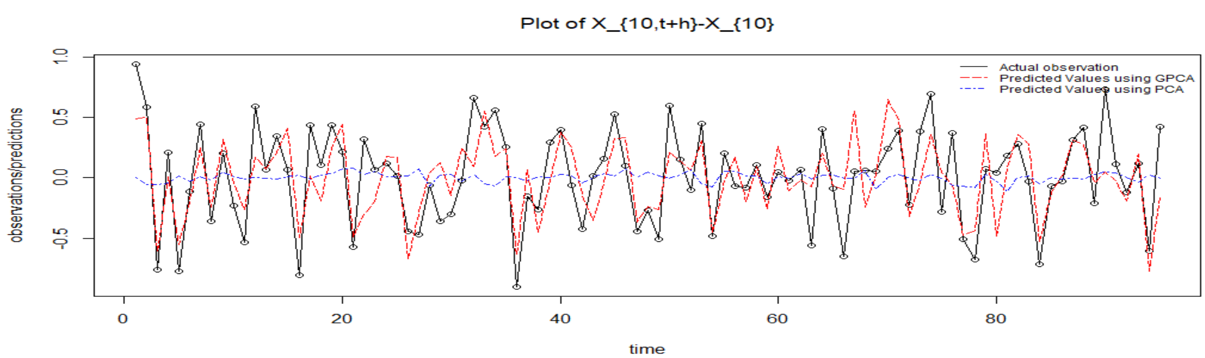
(d) Predicted values vs Actual observation ($h = 4$)



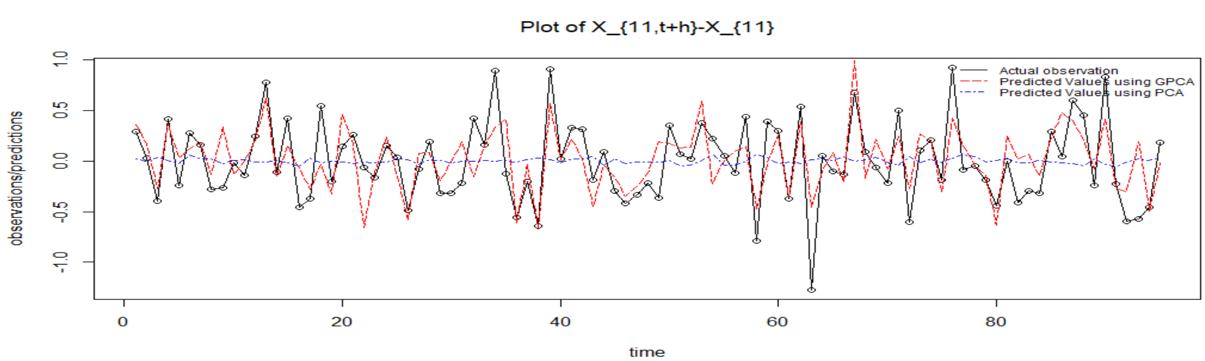
(a) Predicted values vs Actual observation ($h = 4$)



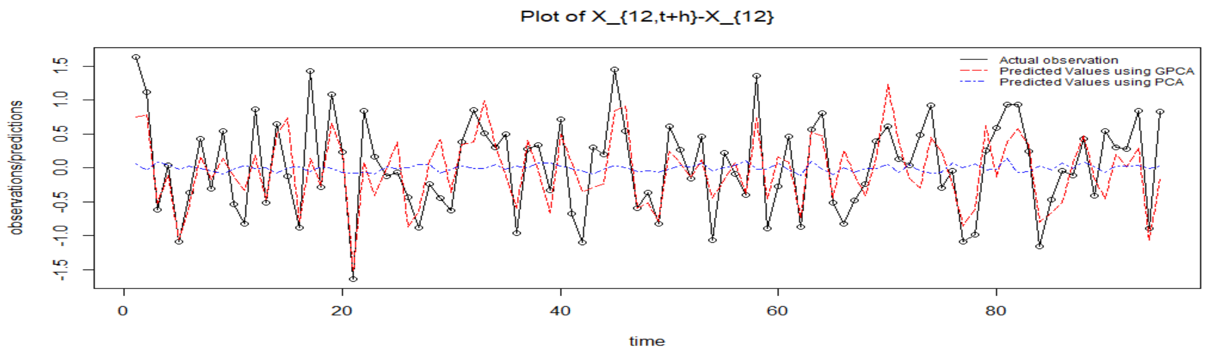
(b) Predicted values vs Actual observation ($h = 4$)



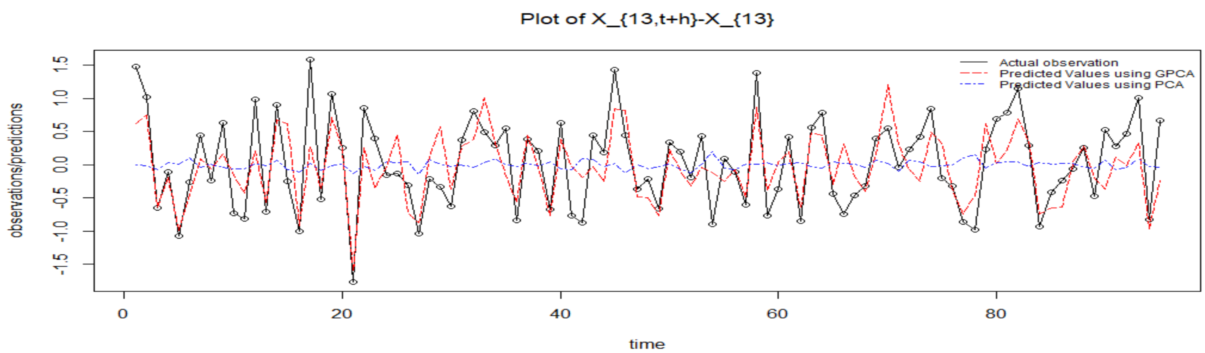
(c) Predicted values vs Actual observation ($h = 4$)



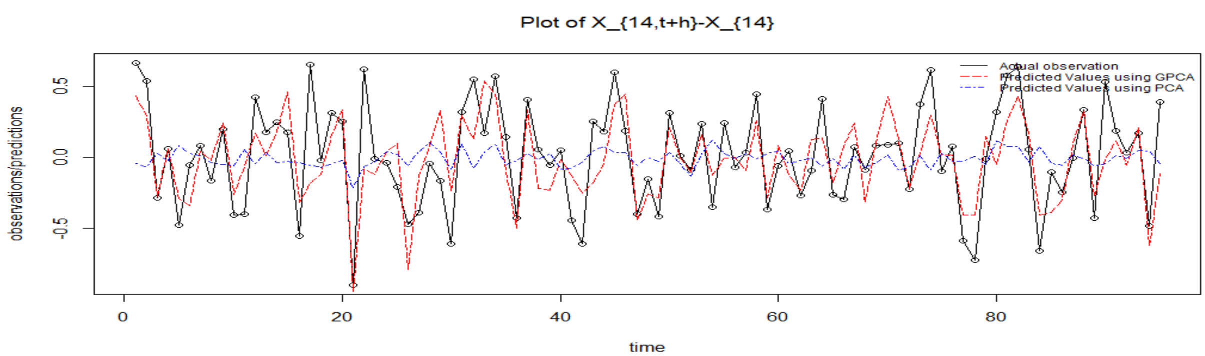
(d) Predicted values vs Actual observation ($h = 4$)



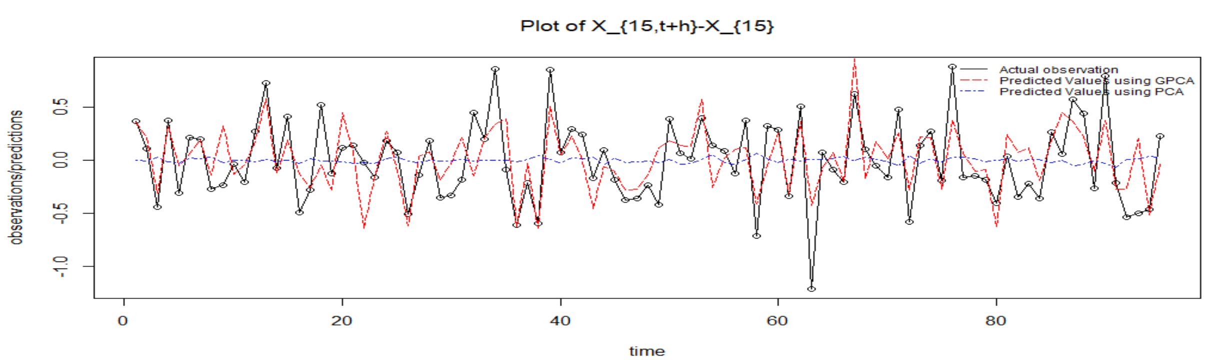
(a) Predicted values vs Actual observation ($h = 4$)



(b) Predicted values vs Actual observation ($h = 4$)

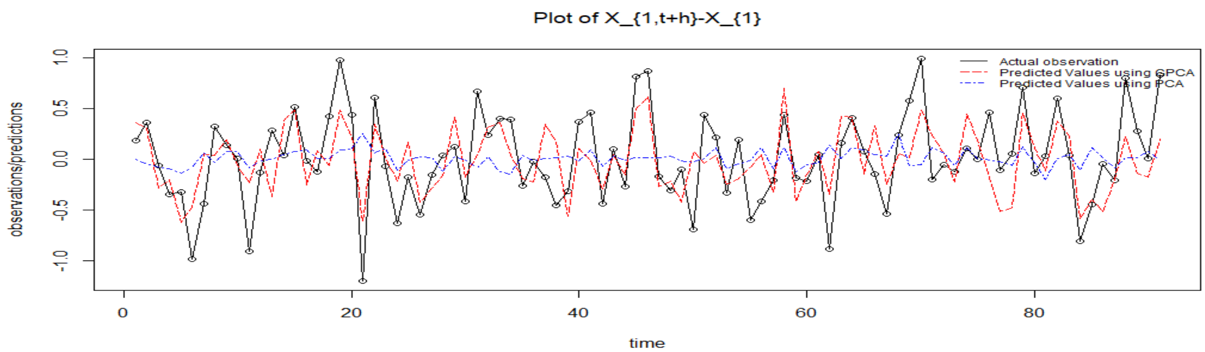


(c) Predicted values vs Actual observation ($h = 4$)

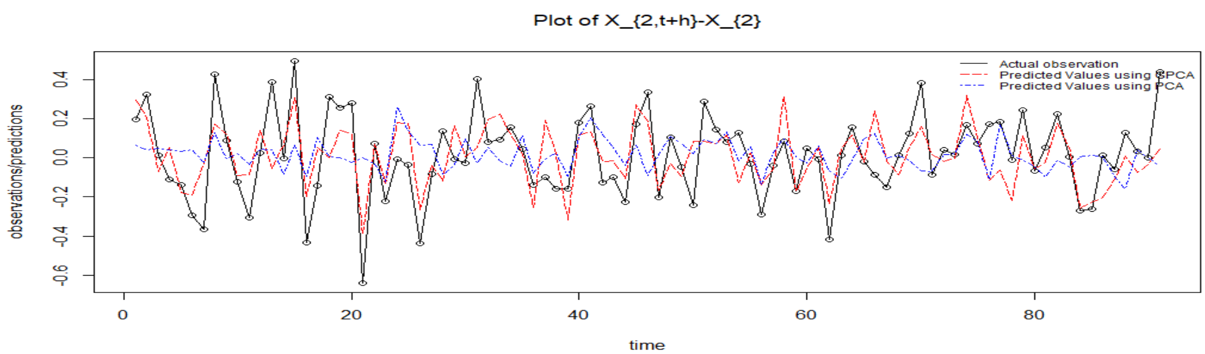


(d) Predicted values vs Actual observation ($h = 4$)

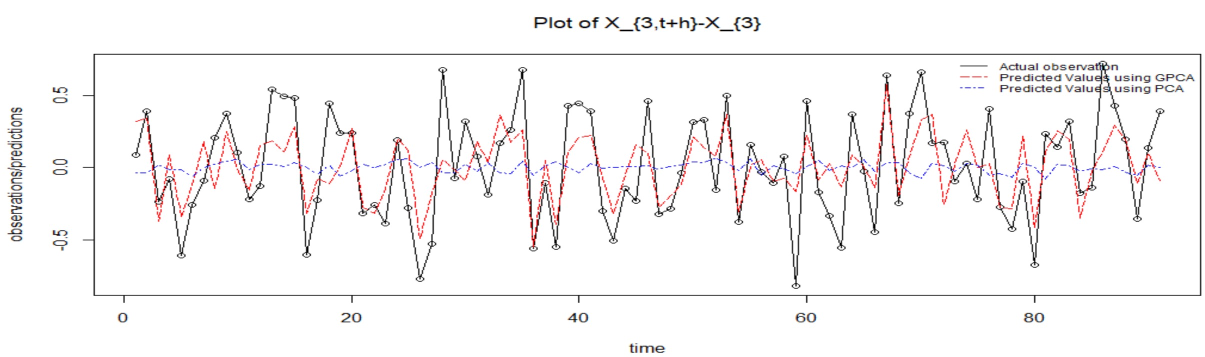
A.3.3 Horizon $h = 8$



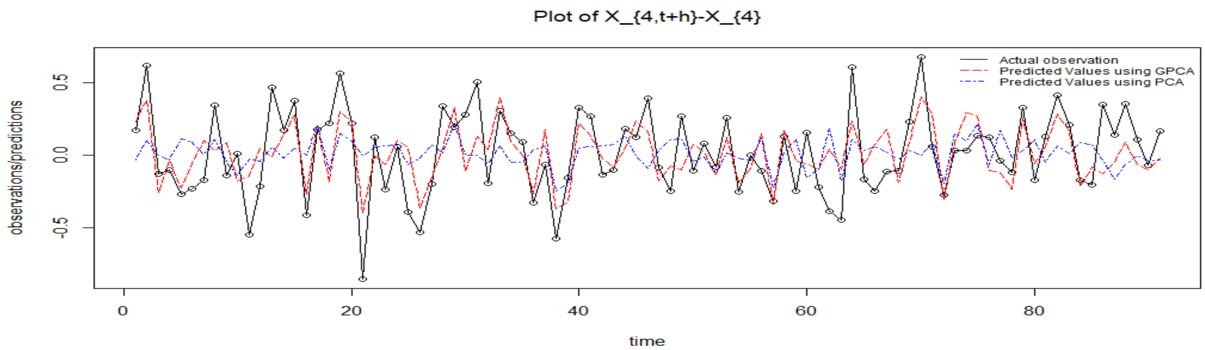
(a) Predicted values vs Actual observation ($h = 8$)



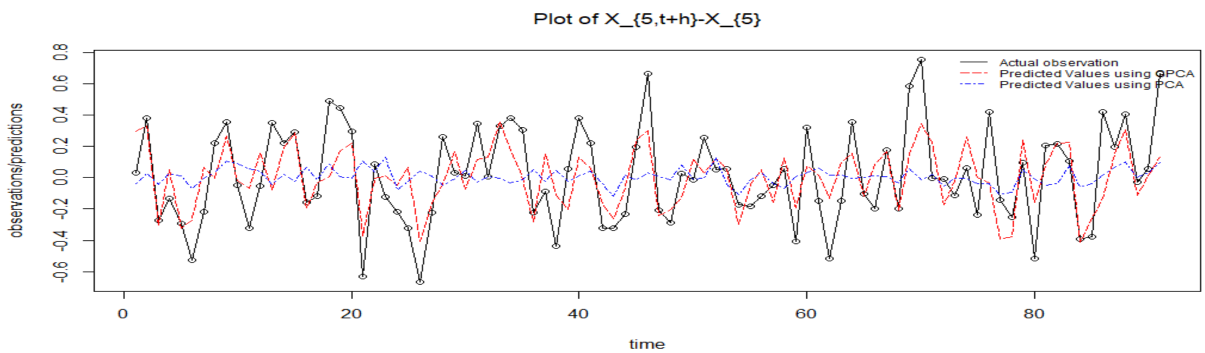
(b) Predicted values vs Actual observation ($h = 8$)



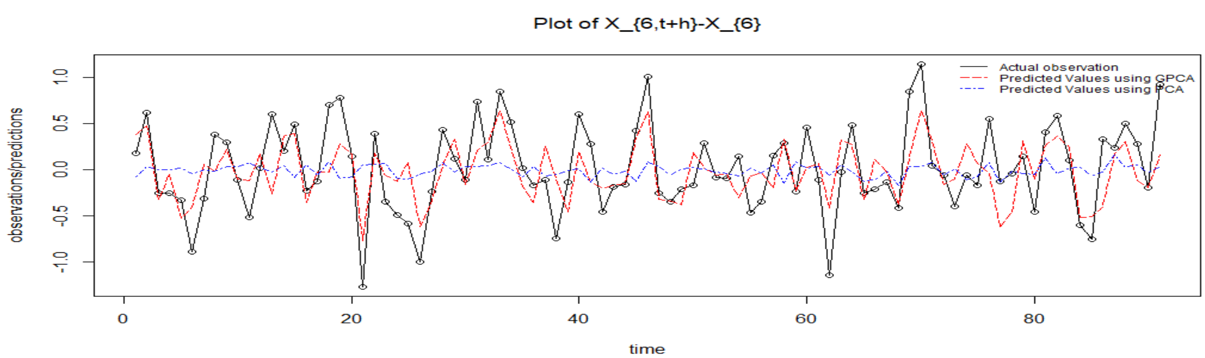
(c) Predicted values vs Actual observation ($h = 8$)



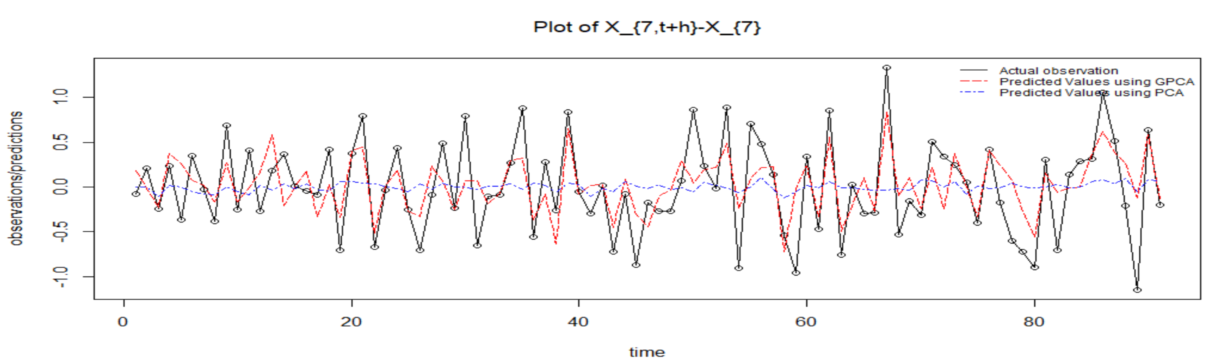
(a) Predicted values vs Actual observation ($h = 8$)



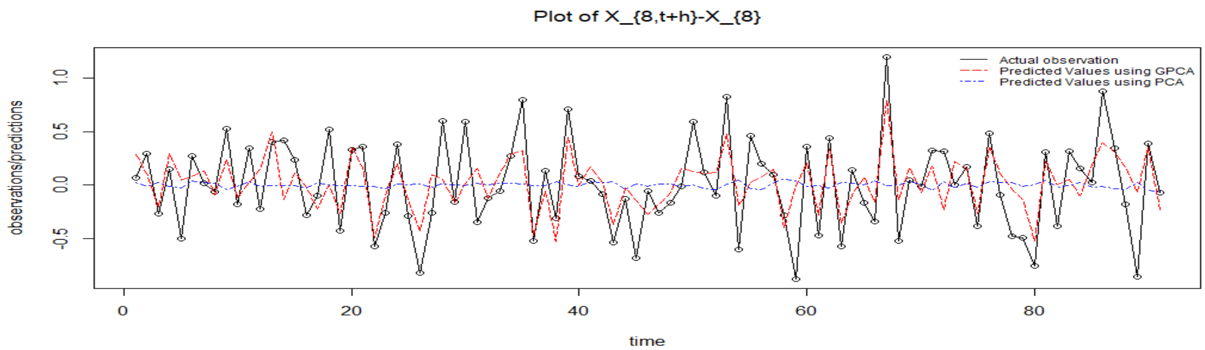
(b) Predicted values vs Actual observation ($h = 8$)



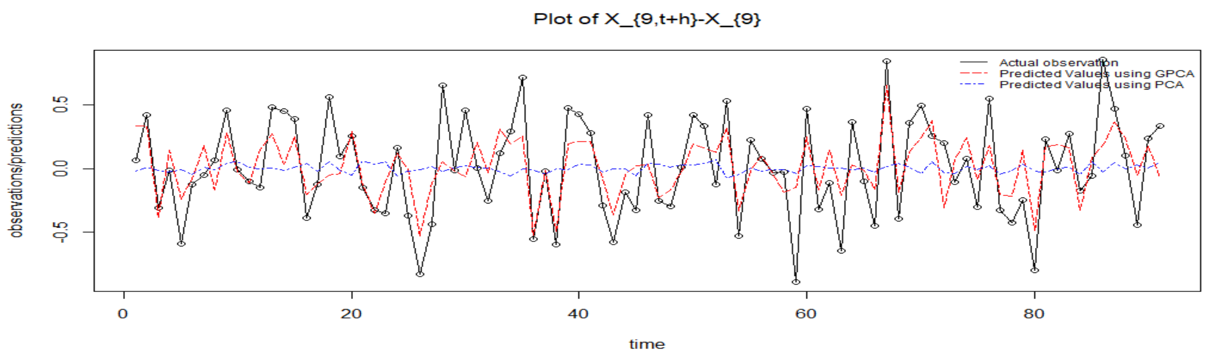
(c) Predicted values vs Actual observation ($h = 8$)



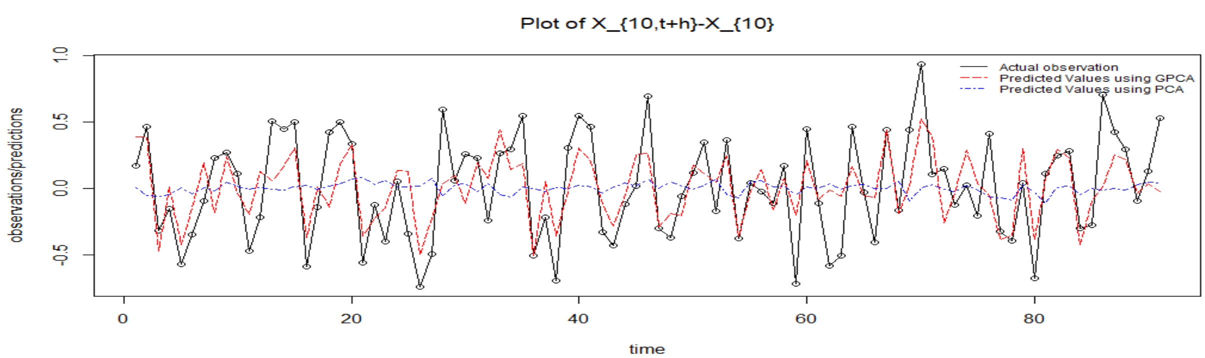
(d) Predicted values vs Actual observation ($h = 8$)



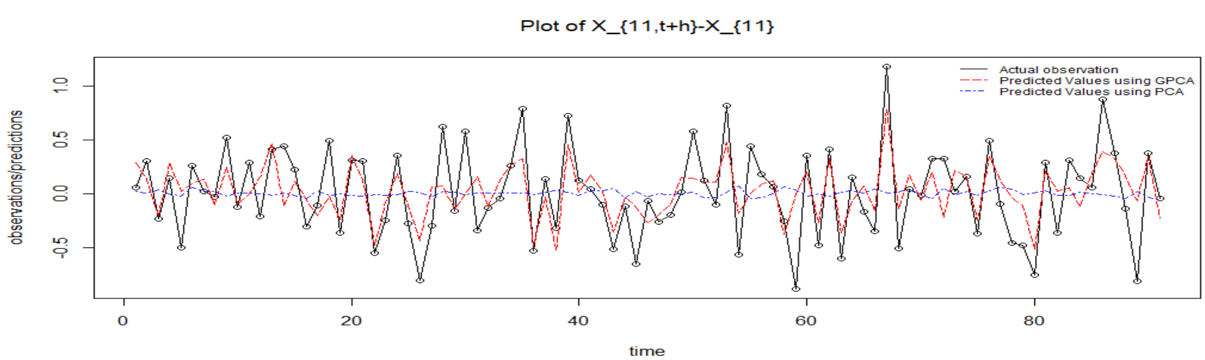
(a) Predicted values vs Actual observation ($h = 8$)



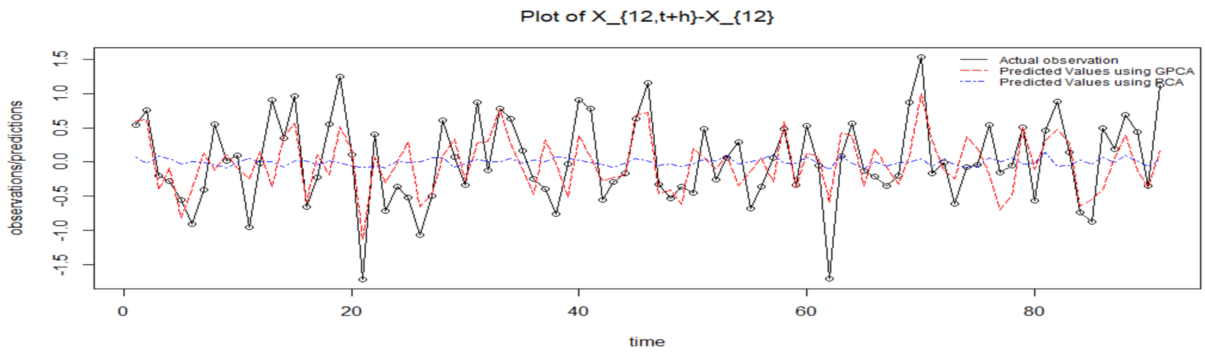
(b) Predicted values vs Actual observation ($h = 8$)



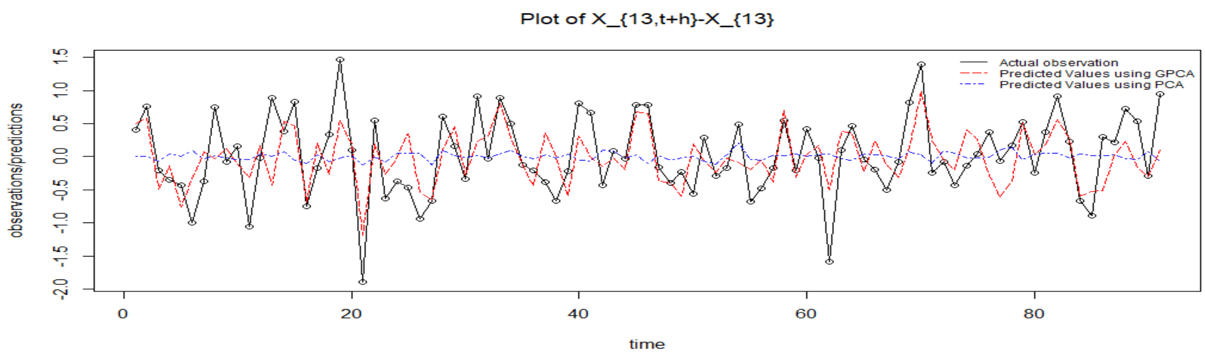
(c) Predicted values vs Actual observation ($h = 8$)



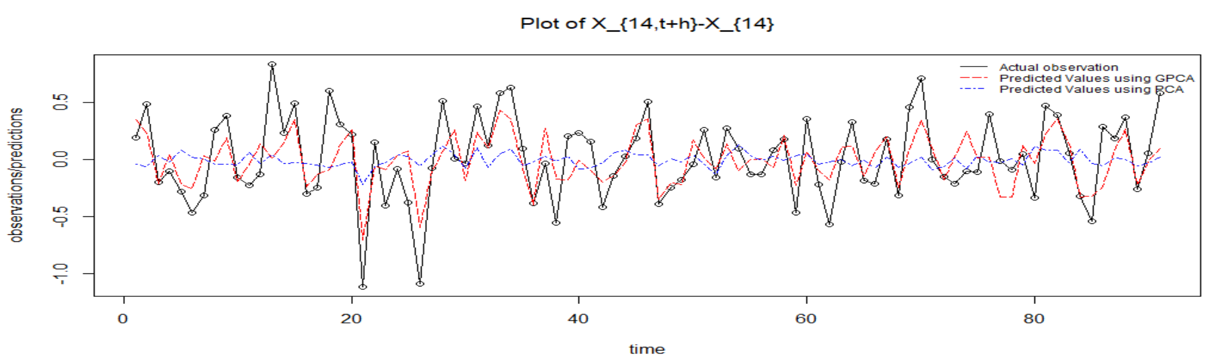
(d) Predicted values vs Actual observation ($h = 8$)



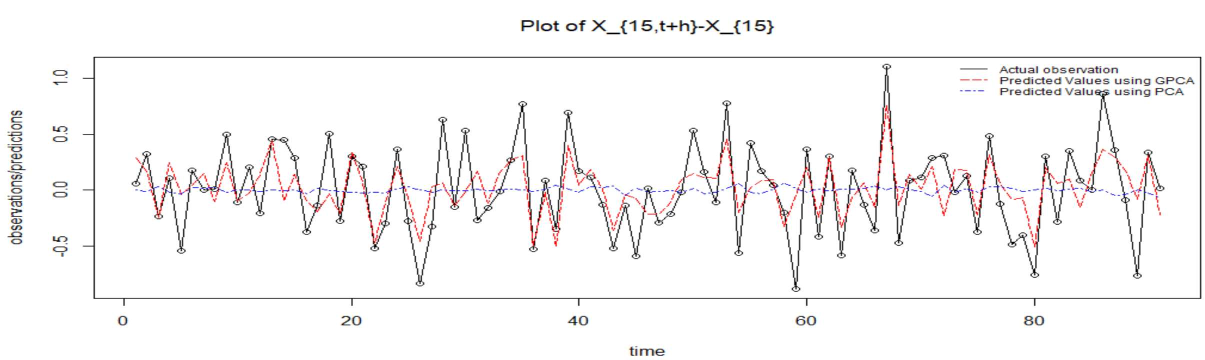
(a) Predicted values vs Actual observation ($h = 8$)



(b) Predicted values vs Actual observation ($h = 8$)

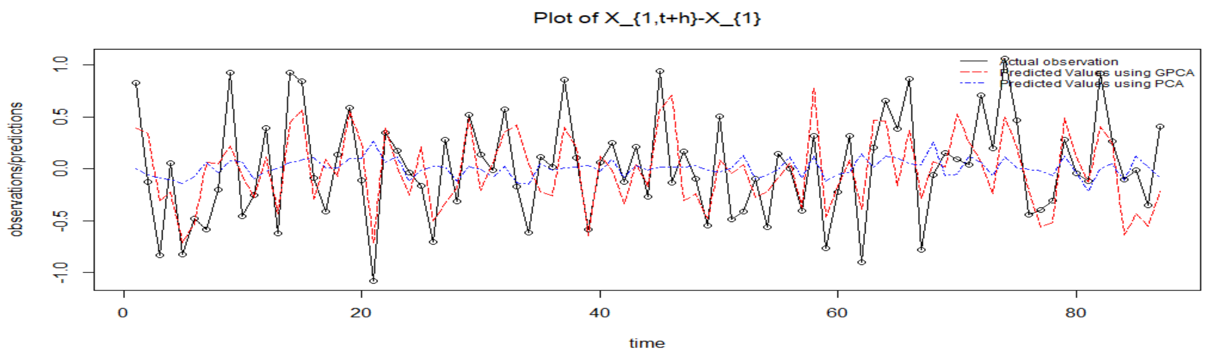


(c) Predicted values vs Actual observation ($h = 8$)

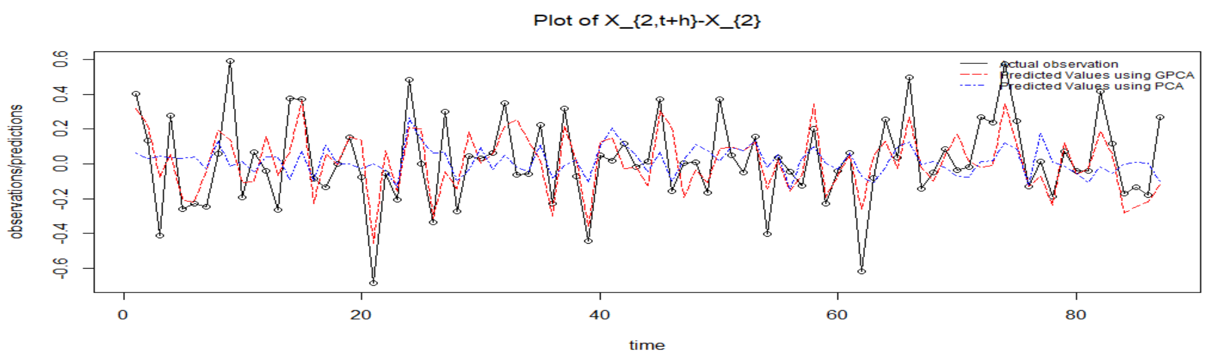


(d) Predicted values vs Actual observation ($h = 8$)

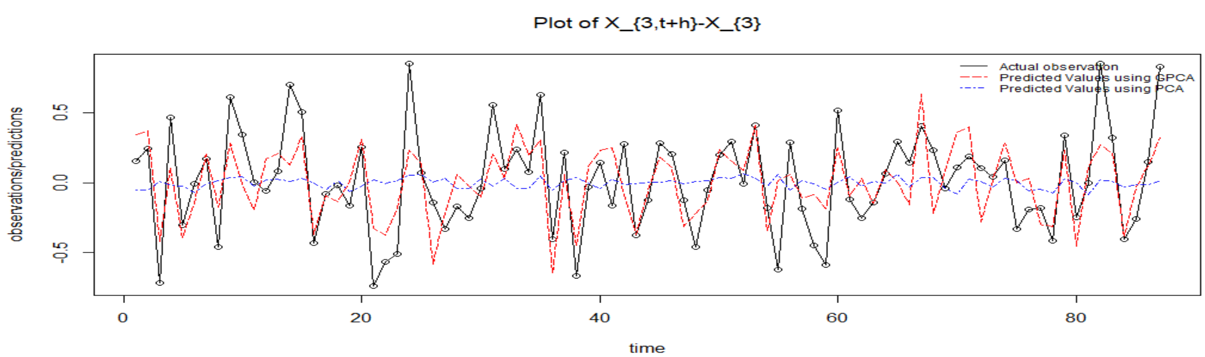
A.3.4 Horizon $h = 12$



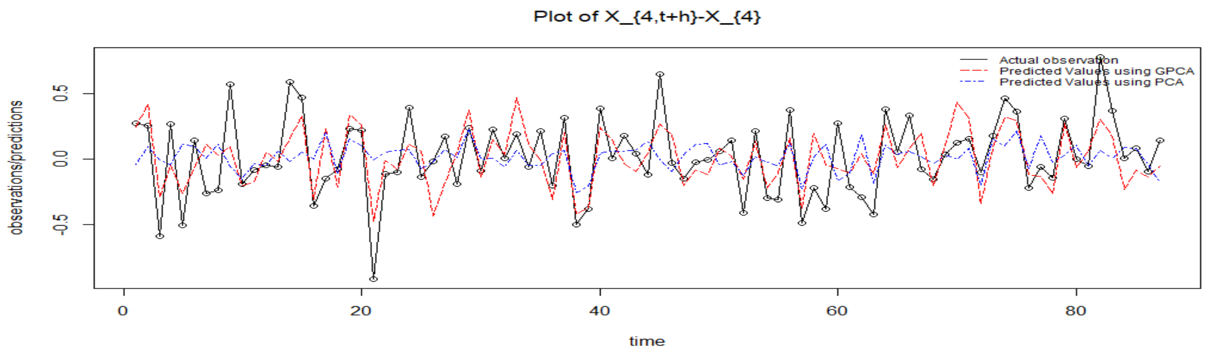
(a) Predicted values vs Actual observation ($h = 12$)



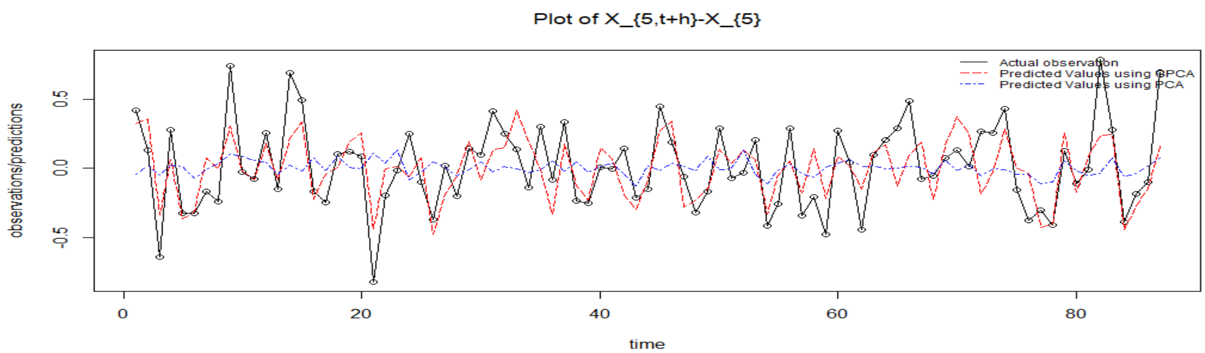
(b) Predicted values vs Actual observation ($h = 12$)



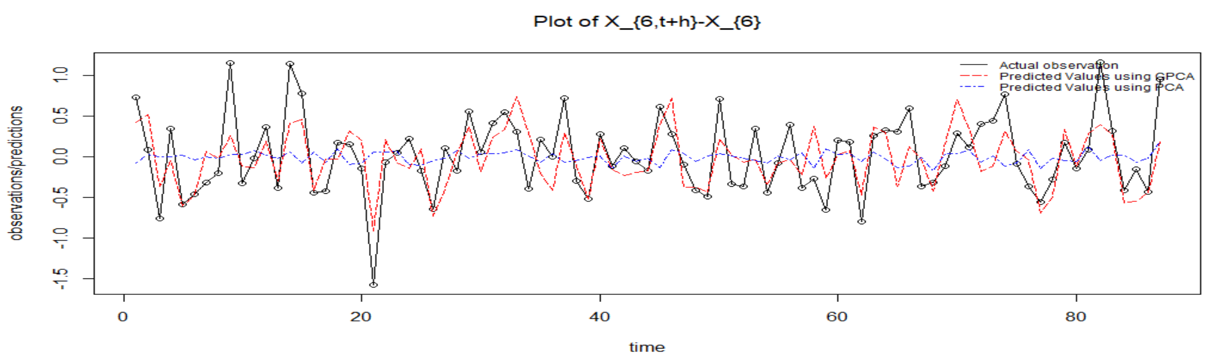
(c) Predicted values vs Actual observation ($h = 12$)



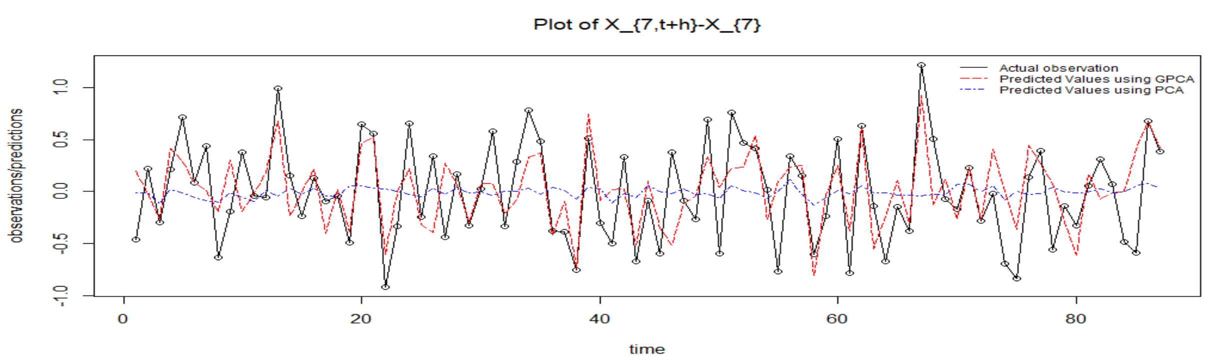
(a) Predicted values vs Actual observation ($h = 12$)



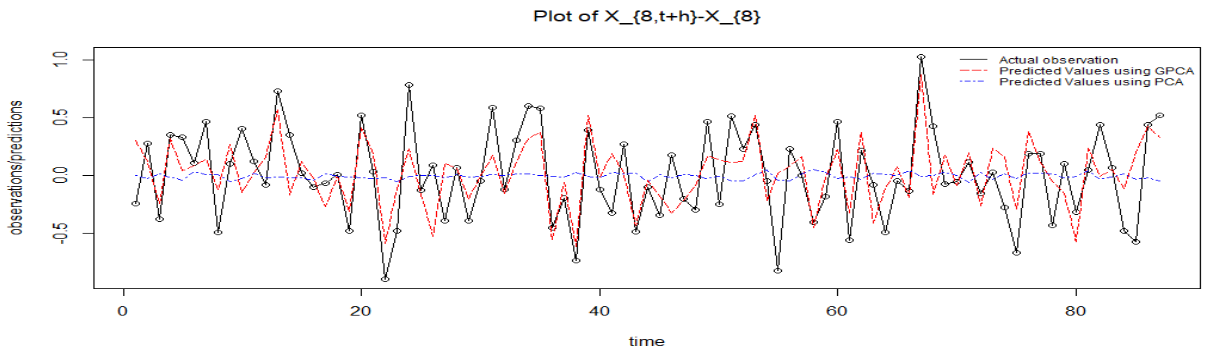
(b) Predicted values vs Actual observation ($h = 12$)



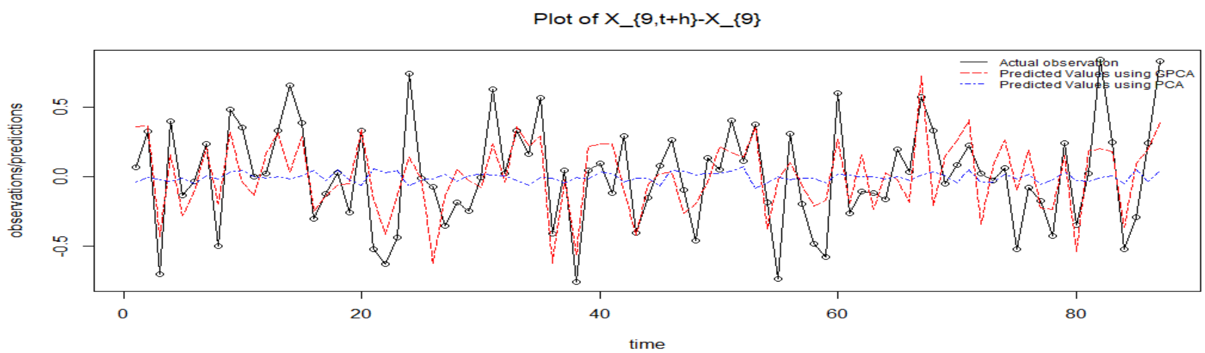
(c) Predicted values vs Actual observation ($h = 12$)



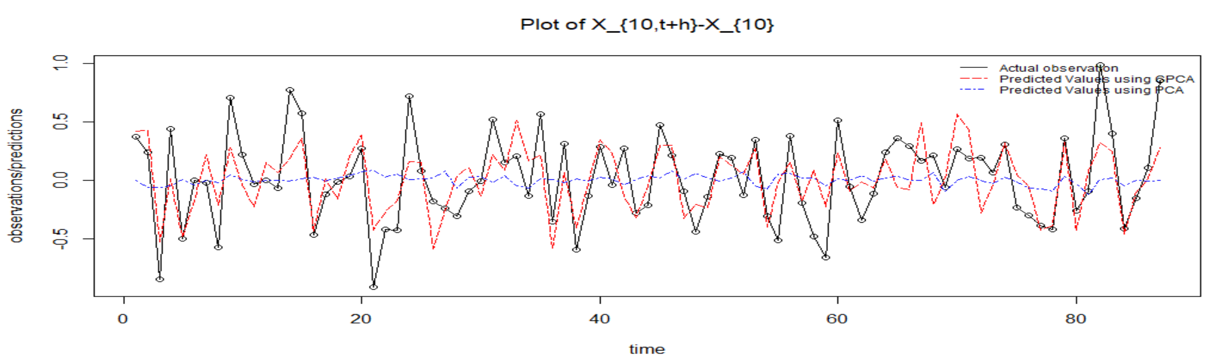
(d) Predicted values vs Actual observation ($h = 12$)



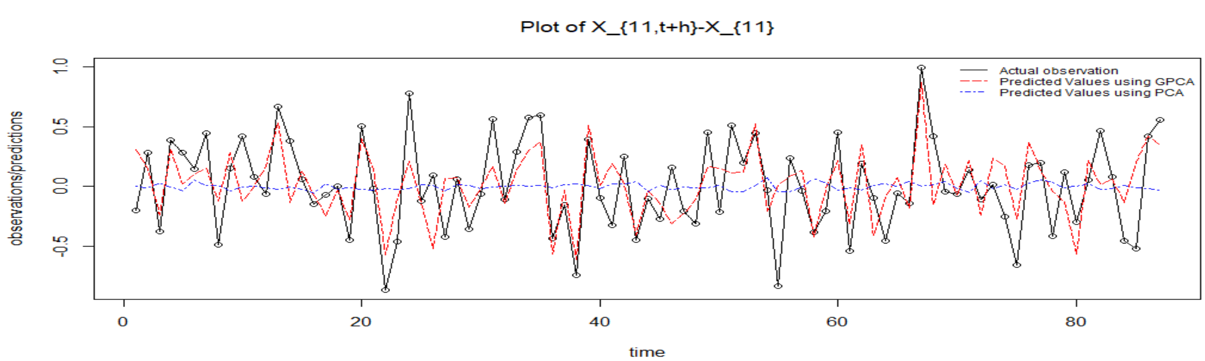
(a) Predicted values vs Actual observation ($h = 12$)



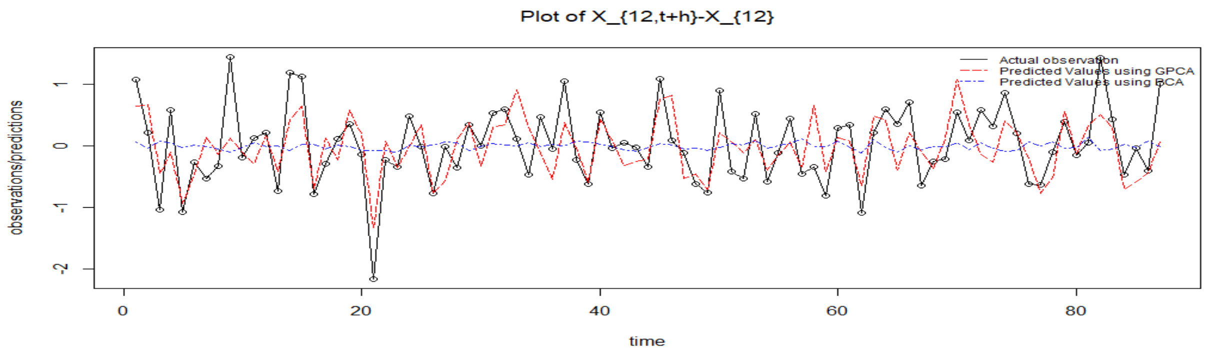
(b) Predicted values vs Actual observation ($h = 12$)



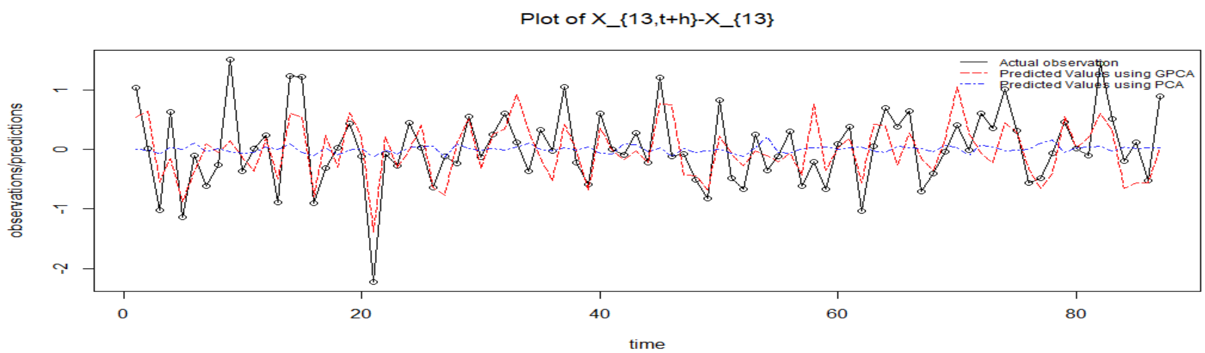
(c) Predicted values vs Actual observation ($h = 12$)



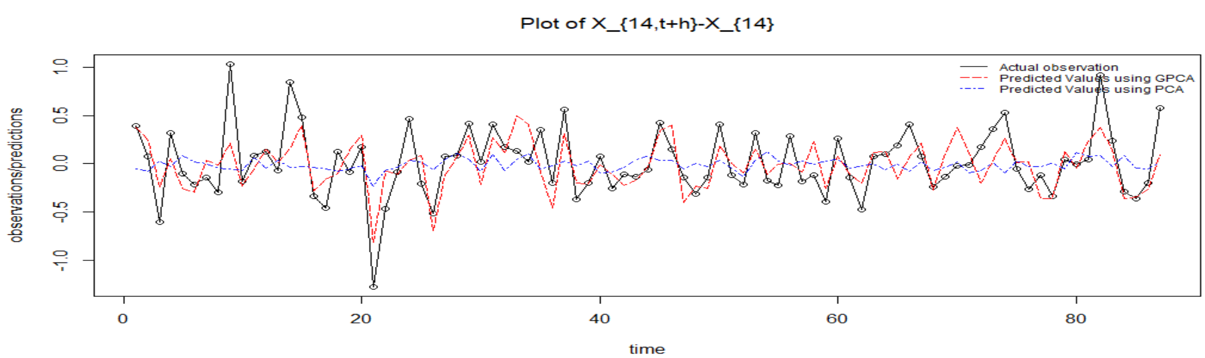
(d) Predicted values vs Actual observation ($h = 12$)



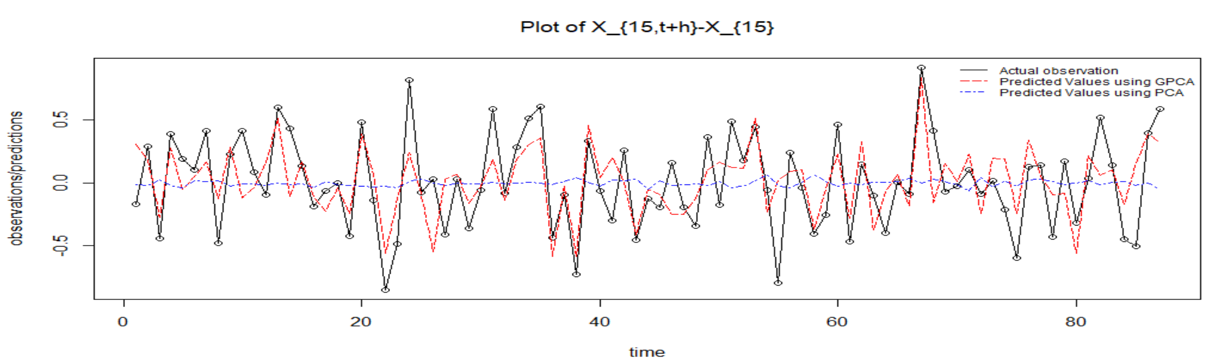
(a) Predicted values vs Actual observation ($h = 12$)



(b) Predicted values vs Actual observation ($h = 12$)



(c) Predicted values vs Actual observation ($h = 12$)



(d) Predicted values vs Actual observation ($h = 12$)

A.4 The Student's t VAR Detailed Forecasting Results

Table A.3: Individual Forecast Evaluation for Student's t GPCA

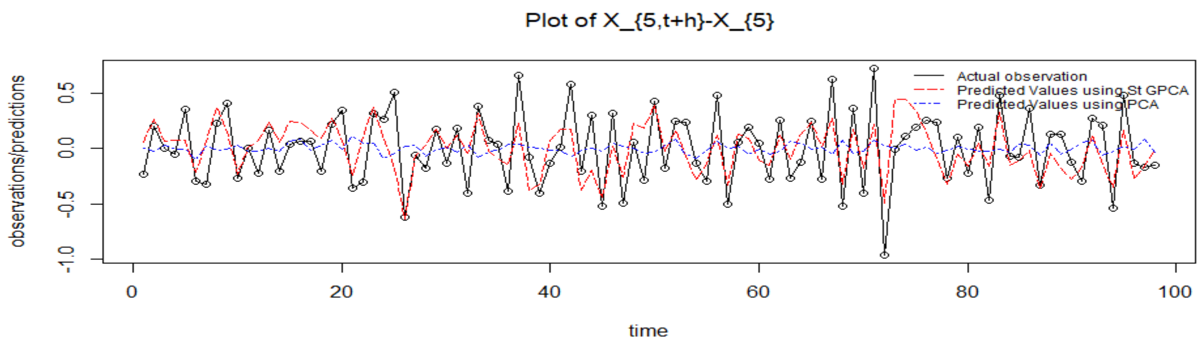
	Theil's U-statistic				Clack and West t-test			
Variables	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
$\mathbf{x}_1$	0.737	0.722	0.727	0.724	2.991	1.681	1.222	1.175
$\mathbf{x}_2$	0.700	0.748	0.703	0.745	3.122	1.642	1.340	0.982
$\mathbf{x}_3$	0.699	0.743	0.766	0.707	3.182	1.492	0.975	1.253
$\mathbf{x}_4$	0.712	0.703	0.718	0.704	2.968	2.030	1.492	1.267
$\mathbf{x}_5$	0.744	0.736	0.736	0.719	2.813	1.730	1.256	1.166
$\mathbf{x}_6$	0.758	0.720	0.737	0.704	2.835	2.096	1.382	1.421
$\mathbf{x}_7$	0.687	0.761	0.779	0.713	3.488	1.196	0.672	0.775
$\mathbf{x}_8$	0.675	0.779	0.793	0.712	3.795	1.117	0.581	0.736
$\mathbf{x}_9$	0.682	0.772	0.831	0.702	3.555	1.149	0.570	1.079
$\mathbf{x}_{10}$	0.720	0.731	0.754	0.705	2.823	1.707	1.128	1.324
$\mathbf{x}_{11}$	0.672	0.781	0.794	0.708	3.892	1.099	0.577	0.748
$\mathbf{x}_{12}$	0.738	0.718	0.728	0.710	2.893	2.196	1.477	1.505
$\mathbf{x}_{13}$	0.745	0.705	0.728	0.708	2.946	2.268	1.438	1.487
$\mathbf{x}_{14}$	0.738	0.744	0.752	0.717	2.706	1.562	1.348	1.391
$\mathbf{x}_{15}$	0.670	0.785	0.802	0.710	3.920	1.088	0.563	0.760

Table A.4: Individual Forecast Evaluation for PCA

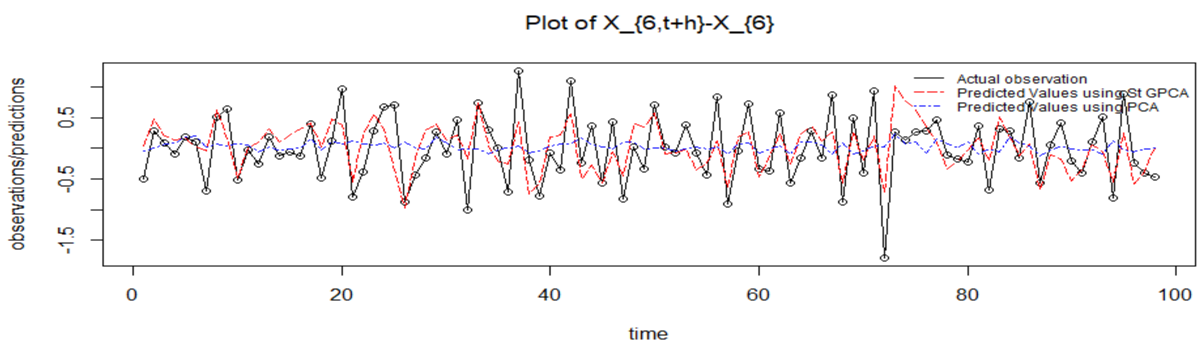
	Theil's U-statistic				Clack and West t-test			
Variables	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
$\mathbf{x}_1$	1.004	1.004	0.996	1.011	-0.276	-0.145	0.084	-0.180
$\mathbf{x}_2$	0.930	0.966	0.961	0.983	2.496	0.506	0.452	0.143
$\mathbf{x}_3$	0.989	1.009	0.980	0.997	0.951	-0.309	0.402	0.068
$\mathbf{x}_4$	0.950	0.981	0.949	0.951	1.214	0.283	0.408	0.291
$\mathbf{x}_5$	1.021	1.003	0.992	1.016	-1.503	-0.157	0.247	-0.434
$\mathbf{x}_6$	0.999	0.979	0.990	0.980	0.066	0.713	0.211	0.465
$\mathbf{x}_7$	0.993	0.999	1.015	1.002	0.771	0.042	-0.510	-0.065
$\mathbf{x}_8$	1.013	1.010	1.004	1.017	-1.744	-0.517	-0.096	-0.438
$\mathbf{x}_9$	0.990	1.002	0.996	0.979	1.230	-0.116	0.117	0.582
$\mathbf{x}_{10}$	0.999	1.014	0.982	1.000	0.082	-0.458	0.358	0.006
$\mathbf{x}_{11}$	1.011	1.010	0.999	1.015	-1.575	-0.544	0.023	-0.428
$\mathbf{x}_{12}$	0.990	0.988	0.989	0.992	1.560	1.389	0.861	0.458
$\mathbf{x}_{13}$	1.002	0.999	1.011	1.005	-0.229	0.072	-0.502	-0.201
$\mathbf{x}_{14}$	0.978	0.988	0.999	0.993	1.347	0.349	0.032	0.176
$\mathbf{x}_{15}$	1.007	1.006	0.997	1.011	-1.154	-0.426	0.077	-0.367

A.5 The Student's t VAR: Predictions vs. Actual Observations Plots

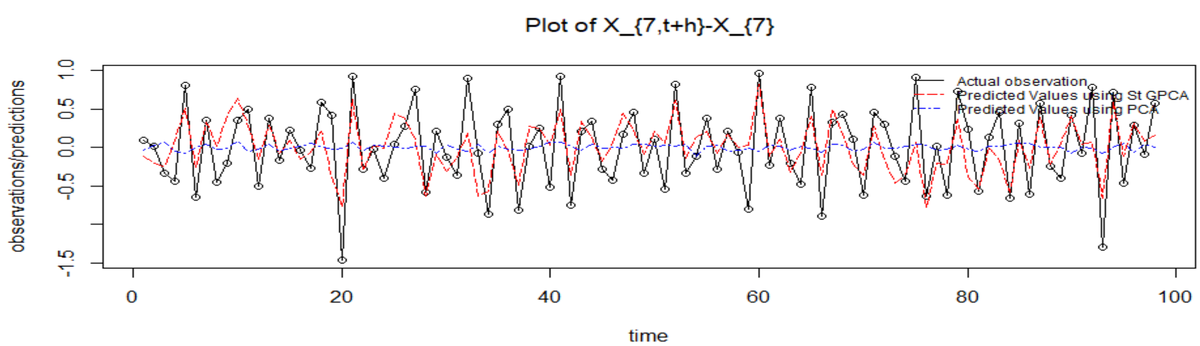
A.5.1 Horizon $h = 1$



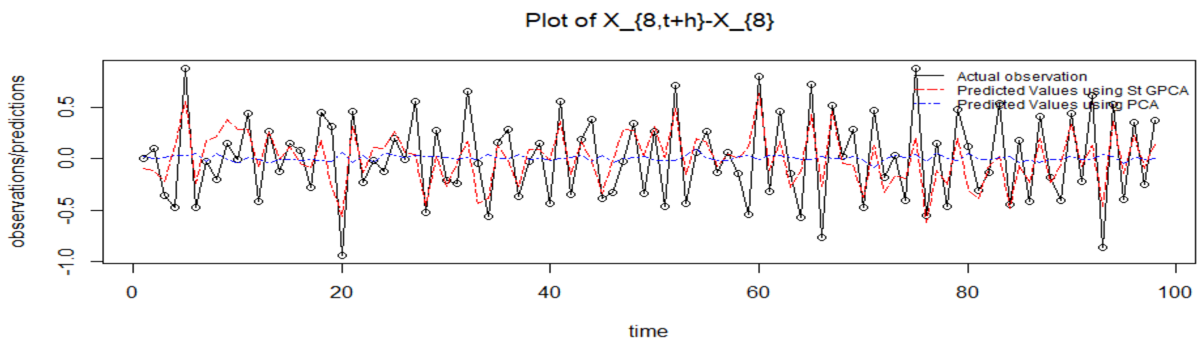
(a) Predicted values vs Actual observation ($h = 1$)



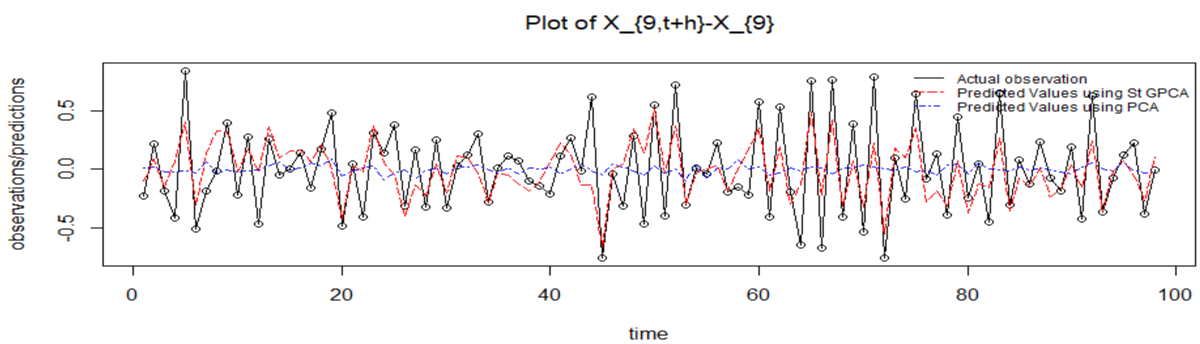
(b) Predicted values vs Actual observation ($h = 1$)



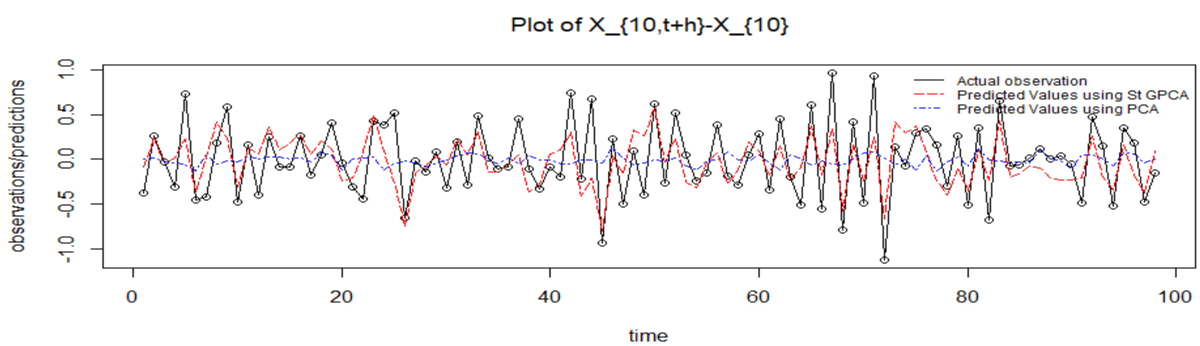
(c) Predicted values vs Actual observation ($h = 1$)



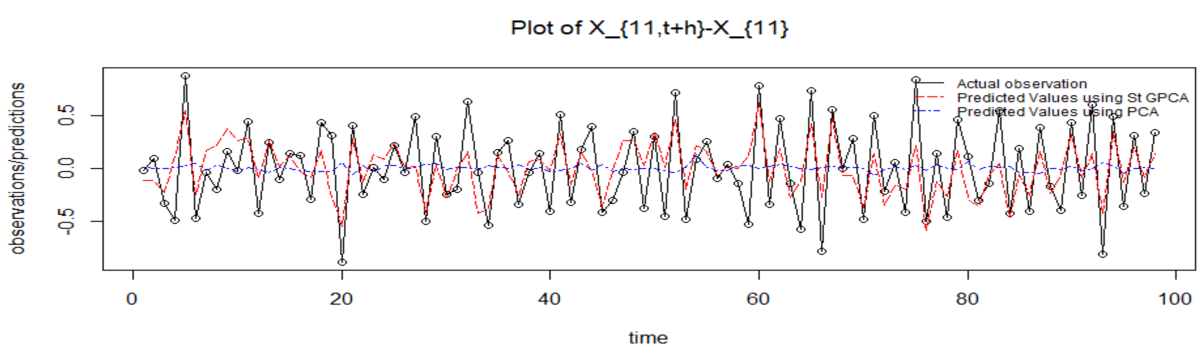
(a) Predicted values vs Actual observation ($h = 1$)



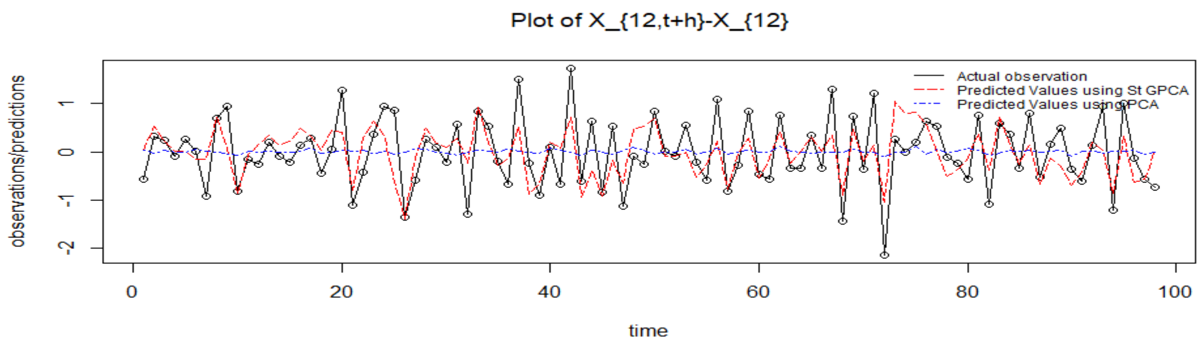
(b) Predicted values vs Actual observation ($h = 1$)



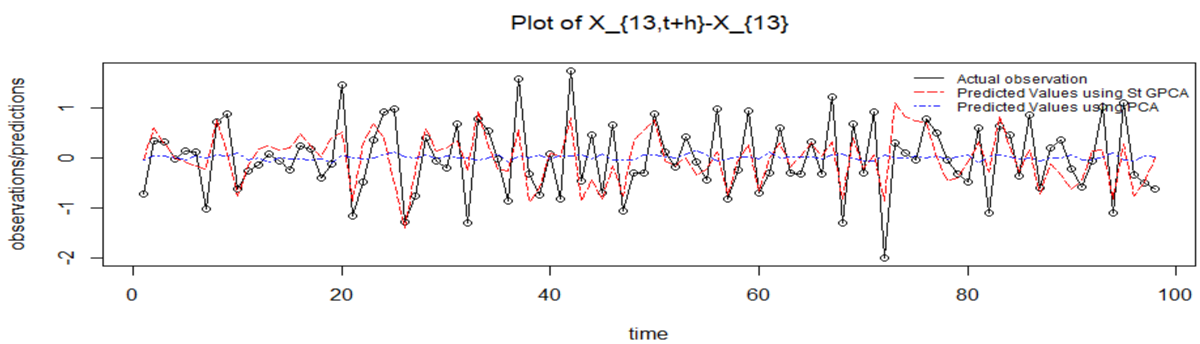
(c) Predicted values vs Actual observation ($h = 1$)



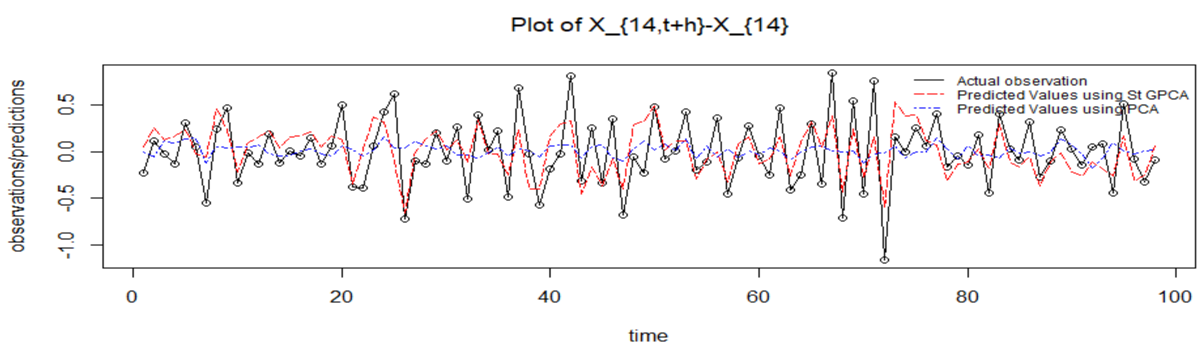
(d) Predicted values vs Actual observation ($h = 1$)



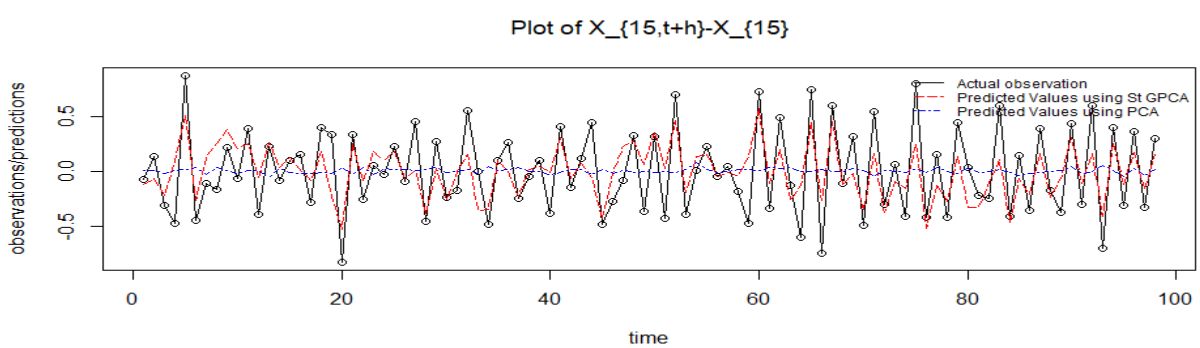
(a) Predicted values vs Actual observation ($h = 1$)



(b) Predicted values vs Actual observation ($h = 1$)

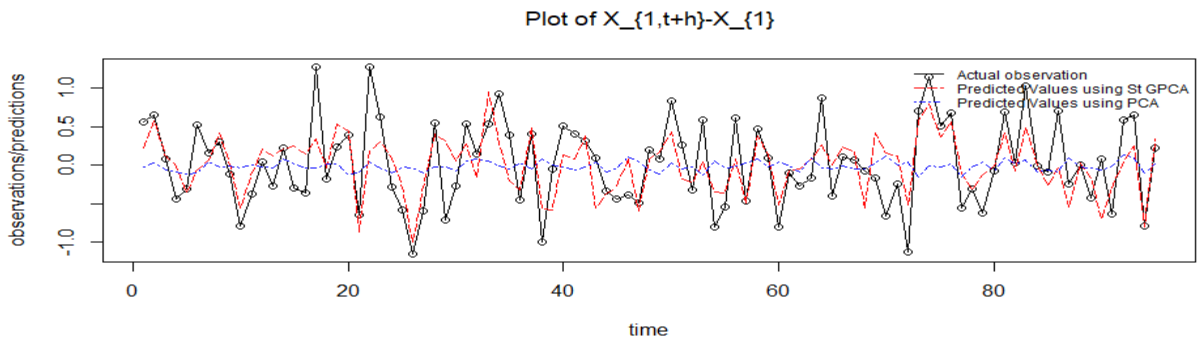


(c) Predicted values vs Actual observation ($h = 1$)

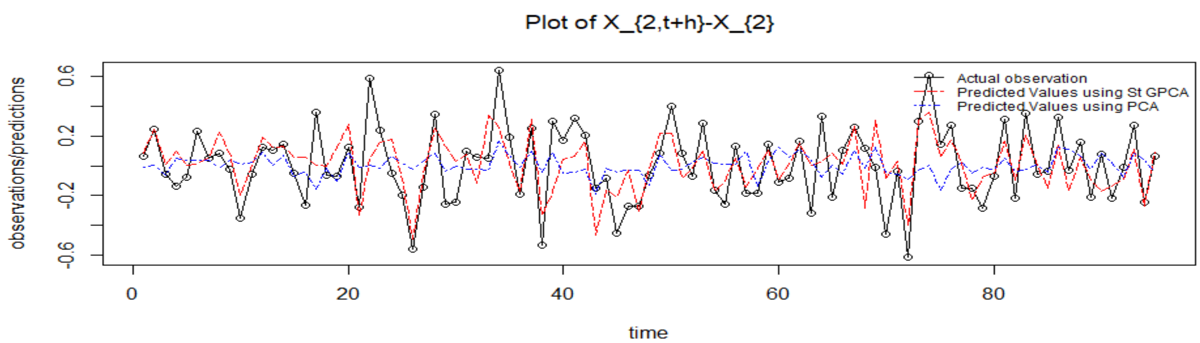


(d) Predicted values vs Actual observation ($h = 1$)

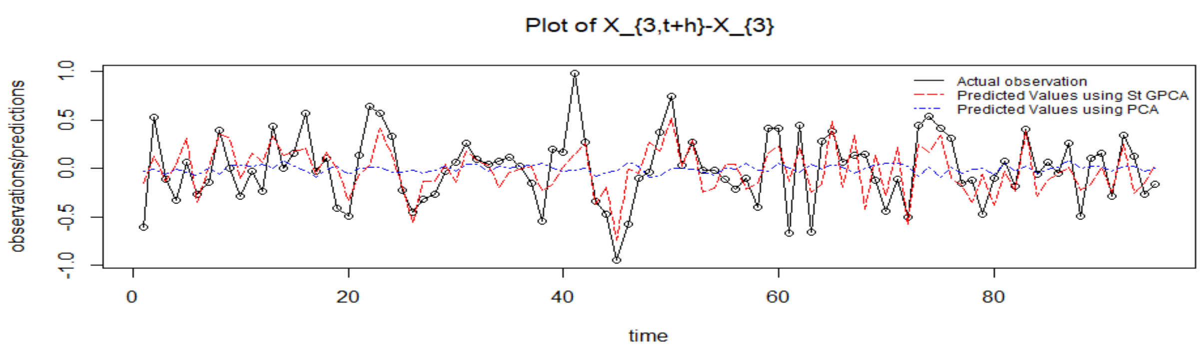
A.5.2 Horizon $h = 4$



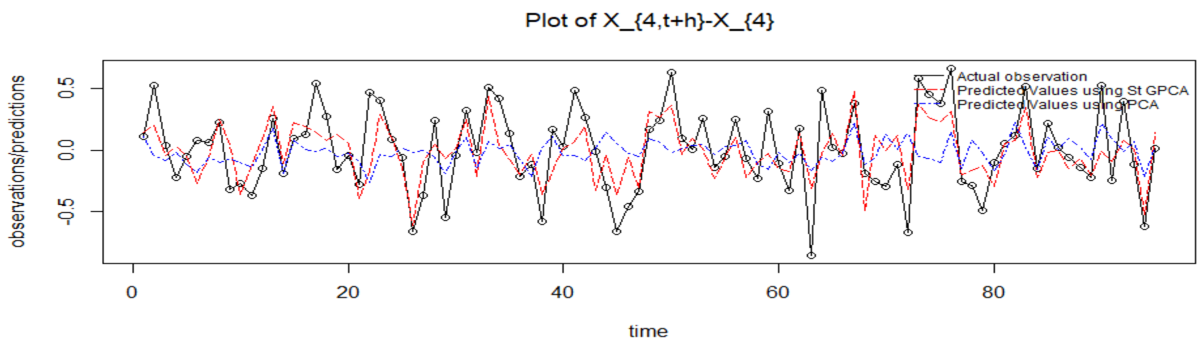
(a) Predicted values vs Actual observation ($h = 4$)



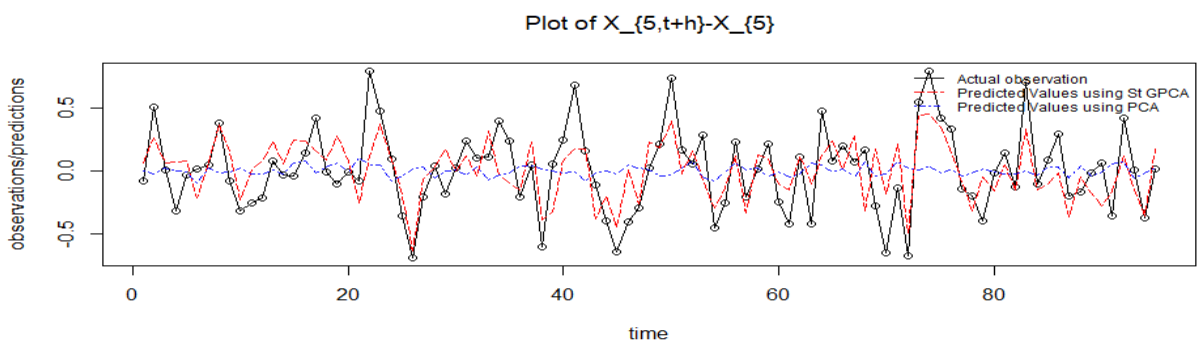
(b) Predicted values vs Actual observation ($h = 4$)



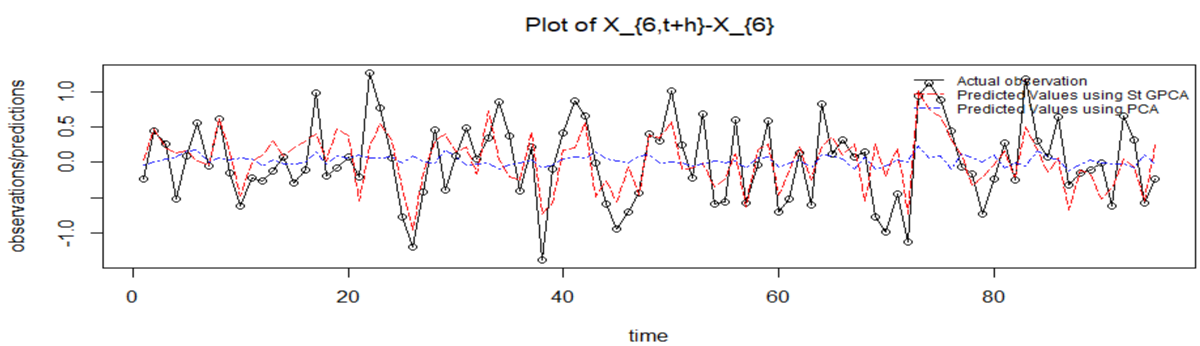
(c) Predicted values vs Actual observation ($h = 4$)



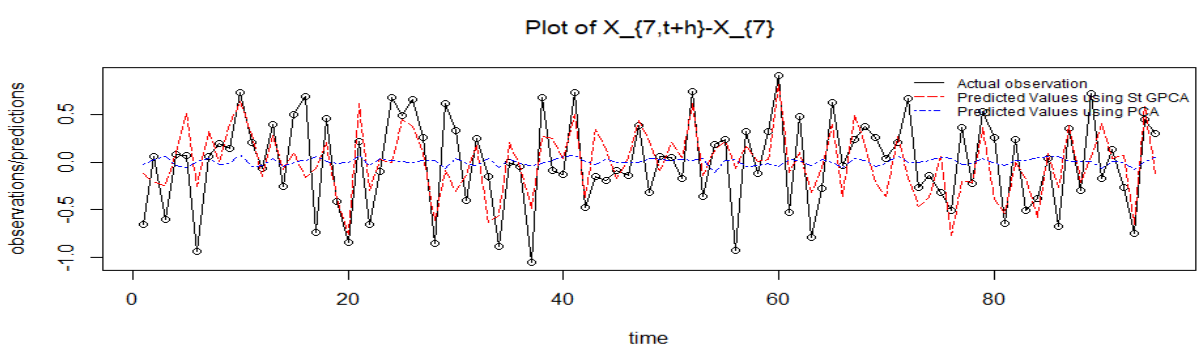
(a) Predicted values vs Actual observation ($h = 4$)



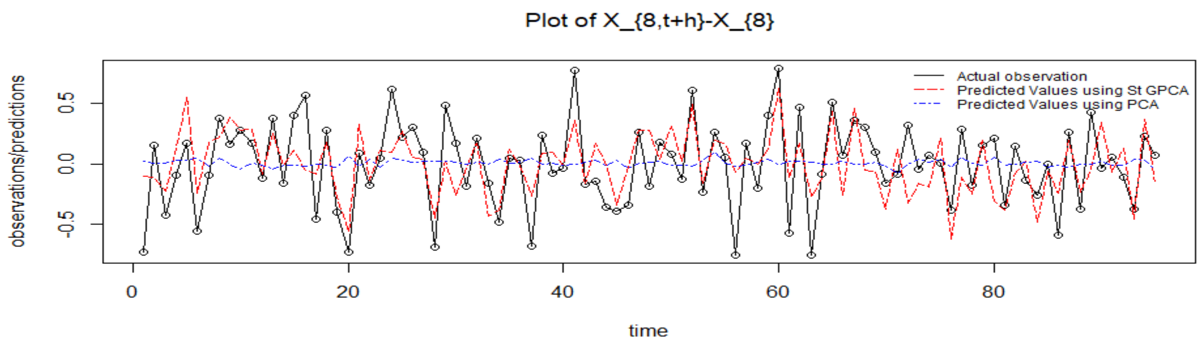
(b) Predicted values vs Actual observation ($h = 4$)



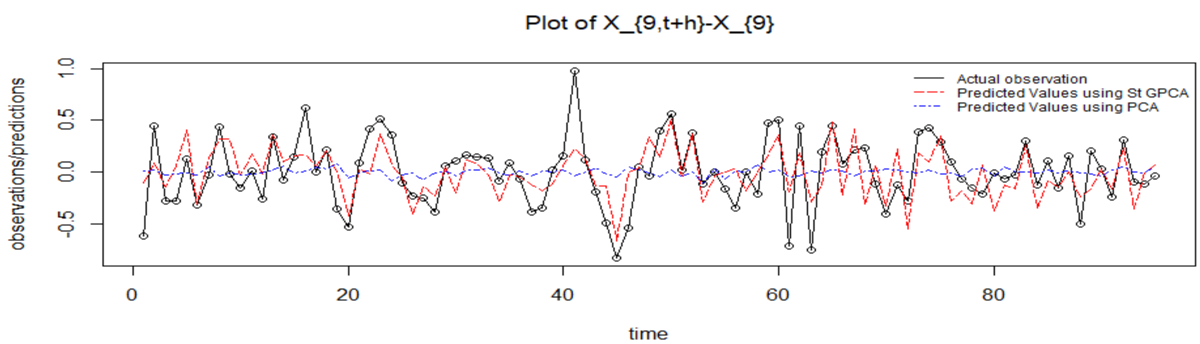
(c) Predicted values vs Actual observation ($h = 4$)



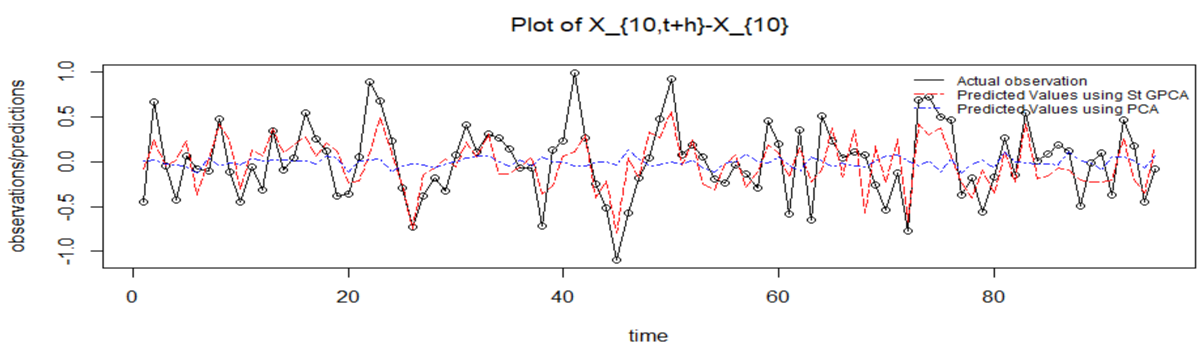
(d) Predicted values vs Actual observation ($h = 4$)



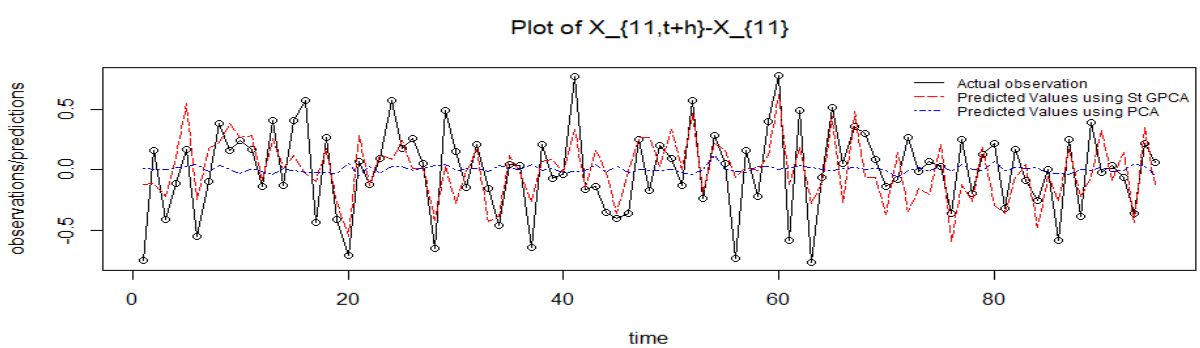
(a) Predicted values vs Actual observation ($h = 4$)



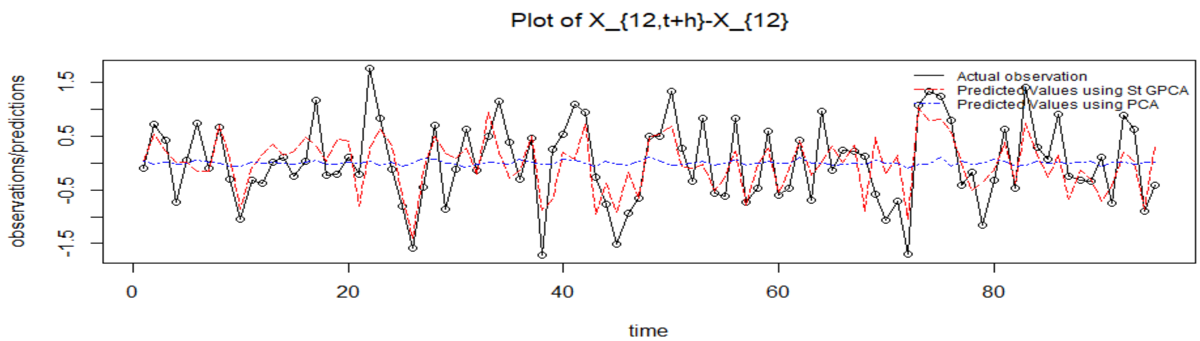
(b) Predicted values vs Actual observation ($h = 4$)



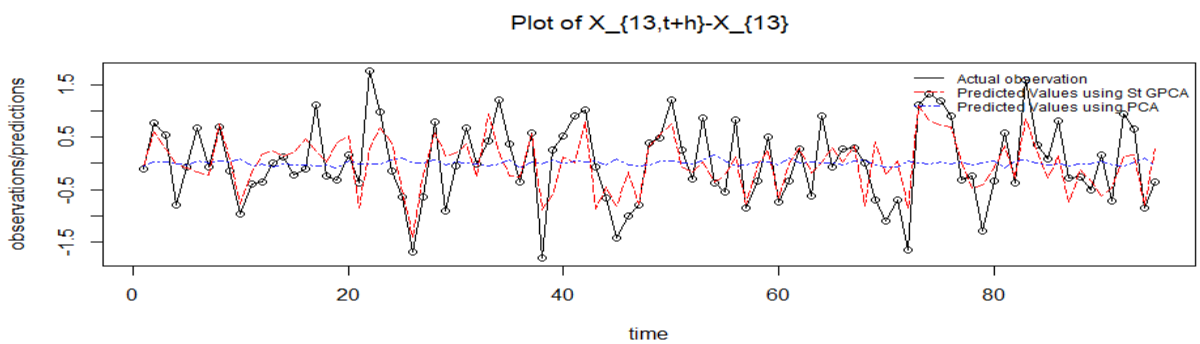
(c) Predicted values vs Actual observation ($h = 4$)



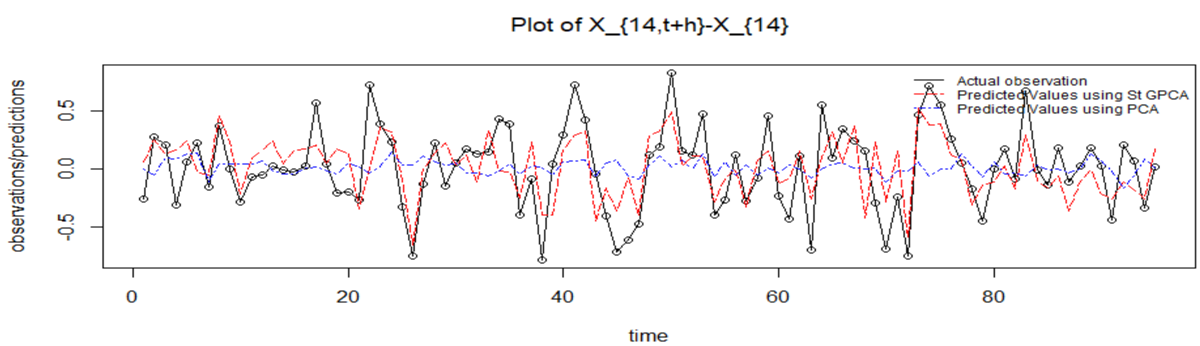
(d) Predicted values vs Actual observation ($h = 4$)



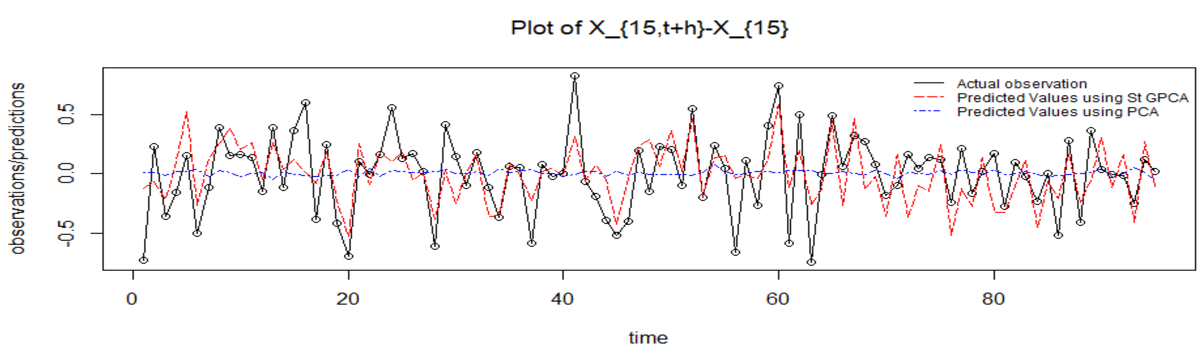
(a) Predicted values vs Actual observation ($h = 4$)



(b) Predicted values vs Actual observation ($h = 4$)

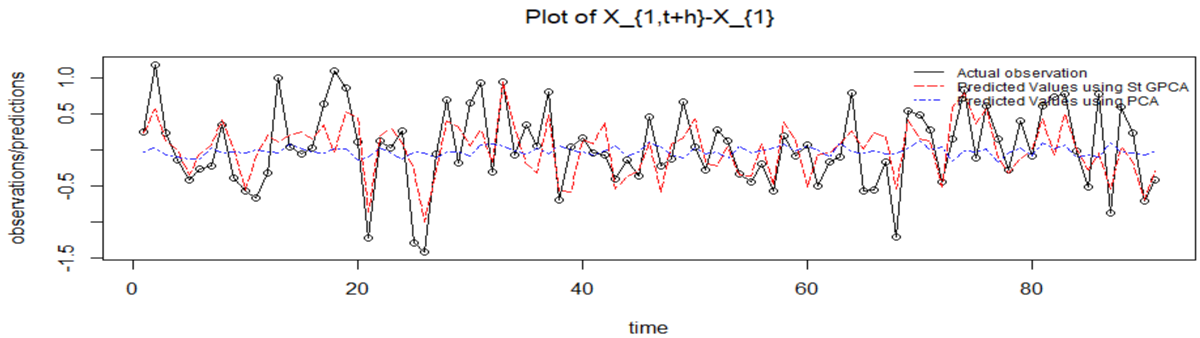


(c) Predicted values vs Actual observation ($h = 4$)

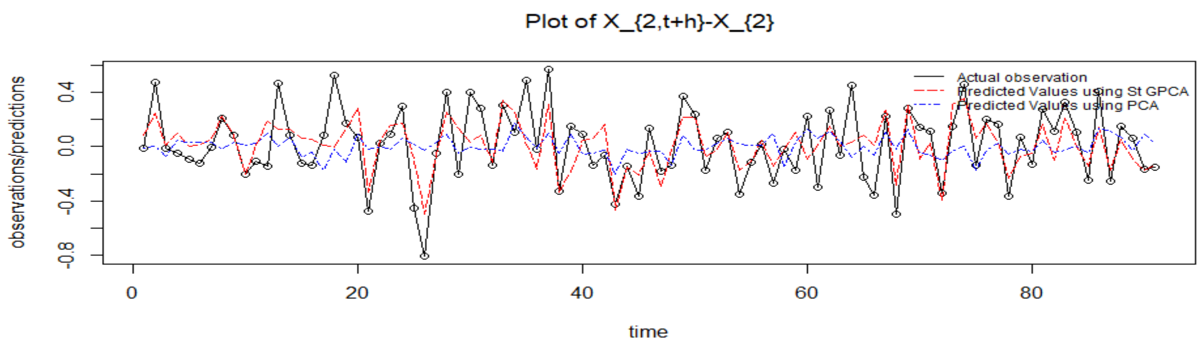


(d) Predicted values vs Actual observation ($h = 4$)

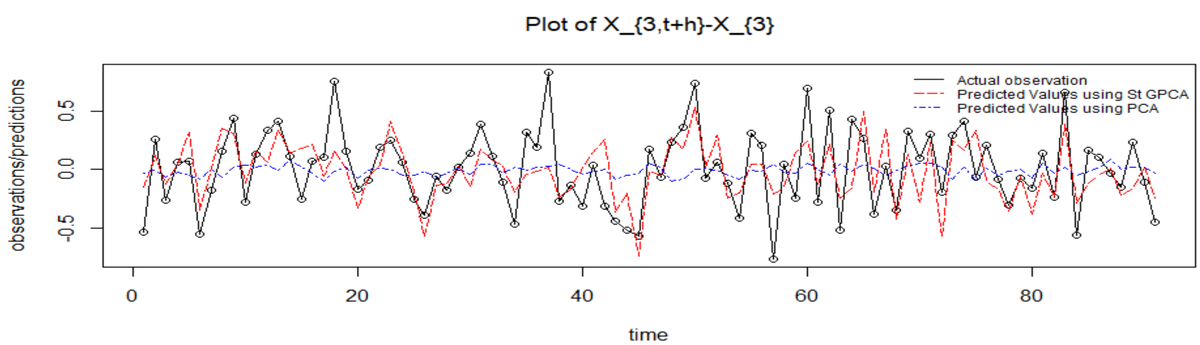
A.5.3 Horizon $h = 8$



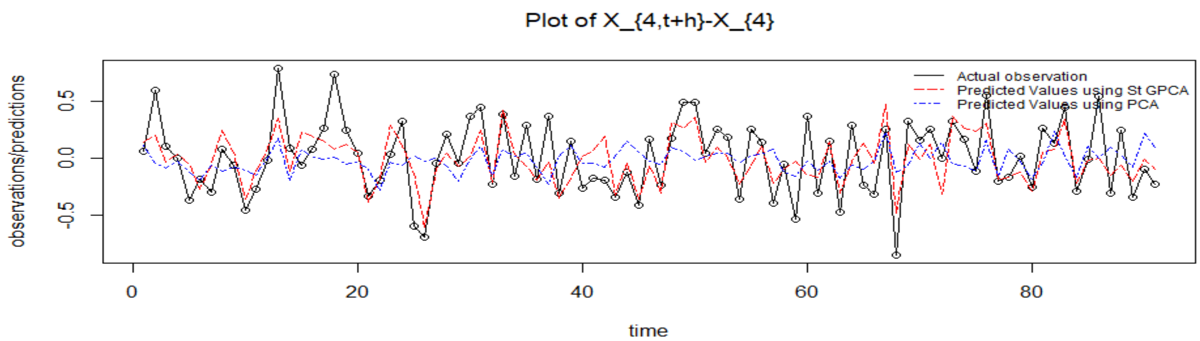
(a) Predicted values vs Actual observation ($h = 8$)



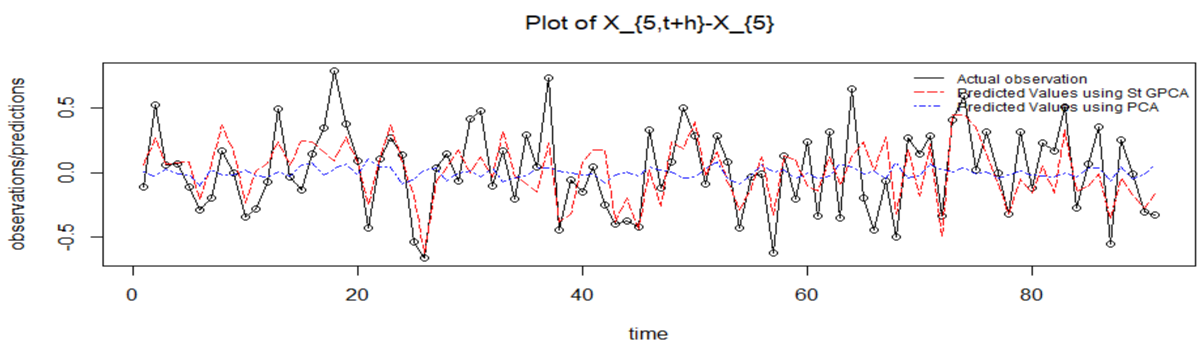
(b) Predicted values vs Actual observation ($h = 8$)



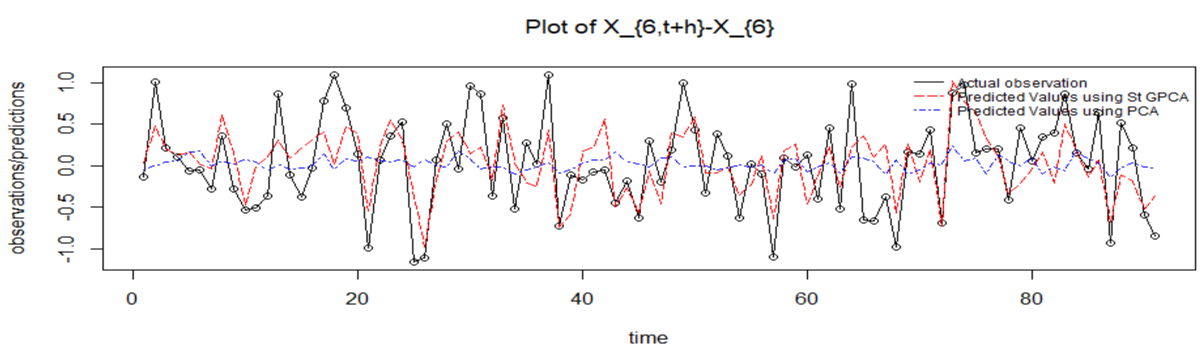
(c) Predicted values vs Actual observation ($h = 8$)



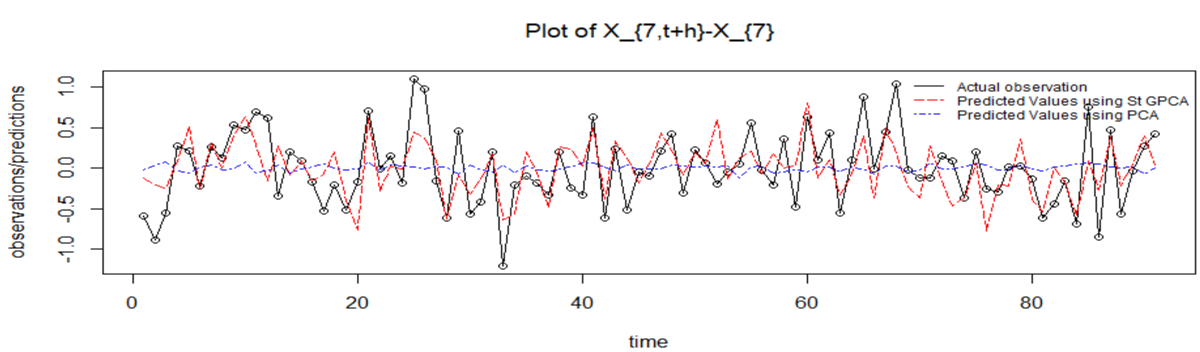
(a) Predicted values vs Actual observation ($h = 8$)



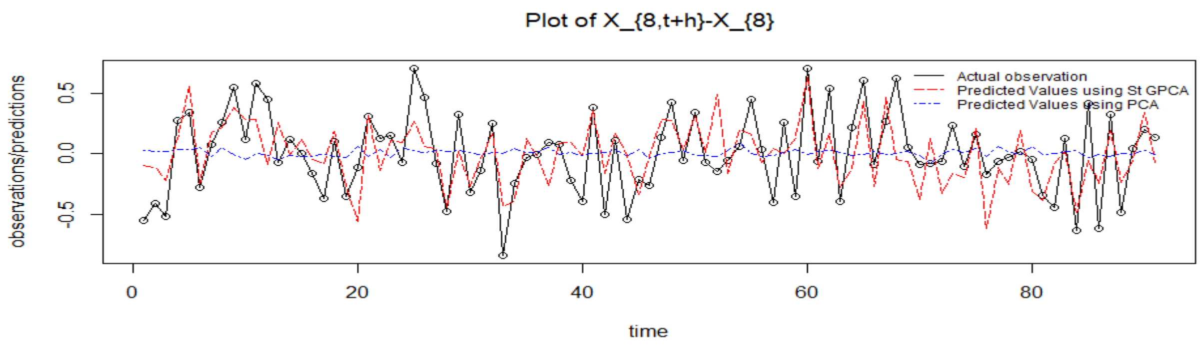
(b) Predicted values vs Actual observation ($h = 8$)



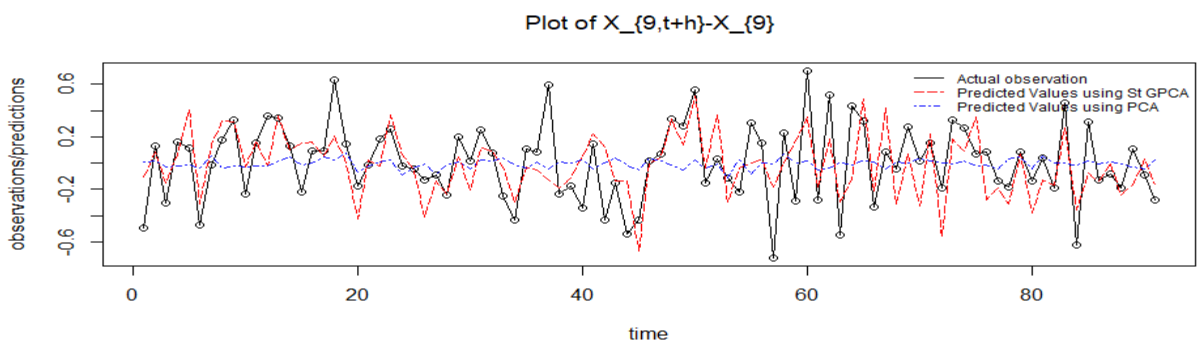
(c) Predicted values vs Actual observation ($h = 8$)



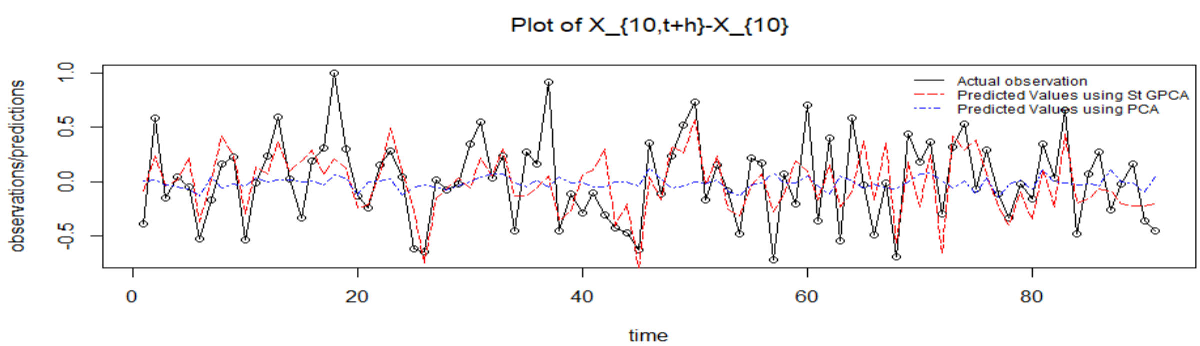
(d) Predicted values vs Actual observation ($h = 8$)



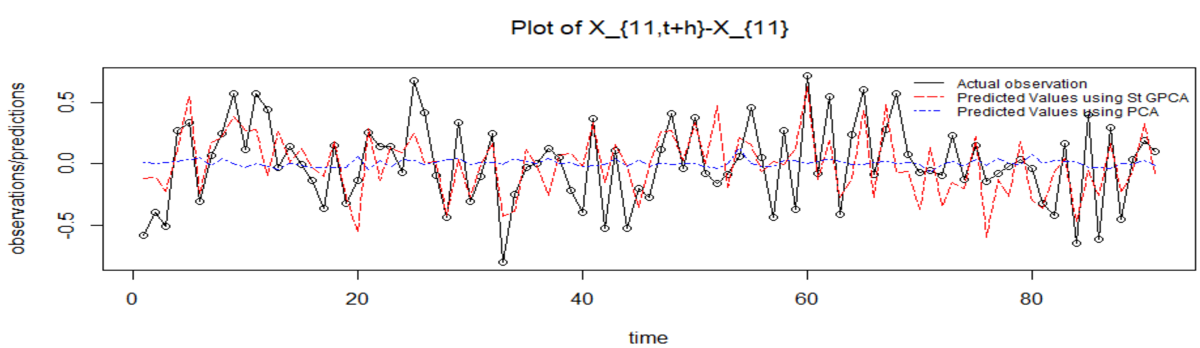
(a) Predicted values vs Actual observation ($h = 8$)



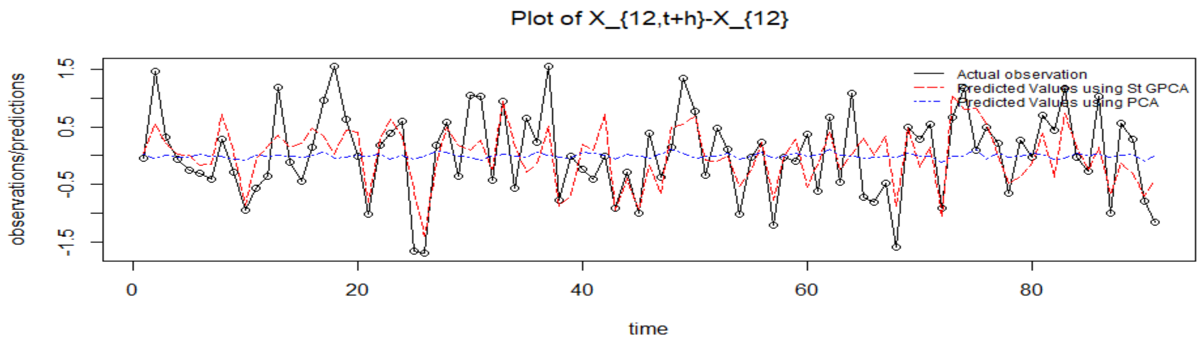
(b) Predicted values vs Actual observation ($h = 8$)



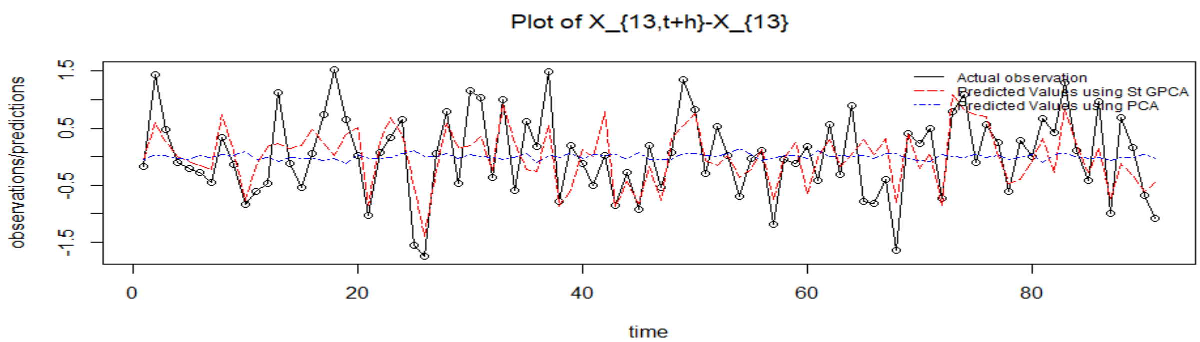
(c) Predicted values vs Actual observation ($h = 8$)



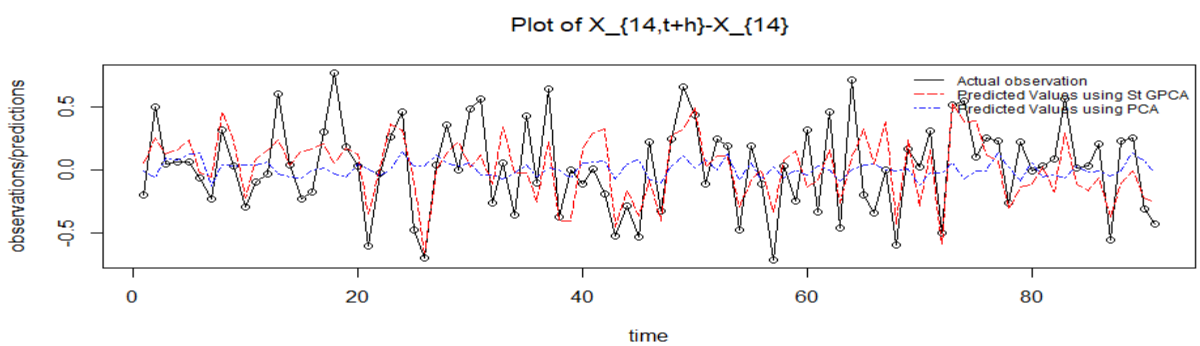
(d) Predicted values vs Actual observation ($h = 8$)



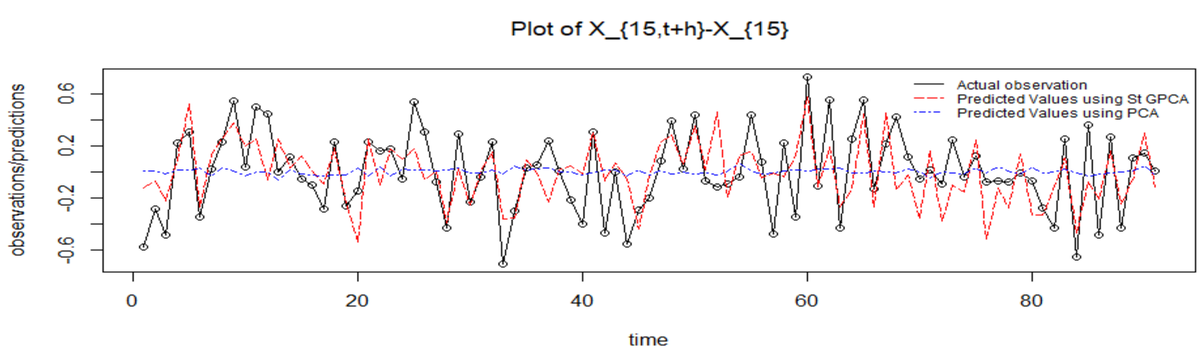
(a) Predicted values vs Actual observation ($h = 8$)



(b) Predicted values vs Actual observation ($h = 8$)

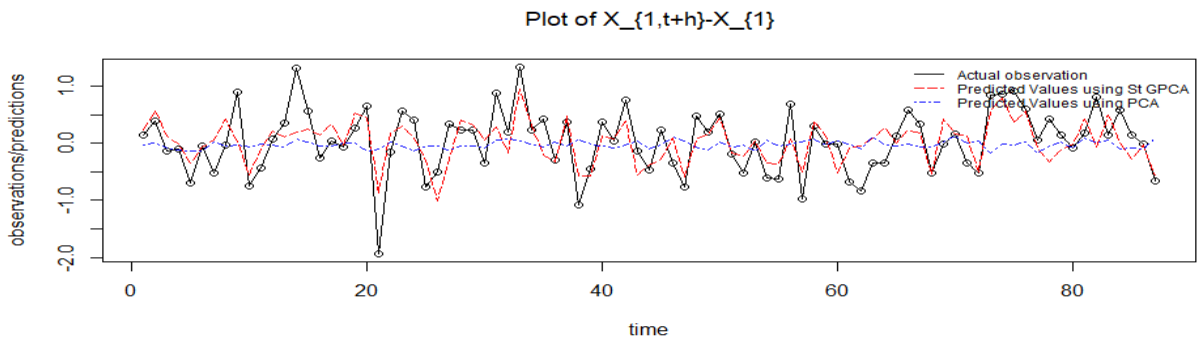


(c) Predicted values vs Actual observation ($h = 8$)

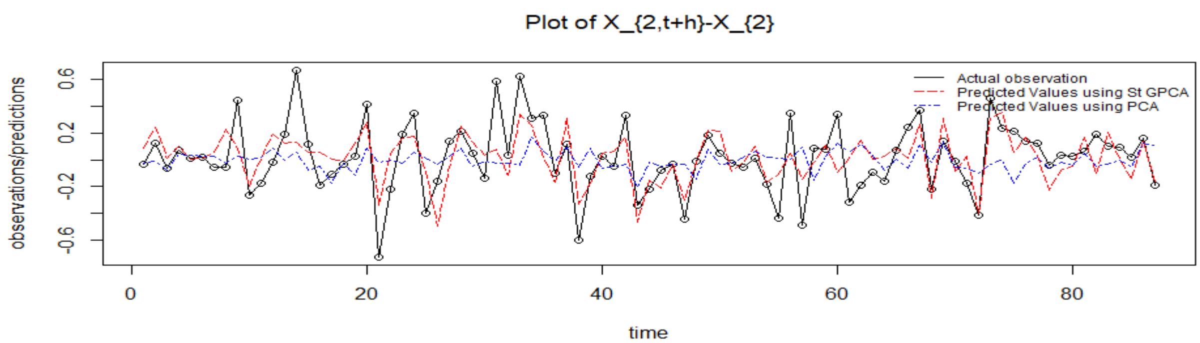


(d) Predicted values vs Actual observation ($h = 8$)

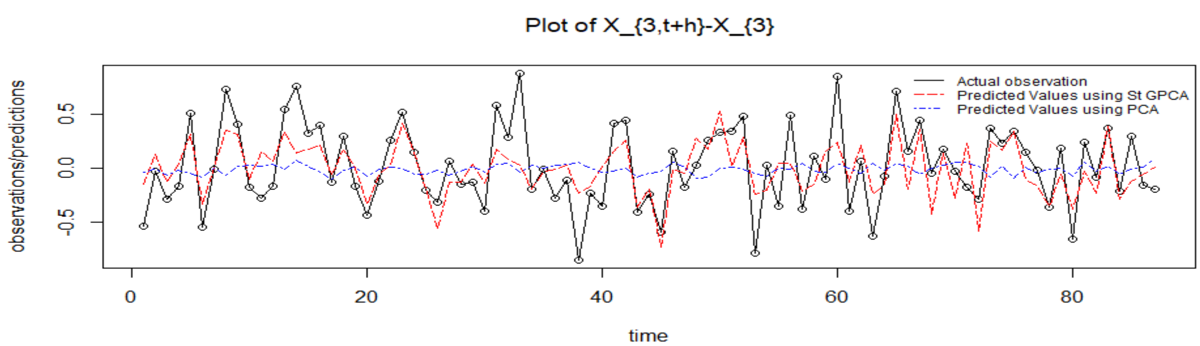
A.5.4 Horizon $h = 12$



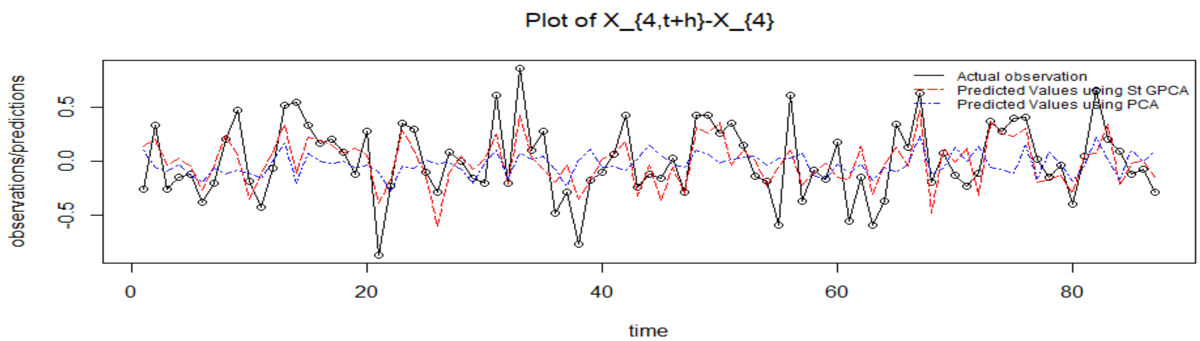
(a) Predicted values vs Actual observation ($h = 12$)



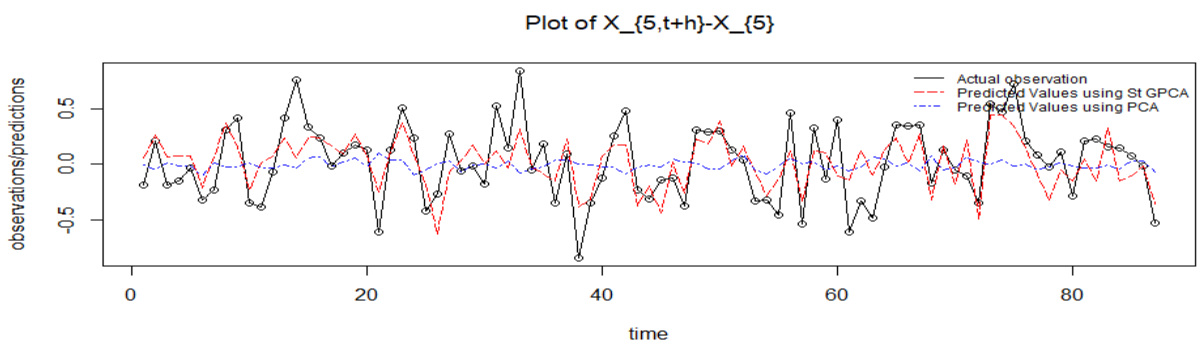
(b) Predicted values vs Actual observation ($h = 12$)



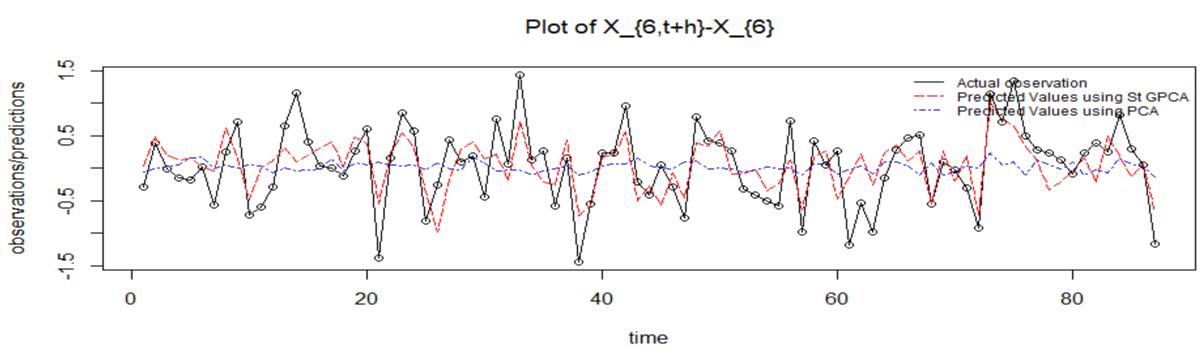
(c) Predicted values vs Actual observation ($h = 12$)



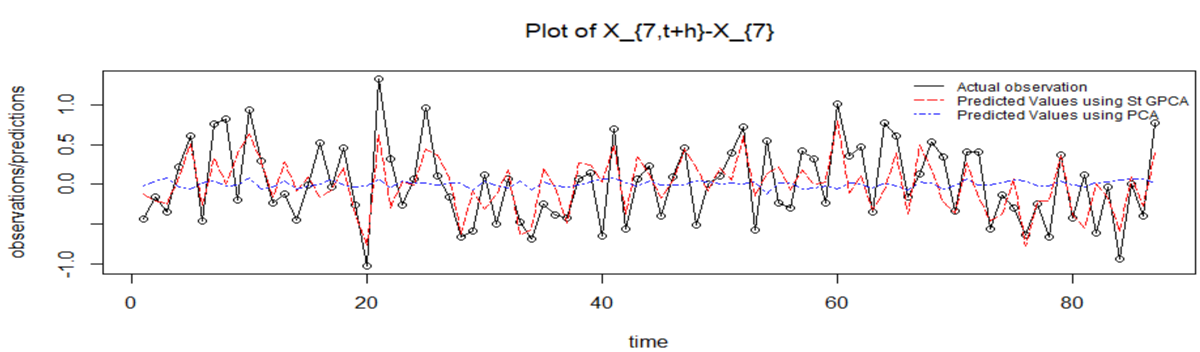
(a) Predicted values vs Actual observation ($h = 12$)



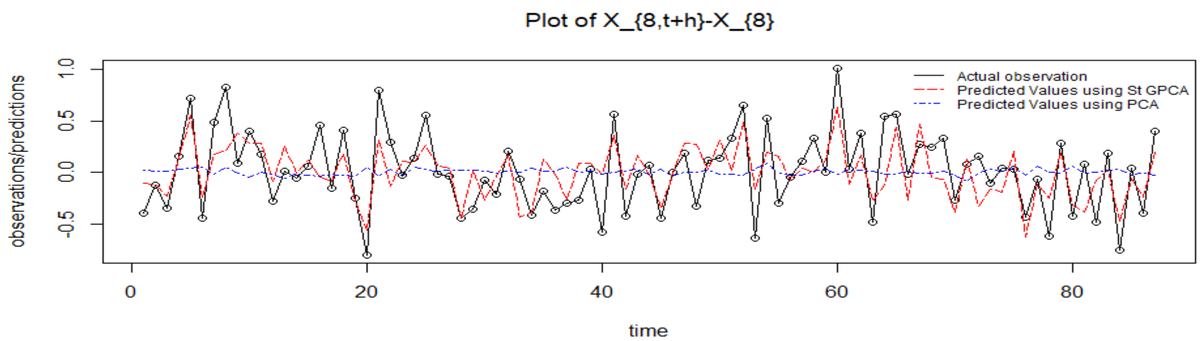
(b) Predicted values vs Actual observation ($h = 12$)



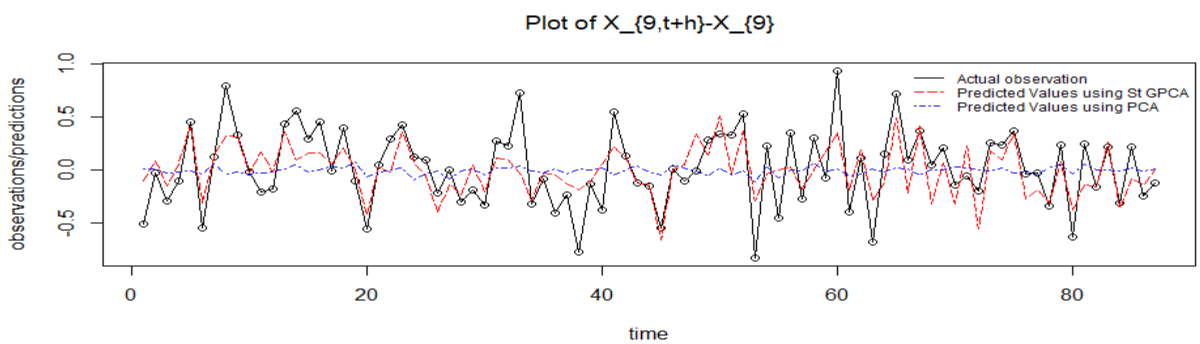
(c) Predicted values vs Actual observation ($h = 12$)



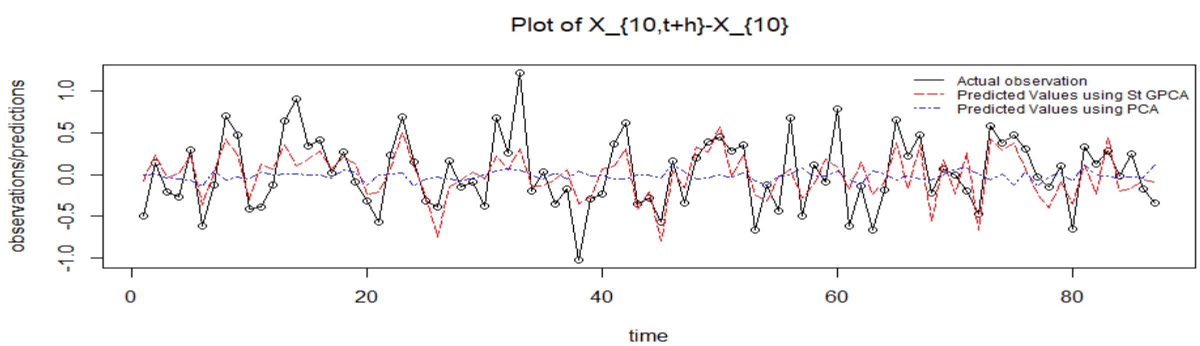
(d) Predicted values vs Actual observation ($h = 12$)



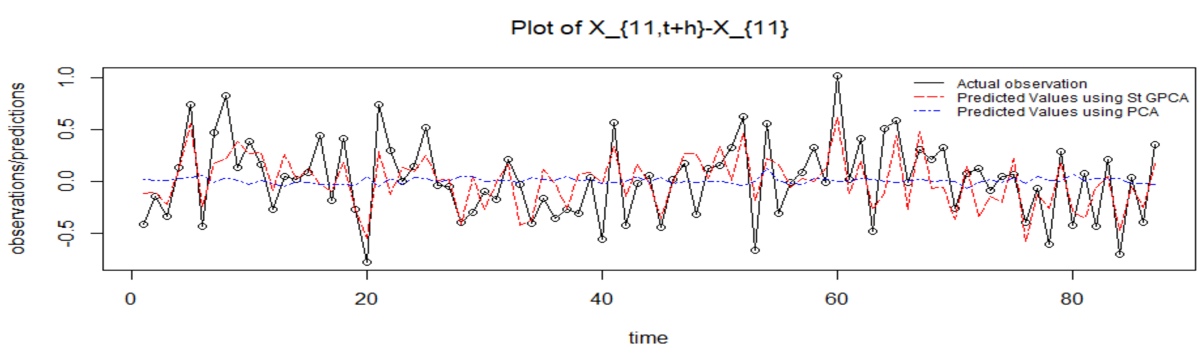
(a) Predicted values vs Actual observation ($h = 12$)



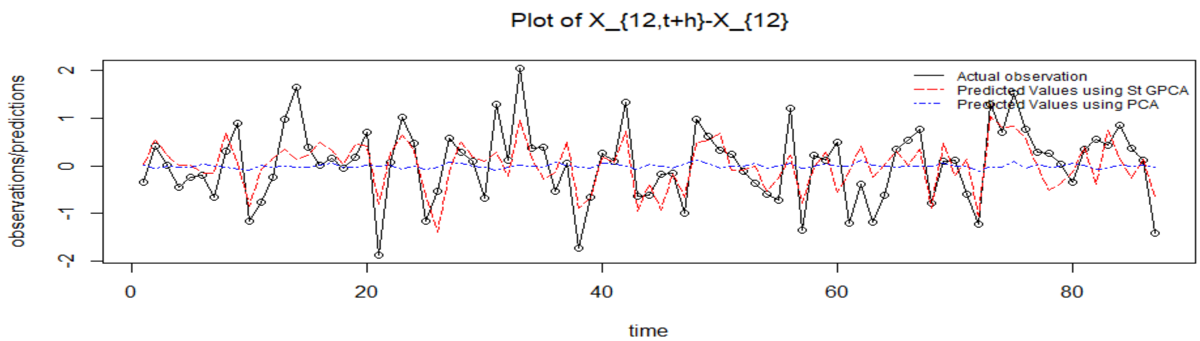
(b) Predicted values vs Actual observation ($h = 12$)



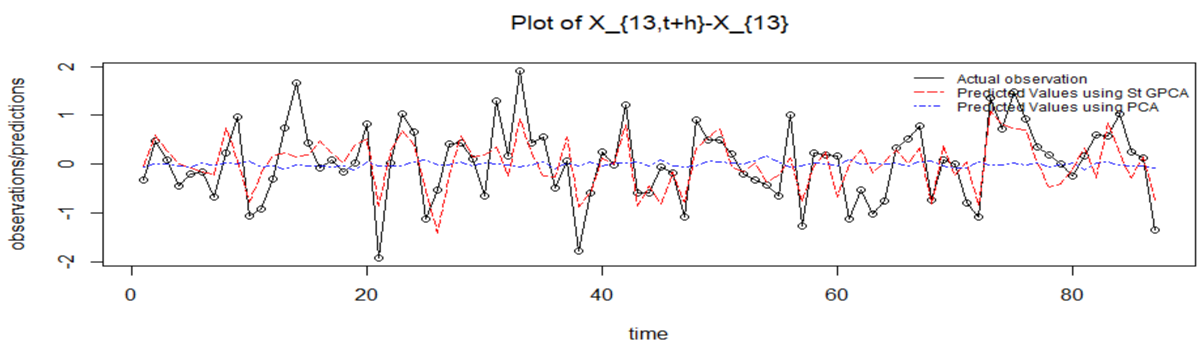
(c) Predicted values vs Actual observation ($h = 12$)



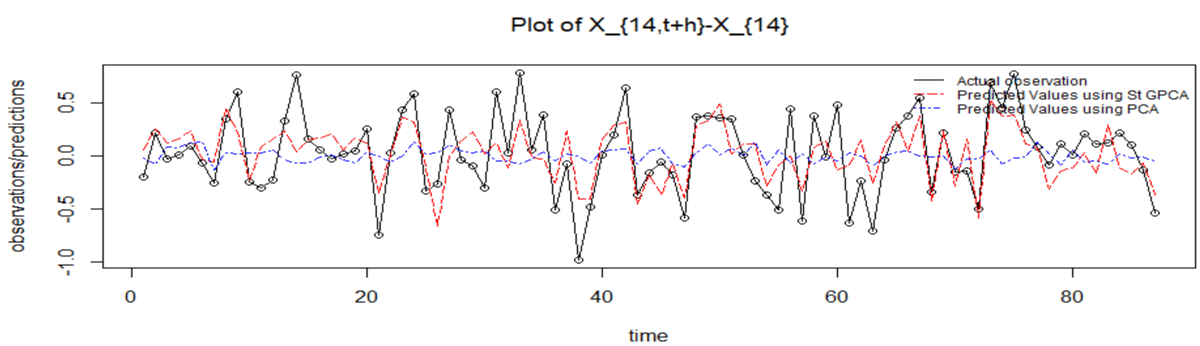
(d) Predicted values vs Actual observation ($h = 12$)



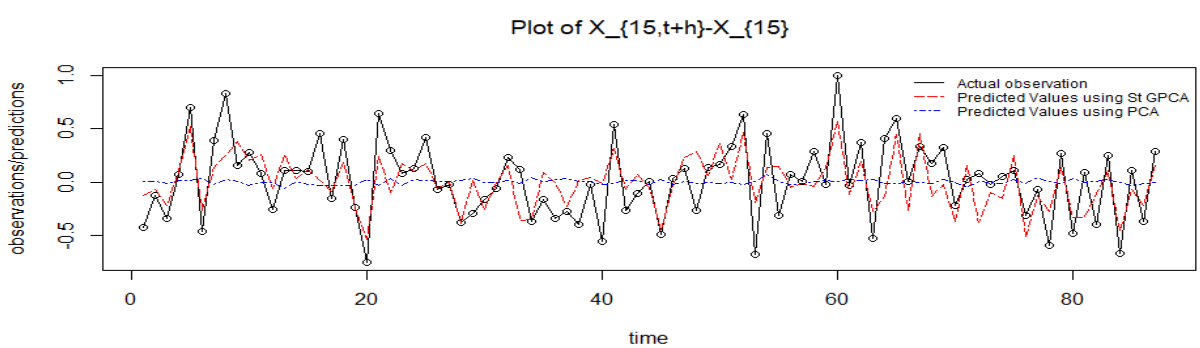
(a) Predicted values vs Actual observation ($h = 12$)



(b) Predicted values vs Actual observation ($h = 12$)



(c) Predicted values vs Actual observation ($h = 12$)



(d) Predicted values vs Actual observation ($h = 12$)

A.6 R Codes

Some part of these codes are based on the E-views codes provided by Charles Engel [1] related to the paper Engel et al. [2015].

A.6.1 The Normal VAR Simulation Design and Forecasting

```
options(tol=10e-40)
library(psych); library(zoo); library(dynlm); library(graphics); library(aod)
library(Quandl); library(nortest); library(car);library(foreign)
library(tidyr); library(nFactors); library(fBasics); library(far)
library(Matrix); library(MCMCpack); library(Hmisc); library(ADGofTest)
library(numDeriv); library(grDevices); library(StVAR); library(stats)
library(mvtnorm); library(plyr); library(reshape2)

############################        Data Generating
set.seed(1234)
phi0<-1.8
a<-0.8
sigmat<-matrix(c(0.072253514,0.029550653,0.018048041,0.030974202,0.035580663,
0.063596492,-0.044353946,-0.023820021,0.007845989,0.031214058,
-0.021647049,0.08288506,0.084255886,0.036116467,-0.015758023,
0.029550653,0.01679944,0.011098948,0.014289844,0.016592454,
0.027956533,-0.015018814,-0.005433569,0.006380243,0.015463976,
-0.004629422,0.037085423,0.037605746,0.018340162,-0.002218735,
0.018048041,0.011098948,0.032512655,0.019745562,0.022764677,
0.028123621,0.016547583,0.022492343,0.031449585,0.033754869,
0.023350481,0.037365467,0.033886629,0.023088821,0.025034264,
0.030974202,0.014289844,0.019745562,0.024720436,0.021428559,
0.033187292,-0.007956688,0.00132638,0.016101257,0.025557668,
0.002543218,0.044172615,0.044523807,0.022379023,0.005155761,
0.035580663,0.016592454,0.022764677,0.021428559,0.026619446,
0.039854456,-0.006240459,0.003382064,0.019284845,0.028382497,
0.004487382,0.050225972,0.048577871,0.026500612,0.007327197,
0.063596492,0.027956533,0.028123621,0.033187292,0.039854456,
0.069823393,-0.029664082,-0.010147566,0.019770641,0.039342479,
-0.008072806,0.087829782,0.087412774,0.042590271,-0.00259529,
```

[1] https://www.ssc.wisc.edu/~cengel/Data/Factor/FactorData.htm

```
-0.044353946,-0.015018814,0.016547583,-0.007956688,-0.006240459,
-0.029664082,0.071597342,0.053611242,0.028972791,0.005395715,
0.052089757,-0.043365436,-0.049071759,-0.006508762,0.046979943,
-0.023820021,-0.005433569,0.022492343,0.00132638,0.003382064,
-0.010147566,0.053611242,0.045146562,0.030542163,0.015094679,
0.044348213,-0.015764937,-0.020798273,0.004410417,0.041410615,
0.007845989,0.006380243,0.031449585,0.016101257,0.019284845,
0.019770641,0.028972791,0.030542163,0.03384728,0.030772889,
0.030988106,0.024702971,0.020158578,0.019159546,0.031331137,
0.031214058,0.015463976,0.033754869,0.025557668,0.028382497,
0.039342479,0.005395715,0.015094679,0.030772889,0.03845654,
0.016369007,0.05235939,0.049424261,0.028235865,0.019238997,
-0.021647049,-0.004629422,0.023350481,0.002543218,0.004487382,
-0.008072806,0.052089757,0.044348213,0.030988106,0.016369007,
0.043707141,-0.01286569,-0.017712702,0.005494672,0.040962468,
0.08288506,0.037085423,0.037365467,0.044172615,0.050225972,
0.087829782,-0.043365436,-0.015764937,0.024702971,0.05235939,
-0.01286569,0.116828351,0.115920941,0.053252761,-0.005108878,
0.084255886,0.037605746,0.033886629,0.044523807,0.048577871,
0.087412774,-0.049071759,-0.020798273,0.020158578,0.049424261,
-0.017712702,0.115920941,0.118666879,0.052494625,-0.009870713,
0.036116467,0.018340162,0.023088821,0.022379023,0.026500612,
0.042590271,-0.006508762,0.004410417,0.019159546,0.028235865,
0.005494672,0.053252761,0.052494625,0.031352219,0.008293733,
-0.015758023,-0.002218735,0.025034264,0.005155761,0.007327197,
-0.00259529,0.046979943,0.041410615,0.031331137,0.019238997,
0.040962468,-0.005108878,-0.009870713,0.008293733,0.038942027)
,nrow=15,ncol=15)
sigma<-kronecker(matrix(c(phi0,phi0*a,phi0*a,phi0),nrow = 2,ncol = 2),sigmat)
meann<-c(2.5,1.9,0.8,0.5,1.3,0.9,3.4,2.3,0.3,0.08,4.5,3.7,1.4,2.9,0.001,
2.5,1.9,0.8,0.5,1.3,0.9,3.4,2.3,0.3,0.08,4.5,3.7,1.4,2.9,0.001)
X = rmvnorm(n=250, mean=meann, sigma=sigma, method="chol")
x1=X[,1]; x2=X[,2]; x3=X[,3]; x4=X[,4]; x5=X[,5]; x6=X[,6]; x7=X[,7]; x8=X[,8];
x9=X[,9]; x10=X[,10]; x11=X[,11]; x12=X[,12]; x13=X[,13]; x14=X[,14]; x15=X[,15];
lx1=X[,16]; lx2=X[,17]; lx3=X[,18]; lx4=X[,19]; lx5=X[,20]; lx6=X[,21]; lx7=X[,22]; lx8=X[,23]
; lx9=X[,24]; lx10=X[,25]; lx11=X[,26]; lx12=X[,27]; lx13=X[,28]; lx14=X[,29]; lx15=X[,30]

Xmat<-data.frame(x1,x2,x3,x4,x5,x6,x7,x8,x9,x10,x11,x12,x13,x14,x15)
mydata<-as.matrix(Xmat)
```

```

#################### Set parameter values
FF=3; NN = 15; R=dim(mydata)[1]; tst = 150; hrzn<-c(1,4,8,12); lh=length(hrzn); P=(R-tst-1)
#################################### Constructing matrices and series
TheilU_CW_statistic<-TheilU_CW_statistic.pc<-matrix(NA,NN,2*lh)

rownames(TheilU_CW_statistic)<-rownames(TheilU_CW_statistic.pc)<-c(colnames(mydata[,1:15]))
colnames(TheilU_CW_statistic)<-colnames(TheilU_CW_statistic.pc)<-c("U stat, h=1","U stat,
h=4","U stat, h=8","U stat, h=12", "CW stat, h=1","CW stat, h=4","CW stat, h=8","CW stat, h=12")

Yhat<-Yhat.pc<-matrix(0,dim(mydata)[1],NN*lh)
pred_error_factor_all <- pred_error_factor_all.pc <- pred_error_rw_all<- matrix(0,dim(mydata)[1],NN*lh)
MSPEadj<-MSPEadj.pc<-c(rep(0,NN))
mydatape<-matrix(NaN,dim(mydata)[1],NN)

for(hh in 1:lh){
  k=hrzn[hh]
  tnd=(R-1)

  c<-c(rep(0,1000))
  loads<-loads.pc<-matrix(NA,NN,3)
  rownames(loads)<-rownames(loads.pc)<-colnames(mydata)
  colnames(loads)<-colnames(loads.pc)<-cbind("Load1","Load2","Load3")
  for(t in tst:tnd){
    mydatagpca<-mydata[1:(1+t),]
    for (i in 1:NN) {
      mydatagpca[,i]<-as.matrix(mydatagpca[,i])-mean(as.matrix(mydatagpca[,i]))
    }
    ########## GPCs

    B<-eigen(cov(t(mydatagpca)))$vectors
    A<-eigen(cov(mydatagpca))$vectors[,1:3]
    sc<-t(t(A)%*%t(mydatagpca)%*%B)
    pc<-t(t(A)%*%t(mydatagpca))
    rownames(sc)<-rownames(pc)<-rownames(mydata[1:(1+t),])
    colnames(sc)<-cbind("GPC1","GPC2","GPC3")
    colnames(pc)<-cbind("PC1","PC2","PC3")
```

```r

    ##################### Revised contemporaneous covariance matrix
    #We can use the MLE of contemporaneous covariance matrix as well.
    #However, when we obtain a statistically adequate model, the sample covariance of order
    #T*T would be fine because in the GLS-type regression (if we have non-Gaussian distribution)
    #The heteroskedastic standard error will do the same as MLE. In this simulation I found out the
    #results from using sample covariance and MLE covariance is the same up to three decimals which can
    #be due to the rounding.
    # M<-matrix(NaN,(fp+t),(fp+t))
    # phi0<-1.8
    # a<-0.8
    # e<-matrix(1,(fp+t),1)
    # for(pp in 1:(fp+t)){
    #   for(qq in 1:(fp+t)){
    #     M[pp,qq]<-phi0*(a^(abs(qq-pp)))
    #   }
    # }
    # Qn<-(inv(M)%*%e%*%t(e)%*%inv(M))
    # Qd<-as.numeric(1/(t(e)%*%inv(M)%*%e))
    # Qa<-Qn*Qd
    # hatsigma<-0.004*t(mydatagpca)%*%(inv(M)-Qa)%*%mydatagpca
    #################################
    for(i in 1:NN)  {
      factorfit<-lm(mydatagpca[,i]~sc)
      loads[i,]<-factorfit$coefficients[2:4]
      factorfit.pc<-lm(mydatagpca[,i]~pc)
      loads.pc[i,]<-factorfit.pc$coefficients[2:4]
    }

  # constructing regressors F(it)-s(it) for 1,...,F factors, i=1,...,NN

    FactorX<-FactorX.pc<-matrix(NA,1+t,NN)
    rownames(FactorX)<-rownames(FactorX.pc)<-rownames(mydata[1:(1+t),])
    colnames(FactorX)<-colnames(FactorX.pc)<-colnames(mydata[,1:NN])
    Ymat<-matrix(NA,(1+t),NN)
    rownames(Ymat)<-rownames(mydata[1:(1+t),])
    colnames(Ymat)<-colnames(mydata[,1:NN])
    for (j in 1:NN){
```

```
      FactorX[,j]=-mydatagpca[,j]
      FactorX.pc[,j]=-mydatagpca[,j]
      for(f in 1:FF){
        FactorX[,j]=FactorX[,j]+(loads[j,f]*sc[,f])
        FactorX.pc[,j]=FactorX.pc[,j]+(loads.pc[j,f]*pc[,f])
      }

      Ymat[,j]<-mydatagpca[,j]-Lag(mydatagpca[,j],shift = k)
    }
    FactorLX<-FactorLX.pc<-matrix(NaN,dim(FactorX)[1],dim(FactorX)
       [2])
    for(j in 1:NN){
      FactorLX[,j]<-Lag(FactorX[,j],shift = k)
      FactorLX.pc[,j]<-Lag(FactorX.pc[,j],shift = k)
    }

    FactorLXlong<-melt(FactorLX)
    FactorLXlong.pc<-melt(FactorLX.pc)
    Ylong<-melt(Ymat)

    Y_FactorLX <- cbind(Ylong,FactorLXlong[,3])
    Y_FactorLX.pc <- cbind(Ylong,FactorLXlong.pc[,3])
    colnames(Y_FactorLX) <- c("time","variables","Y","gpcX")
    colnames(Y_FactorLX.pc) <- c("time","variables","Y","pcX")

    LRMFit <- lm(Y ~ gpcX+factor(variables)-1,data = Y_FactorLX)
    LRMFit.pc <- lm(Y ~ pcX+factor(variables)-1,data = Y_FactorLX.pc
       )

    c[601:615]<-LRMFit$coefficients[2:16]; c[650]<-LRMFit$
       coefficients[1]

    c[401:415]<-LRMFit$coefficients[2:16]; c[450]<-LRMFit$
       coefficients[1]

    for(l in 1:NN){
      Yhat[(1+t),l+((hh-1)*NN)]=c[600+l]+c[650]*FactorX[(1+t),l]
      Yhat.pc[(1+t),l+((hh-1)*NN)]=c[400+l]+c[450]*FactorX.pc[(1+t),
         l]
    }
  }

  ##################################### Forecast evaluation
  ti1=(1+tst)
  ti2=R
```

```
pred_error_factor<-pred_error_factor.pc<-matrix(0,dim(mydata)[1],
    NN)
pred_error_rw<-matrix(0,dim(mydata)[1],NN)
SPE_SPEAdj<-SPE_SPEAdj.pc<-matrix(NA,dim(mydata)[1],NN)
for(o in 1:NN){
  mydatape[1:tst,o]<-as.matrix(mydata[1:tst,o])-mean(as.matrix(
      mydata[1:tst,o]))
  for(t in tst:tnd){
    C<-as.matrix(mydata[1:t,o])-mean(as.matrix(mydata[1:t,o]))
    mydatape[t,o]<-C[t]
 }
  pred_error_factor[ti1:ti2,o]<-(Lag(mydatape[ti1:ti2,o],shift = -
      k)-mydatape[ti1:ti2,o])-Yhat[ti1:ti2,o+((hh-1)*NN)]
  pred_error_factor.pc[ti1:ti2,o]<-(Lag(mydatape[ti1:ti2,o],shift
      = -k)-mydatape[ti1:ti2,o])-Yhat.pc[ti1:ti2,o+((hh-1)*NN)]

  pred_error_rw[ti1:ti2,o]<-Lag(mydatape[ti1:ti2,o],shift = -k)-
      mydatape[ti1:ti2,o]

  pred_error_factor_all[,(((hh-1)*NN)+1):(hh*NN)]<-pred_error_
      factor
  pred_error_factor_all.pc[,(((hh-1)*NN)+1):(hh*NN)]<-pred_error_
      factor.pc
  pred_error_rw_all[,(((hh-1)*NN)+1):(hh*NN)]<-pred_error_rw

  SPE_Factor<-pred_error_factor[,o]*pred_error_factor[,o]
  SPE_Factor.pc<-pred_error_factor.pc[,o]*pred_error_factor.pc[,o]
  SPE_rw<-pred_error_rw[,o]*pred_error_rw[,o]

  SPE_SPEAdj[ti1:ti2,o]<-(SPE_rw[ti1:ti2]-SPE_Factor[ti1:ti2])
   +Lag(Yhat[ti1:ti2,o+((hh-1)*NN)],shift = -k)*Lag(Yhat[ti1:ti2,o
       +((hh-1)*NN)],shift = -k)
  SPE_SPEAdj.pc[ti1:ti2,o]<-(SPE_rw[ti1:ti2]-SPE_Factor.pc[ti1:ti2
      ])
  +Lag(Yhat.pc[ti1:ti2,o+((hh-1)*NN)],shift = -k)*Lag(Yhat.pc[ti1:
      ti2,o+((hh-1)*NN)],shift = -k)

  MSPEadj[o]<-mean(SPE_SPEAdj[,o],na.rm=TRUE)
  MSPEadj.pc[o]<-mean(SPE_SPEAdj.pc[,o],na.rm=TRUE)
  TheilU_CW_statistic[o,hh]<-(mean(SPE_Factor,na.rm=TRUE)/mean(SPE
      _rw,na.rm=TRUE))^0.5
  TheilU_CW_statistic.pc[o,hh]<-(mean(SPE_Factor.pc,na.rm=TRUE)/
      mean(SPE_rw,na.rm=TRUE))^0.5
}
```

```
  #Univariate case: Standard errors and CW stats
    P1=P-k+1
    P2=P-(2*(k-1))
    t_1=1+tst
    t_2=dim(mydata)[1]-k+1

  Yhatrec<-Yhatrec.pc<-matrix(0,dim(mydata)[1],NN)
  dist_adj<-dist_adj.pc<-matrix(NA,dim(mydata)[1],NN)
  mean_dist<-mean_dist.pc<-c(rep(0,NN))
  sq_dist_adj<-sq_dist_adj.pc<-c(rep(0,NN))
  CW_statistic<-CW_statistic.pc<-c(rep(0,NN))
  mean_dist_cent<-mean_dist_cent.pc<-matrix(NA,dim(mydata)[1],NN)

  for(jj in 1:NN){
    for(g in 1:k){
       Yhatrec[(t_1:t_2),jj]<-Yhatrec[(t_1:t_2),jj]+Lag(Yhat[(t_1:t_2),jj+((hh-1)*NN)],shift = g)
       Yhatrec.pc[(t_1:t_2),jj]<-Yhatrec.pc[(t_1:t_2),jj]+Lag(Yhat.pc[(t_1:t_2),jj+((hh-1)*NN)],shift = g)
    }
    dist_adj[(t_1:t_2),jj]<-2*(mydatape[(t_1:t_2),jj]-Lag(mydatape[(t_1:t_2),jj],shift = 1))*Yhatrec[(t_1:t_2),jj]
    mean_dist[jj]<-mean(dist_adj[,jj],na.rm=TRUE)
    mean_dist_cent[(t_1:t_2),jj]<-dist_adj[(t_1:t_2),jj]-mean_dist[jj]
    sq_dist_adj[jj]<-(1/P2)*sum(mean_dist_cent[,jj]^2,na.rm = TRUE)

    dist_adj.pc[(t_1:t_2),jj]<-2*(mydatape[(t_1:t_2),jj]-Lag(mydatape[(t_1:t_2),jj],shift = 1))*Yhatrec.pc[(t_1:t_2),jj]
    mean_dist.pc[jj]<-mean(dist_adj.pc[,jj],na.rm=TRUE)
    mean_dist_cent.pc[(t_1:t_2),jj]<-dist_adj.pc[(t_1:t_2),jj]-mean_dist.pc[jj]
    sq_dist_adj.pc[jj]<-(1/P2)*sum(mean_dist_cent.pc[,jj]^2,na.rm = TRUE)

    ###################Univariate Clark-West stats
    CW_statistic[jj]<-sqrt(P1)*(MSPEadj[jj]/sqrt(sq_dist_adj[jj]))
    CW_statistic.pc[jj]<-sqrt(P1)*(MSPEadj.pc[jj]/sqrt(sq_dist_adj.pc[jj]))
    TheilU_CW_statistic[jj,hh+lh]=CW_statistic[jj]
    TheilU_CW_statistic.pc[jj,hh+lh]=CW_statistic.pc[jj]
  }
}
```

A.6.2 The Student's t VAR Simulation Design and Forecasting

```
options(tol=10e-40)
library(psych); library(zoo); library(dynlm); library(graphics); library(aod)
library(Quandl); library(nortest); library(car);library(foreign)
library(tidyr); library(nFactors); library(fBasics); library(far)
library(Matrix); library(MCMCpack); library(Hmisc); library(ADGofTest)
library(numDeriv); library(grDevices); library(StVAR); library(stats)
library(mvtnorm); library(plyr); library(reshape2); library(dummies)

############################            Data Generating
set.seed(1234)
phi0<-1.8
a<-0.8
sigmat<-matrix(c(0.072253514,0.029550653,0.018048041,0.030974202,0.035580663,
0.063596492,-0.044353946,-0.023820021,0.007845989,0.031214058,
-0.021647049,0.08288506,0.084255886,0.036116467,-0.015758023,
0.029550653,0.01679944,0.011098948,0.014289844,0.016592454,
0.027956533,-0.015018814,-0.005433569,0.006380243,0.015463976,
-0.004629422,0.037085423,0.037605746,0.018340162,-0.002218735,
0.018048041,0.011098948,0.032512655,0.019745562,0.022764677,
0.028123621,0.016547583,0.022492343,0.031449585,0.033754869,
0.023350481,0.037365467,0.033886629,0.023088821,0.025034264,
0.030974202,0.014289844,0.019745562,0.024720436,0.021428559,
0.033187292,-0.007956688,0.00132638,0.016101257,0.025557668,
0.002543218,0.044172615,0.044523807,0.022379023,0.005155761,
0.035580663,0.016592454,0.022764677,0.021428559,0.026619446,
0.039854456,-0.006240459,0.003382064,0.019284845,0.028382497,
0.004487382,0.050225972,0.048577871,0.026500612,0.007327197,
0.063596492,0.027956533,0.028123621,0.033187292,0.039854456,
0.069823393,-0.029664082,-0.010147566,0.019770641,0.039342479,
-0.008072806,0.087829782,0.087412774,0.042590271,-0.00259529,
-0.044353946,-0.015018814,0.016547583,-0.007956688,-0.006240459,
-0.029664082,0.071597342,0.053611242,0.028972791,0.005395715,
0.052089757,-0.043365436,-0.049071759,-0.006508762,0.046979943,
-0.023820021,-0.005433569,0.022492343,0.00132638,0.003382064,
-0.010147566,0.053611242,0.045146562,0.030542163,0.015094679,
0.044348213,-0.015764937,-0.020798273,0.004410417,0.041410615,
0.007845989,0.006380243,0.031449585,0.016101257,0.019284845,
0.019770641,0.028972791,0.030542163,0.03384728,0.030772889,
0.030988106,0.024702971,0.020158578,0.019159546,0.031331137,
```

```
0.031214058,0.015463976,0.033754869,0.025557668,0.028382497,
0.039342479,0.005395715,0.015094679,0.030772889,0.03845654,
0.016369007,0.05235939,0.049424261,0.028235865,0.019238997,
-0.021647049,-0.004629422,0.023350481,0.002543218,0.004487382,
-0.008072806,0.052089757,0.044348213,0.030988106,0.016369007,
0.043707141,-0.01286569,-0.017712702,0.005494672,0.040962468,
0.08288506,0.037085423,0.037365467,0.044172615,0.050225972,
0.087829782,-0.043365436,-0.015764937,0.024702971,0.05235939,
-0.01286569,0.116828351,0.115920941,0.053252761,-0.005108878,
0.084255886,0.037605746,0.033886629,0.044523807,0.048577871,
0.087412774,-0.049071759,-0.020798273,0.020158578,0.049424261,
-0.017712702,0.115920941,0.118666879,0.052494625,-0.009870713,
0.036116467,0.018340162,0.023088821,0.022379023,0.026500612,
0.042590271,-0.006508762,0.004410417,0.019159546,0.028235865,
0.005494672,0.053252761,0.052494625,0.031352219,0.008293733,
-0.015758023,-0.002218735,0.025034264,0.005155761,0.007327197,
-0.00259529,0.046979943,0.041410615,0.031331137,0.019238997,
0.040962468,-0.005108878,-0.009870713,0.008293733,0.038942027)
,nrow=15,ncol=15)
sigma<-kronecker(matrix(c(phi0,phi0*a,phi0*a,phi0),nrow = 2,ncol = 2),sigmat)
meann<-c(2.5,1.9,0.8,0.5,1.3,0.9,3.4,2.3,0.3,0.08,4.5,3.7,1.4,2.9,0.001,
2.5,1.9,0.8,0.5,1.3,0.9,3.4,2.3,0.3,0.08,4.5,3.7,1.4,2.9,0.001)
X = rmvt(n=250, sigma=sigma, df=30, delta=meann,type="shifted")
x1=X[,1]; x2=X[,2]; x3=X[,3]; x4=X[,4]; x5=X[,5]; x6=X[,6]; x7=X[,7]; x8=X[,8];
x9=X[,9]; x10=X[,10]; x11=X[,11]; x12=X[,12]; x13=X[,13]; x14=X[,14]; x15=X[,15];
lx1=X[,16]; lx2=X[,17]; lx3=X[,18]; lx4=X[,19]; lx5=X[,20]; lx6=X[,21]; lx7=X[,22]; lx8=X[,23]
; lx9=X[,24]; lx10=X[,25]; lx11=X[,26]; lx12=X[,27]; lx13=X[,28]; lx14=X[,29]; lx15=X[,30]

Xmat<-data.frame(x1,x2,x3,x4,x5,x6,x7,x8,x9,x10,x11,x12,x13,x14,x15)
mydata<-as.matrix(Xmat)

#################### Set parameter values
FF=3; NN = 15; R=dim(mydata)[1]; tst = 150; hrzn<-c(1,4,8,12); lh=length(hrzn); P=(R-tst-1)
#####################################  Constructing matrices and series
TheilU_CW_statistic<-TheilU_CW_statistic.pc<-matrix(NA,NN,2*lh)

```

```
rownames(TheilU_CW_statistic)<-rownames(TheilU_CW_statistic.pc)<-c(
   colnames(mydata[,1:15]))
colnames(TheilU_CW_statistic)<-colnames(TheilU_CW_statistic.pc)<-c("
   U stat, h=1","U stat,
h=4","U stat, h=8","U stat, h=12", "CW stat, h=1","CW stat, h=4","CW
    stat, h=8","CW stat, h=12")
Yhat<-Yhat.pc<-matrix(0,dim(mydata)[1],NN*lh)

pred_error_factor_all <- pred_error_factor_all.pc <- pred_error_rw_
   all<- matrix(0,dim(mydata)[1],NN*lh)

MSPEadj<-MSPEadj.pc<-c(rep(0,NN))

mydatape<-matrix(NaN,dim(mydata)[1],NN)

for(hh in 1:lh){
  k=hrzn[hh]
  tnd=(R-1)

  c<-c(rep(0,1000))
  loads<-loads.pc<-matrix(NA,NN,3)
  rownames(loads)<-rownames(loads.pc)<-colnames(mydata)
  colnames(loads)<-colnames(loads.pc)<-cbind("Load1","Load2","Load3"
    )
  for(t in tst:tnd){
    mydatagpca<-mydata[1:(1+t),]
    for (i in 1:NN) {
      mydatagpca[,i]<-as.matrix(mydatagpca[,i])-mean(as.matrix(
         mydatagpca[,i]))
    }

    B<-eigen(cov(t(mydatagpca)))$vectors
    A<-eigen(cov(mydatagpca))$vectors[,1:3]
    sc<-t(t(A)%*%t(mydatagpca)%*%B)
    pc<-t(t(A)%*%t(mydatagpca))
    rownames(sc)<-rownames(pc)<-rownames(mydata[1:(1+t),])
    colnames(sc)<-cbind("GPC1","GPC2","GPC3")
    colnames(pc)<-cbind("PC1","PC2","PC3")

    for(i in 1:NN)  {
      factorfit<-lm(mydatagpca[,i]~sc)
      loads[i,]<-factorfit$coefficients[2:4]
      factorfit.pc<-lm(mydatagpca[,i]~pc)
      loads.pc[i,]<-factorfit.pc$coefficients[2:4]
    }
```

```
# constructing regressors F(it)-s(it) for 1,...,F factors, i
    =1,...,NN

  FactorX<-FactorX.pc<-matrix(NA,1+t,NN)
  rownames(FactorX)<-rownames(FactorX.pc)<-rownames(mydata[1:(1+t)
      ,])
  colnames(FactorX)<-colnames(FactorX.pc)<-colnames(mydata[,1:NN])
  Ymat<-matrix(NA,(1+t),NN)
  rownames(Ymat)<-rownames(mydata[1:(1+t),])
  colnames(Ymat)<-colnames(mydata[,1:NN])
  for (j in 1:NN){
    FactorX[,j]=-mydatagpca[,j]
    FactorX.pc[,j]=-mydatagpca[,j]
    for(f in 1:FF){
      FactorX[,j]=FactorX[,j]+(loads[j,f]*sc[,f])
      FactorX.pc[,j]=FactorX.pc[,j]+(loads.pc[j,f]*pc[,f])
    }

    Ymat[,j]<-mydatagpca[,j]-Lag(mydatagpca[,j],shift = k)
  }
  FactorLX<-FactorLX.pc<-matrix(NaN,dim(FactorX)[1],dim(FactorX)
      [2])
  for(j in 1:NN){
    FactorLX[,j]<-Lag(FactorX[,j],shift = k)
    FactorLX.pc[,j]<-Lag(FactorX.pc[,j],shift = k)
  }

  FactorLXlong<-melt(FactorLX)
  FactorLXlong.pc<-melt(FactorLX.pc)
  Ylong<-melt(Ymat)

  Y_FactorLX <- cbind(Ylong,FactorLXlong[,3])
  Y_FactorLX.pc <- cbind(Ylong,FactorLXlong.pc[,3])
  colnames(Y_FactorLX) <- c("time","variables","Y","gpcX")
  colnames(Y_FactorLX.pc) <- c("time","variables","Y","pcX")

  y = Y_FactorLX$Y ; X = cbind(Y_FactorLX$gpcX)
  Trendd = cbind(dummy(Y_FactorLX$variables))
  XX <- na.omit(cbind(y,X,Trendd))
  y1 <- XX[,1] ; X1 <-as.matrix(XX[,2]) ; Trend1 <- XX[,3:17]
  colnames(Trend1) <- colnames(mydata)[1:15]; colnames(X1) <- "
      gpcX" ; lag <- 0 ; ll <- ncol(X1)

  LRMFit <- StDLRM(y1, X1 ,v=30,Trend=Trend1,lag=0,hes="TRUE")
```

```r
    LRMFit.pc <- lm(Y ~ pcX+factor(variables)-1,data = Y_FactorLX.pc)

    c[601:615]<-LRMFit$coefficients[1:15]; c[650]<-LRMFit$coefficients[16]

    c[401:415]<-LRMFit$coefficients[2:16]; c[450]<-LRMFit$coefficients[1]

    for(l in 1:NN){
      Yhat[(1+t),l+((hh-1)*NN)]=c[600+l]+c[650]*FactorX[(1+t),l]
      Yhat.pc[(1+t),l+((hh-1)*NN)]=c[400+l]+c[450]*FactorX.pc[(1+t),l]
    }

  }

  ##################################### Forecast evaluation
  ti1=(1+tst)
  ti2=R

  pred_error_factor<-pred_error_factor.pc<-matrix(0,dim(mydata)[1],NN)
  pred_error_rw<-matrix(0,dim(mydata)[1],NN)
  SPE_SPEAdj<-SPE_SPEAdj.pc<-matrix(NA,dim(mydata)[1],NN)
  for(o in 1:NN){
    mydatape[1:tst,o]<-as.matrix(mydata[1:tst,o])-mean(as.matrix(mydata[1:tst,o]))
    for(t in tst:tnd){
      C<-as.matrix(mydata[1:t,o])-mean(as.matrix(mydata[1:t,o]))
      mydatape[t,o]<-C[t]
   }
    pred_error_factor[ti1:ti2,o]<-(Lag(mydatape[ti1:ti2,o],shift = -k)-mydatape[ti1:ti2,o])-Yhat[ti1:ti2,o+((hh-1)*NN)]
    pred_error_factor.pc[ti1:ti2,o]<-(Lag(mydatape[ti1:ti2,o],shift = -k)-mydatape[ti1:ti2,o])-Yhat.pc[ti1:ti2,o+((hh-1)*NN)]

    pred_error_rw[ti1:ti2,o]<-Lag(mydatape[ti1:ti2,o],shift = -k)-mydatape[ti1:ti2,o]

    pred_error_factor_all[,(((hh-1)*NN)+1):(hh*NN)]<-pred_error_factor
    pred_error_factor_all.pc[,(((hh-1)*NN)+1):(hh*NN)]<-pred_error_factor.pc
    pred_error_rw_all[,(((hh-1)*NN)+1):(hh*NN)]<-pred_error_rw
```

```

    SPE_Factor<-pred_error_factor[,o]*pred_error_factor[,o]
    SPE_Factor.pc<-pred_error_factor.pc[,o]*pred_error_factor.pc[,o]
    SPE_rw<-pred_error_rw[,o]*pred_error_rw[,o]

    SPE_SPEAdj[ti1:ti2,o]<-(SPE_rw[ti1:ti2]-SPE_Factor[ti1:ti2])
     +Lag(Yhat[ti1:ti2,o+((hh-1)*NN)],shift = -k)*Lag(Yhat[ti1:ti2,o+((hh-1)*NN)],shift = -k)
    SPE_SPEAdj.pc[ti1:ti2,o]<-(SPE_rw[ti1:ti2]-SPE_Factor.pc[ti1:ti2])
    +Lag(Yhat.pc[ti1:ti2,o+((hh-1)*NN)],shift = -k)*Lag(Yhat.pc[ti1:ti2,o+((hh-1)*NN)],shift = -k)

    MSPEadj[o]<-mean(SPE_SPEAdj[,o],na.rm=TRUE)
    MSPEadj.pc[o]<-mean(SPE_SPEAdj.pc[,o],na.rm=TRUE)
    TheilU_CW_statistic[o,hh]<-(mean(SPE_Factor,na.rm=TRUE)/mean(SPE_rw,na.rm=TRUE))^0.5
    TheilU_CW_statistic.pc[o,hh]<-(mean(SPE_Factor.pc,na.rm=TRUE)/mean(SPE_rw,na.rm=TRUE))^0.5
  }
  #Univariate case: Standard errors and CW stats
    P1=P-k+1
    P2=P-(2*(k-1))
    t_1=1+tst
    t_2=dim(mydata)[1]-k+1

  Yhatrec<-Yhatrec.pc<-matrix(0,dim(mydata)[1],NN)
  dist_adj<-dist_adj.pc<-matrix(NA,dim(mydata)[1],NN)
  mean_dist<-mean_dist.pc<-c(rep(0,NN))
  sq_dist_adj<-sq_dist_adj.pc<-c(rep(0,NN))
  CW_statistic<-CW_statistic.pc<-c(rep(0,NN))
  mean_dist_cent<-mean_dist_cent.pc<-matrix(NA,dim(mydata)[1],NN)

  for(jj in 1:NN){
    for(g in 1:k){
       Yhatrec[(t_1:t_2),jj]<-Yhatrec[(t_1:t_2),jj]+Lag(Yhat[(t_1:t_2),jj+((hh-1)*NN)],shift = g)
       Yhatrec.pc[(t_1:t_2),jj]<-Yhatrec.pc[(t_1:t_2),jj]+Lag(Yhat.pc[(t_1:t_2),jj+((hh-1)*NN)],shift = g)
    }
    dist_adj[(t_1:t_2),jj]<-2*(mydatape[(t_1:t_2),jj]-Lag(mydatape[(t_1:t_2),jj],shift = 1))*Yhatrec[(t_1:t_2),jj]
    mean_dist[jj]<-mean(dist_adj[,jj],na.rm=TRUE)
    mean_dist_cent[(t_1:t_2),jj]<-dist_adj[(t_1:t_2),jj]-mean_dist[jj]
```

```
    sq_dist_adj[jj]<-(1/P2)*sum(mean_dist_cent[,jj]^2,na.rm = TRUE)

    dist_adj.pc[(t_1:t_2),jj]<-2*(mydatape[(t_1:t_2),jj]-Lag(mydatape[(t_1:t_2),jj],shift = 1))*Yhatrec.pc[(t_1:t_2),jj]
    mean_dist.pc[jj]<-mean(dist_adj.pc[,jj],na.rm=TRUE)
    mean_dist_cent.pc[(t_1:t_2),jj]<-dist_adj.pc[(t_1:t_2),jj]-mean_dist.pc[jj]
    sq_dist_adj.pc[jj]<-(1/P2)*sum(mean_dist_cent.pc[,jj]^2,na.rm = TRUE)

    #Univariate Clark-West stats
    CW_statistic[jj]<-sqrt(P1)*(MSPEadj[jj]/sqrt(sq_dist_adj[jj]))
    CW_statistic.pc[jj]<-sqrt(P1)*(MSPEadj.pc[jj]/sqrt(sq_dist_adj.pc[jj]))
    TheilU_CW_statistic[jj,hh+lh]=CW_statistic[jj]
    TheilU_CW_statistic.pc[jj,hh+lh]=CW_statistic.pc[jj]
  }

}
```

A.6.3 Exchange Rate Forecasting

```
options(width=60, keep.space=TRUE, scipen = 999)
library(plm); library(psych); library(zoo); library(nlme); library(
    dynlm); library(graphics)
library(aod); library(foreign); library(mvtnorm); library(Quandl);
    library(ConvergenceConcepts)
library(tseries); library(nortest); library(car); library(tidyr);
    library(nFactors); library(quantmod)
library(fBasics); library(far); library(ADGofTest); library(matlab);
     library(rms); library(ggplot2)
library(Hmisc); library(ggpubr); library(Matrix); library(forecast);
     library(MCMCpack); library(numDeriv)
library(grDevices); library(rgl); library(heavy); library(glmnet);
    library(rpart)
library(randomForest); library(leaps); library(rpart.plot); library(
    mFilter)
library(stats); library(tidyr); library(reshape2)

################# PROCESSING THE DATA
setwd("PATH")
mydata<-read.csv("Data - Updated.csv",header = TRUE)
#################### Set parameter values
S=3 #1=early sample (pre Euro), 2=late smpl (post), 3=full smpl
EUR=1  #1 if forecast of Euro is needed
FF=3; NN = 17; R=dim(mydata)[1]; tst = 55; hrzn<-c(1,4,8,12); lh=
    length(hrzn)

if(S==1){
  P=49
}
if(S==2){
  P=75
}
if(S==3){
  P=(R-tst-1)
}
##################################### 3. Constructing matrices and
    series
TheilU_CW_statistic<-matrix(NA,NN,2*lh)

rownames(TheilU_CW_statistic)<-c(colnames(mydata[,1:17]))
colnames(TheilU_CW_statistic)<-c("U stat, h=1","U stat,h=4","U stat,
     h=8","U stat, h=12","CW stat, h=1","CW stat, h=4","CW stat, h=8"
    ,"CW stat, h=12")
```

```r
Yhat<-matrix(0,dim(mydata)[1],NN*lh); Yhat_euro<-matrix(0,dim(mydata)[1],lh); Yhateuro<-matrix(0,dim(mydata)[1],lh)

pred_error_factor_all <- pred_error_rw_all<- matrix(0,dim(mydata)[1],NN*lh)

MSPEadj<-c(rep(0,NN))

mydatape<-matrix(NaN,dim(mydata)[1],NN)

for(hh in 1:lh){
  k=hrzn[hh]

  if (S==1){
    tnd=(tst+P-1)
  }
  if (S==2){
    tst=104
    tnd=(tst+P-1)
  }
  if (S==3 && EUR==1){
    tnd=(R-1)
  }

  c<-c(rep(0,1000))
  loads<-matrix(NA,NN,3)
  rownames(loads)<-colnames(mydata[,1:17])
  colnames(loads)<-cbind("Load1","Load2","Load3")
  for(t in tst:tnd){
    mydatagpca<-mydata[1:(1+t),]
    for (i in 1:NN) {
      mydatagpca[,i]<-as.matrix(mydatagpca[,i])-mean(as.matrix(mydatagpca[,i]))
    }

    B<-eigen(cov(t(mydatagpca)))$vectors
    A<-eigen(cov(mydatagpca))$vectors[,1:3]
    sc<-t(t(A)%*%t(mydatagpca)%*%B)
    rownames(sc)<-rownames(mydata[1:(1+t),])
    colnames(sc)<-cbind("GPC1","GPC2","GPC3")
    for(i in 1:NN)  {
      factorfit<-lm(mydatagpca[,i]~sc)
      loads[i,]<-factorfit$coefficients[2:4]
    }

```

```
  # constructing regressors F(it)-s(it) for 1,...,F factors, i =1,...,NN

    FactorX<-matrix(NA,1+t,NN)
    rownames(FactorX)<-rownames(mydata[1:(1+t),])
    colnames(FactorX)<-colnames(mydata[,1:NN])
    Ymat<-matrix(NA,(1+t),NN)
    rownames(Ymat)<-rownames(mydata[1:(1+t),])
    colnames(Ymat)<-colnames(mydata[,1:NN])
    for (j in 1:NN){
      FactorX[,j]=-mydatagpca[,j]
      for(f in 1:FF){
        FactorX[,j]=FactorX[,j]+(loads[j,f]*sc[,f])
      }

      Ymat[,j]<-mydatagpca[,j]-Lag(mydatagpca[,j],shift = k)
    }
    FactorLX<-matrix(NaN,dim(FactorX)[1],dim(FactorX)[2])
    for(j in 1:NN){
      FactorLX[,j]<-Lag(FactorX[,j],shift = k)
      }

    FactorLXlong<-melt(FactorLX)
    Ylong<-melt(Ymat)

    Y_FactorLX <- cbind(Ylong,FactorLXlong[,3])
    colnames(Y_FactorLX) <- c("time","country","Y","gpcX")

    LRMFit <- lm(Y ~ gpcX+factor(country)-1,data = Y_FactorLX)

    c[601:617]<-LRMFit$coefficients[2:18]; c[650]<-LRMFit$coefficients[1]

    for(l in 1:NN){
      Yhat[(1+t),l+((hh-1)*NN)]=c[600+l]+c[650]*FactorX[(1+t),l]
    }
  }

  if(S==2){
    for(e in 10:NN){
      Yhat_euro[104:(R-k),hh]<-Yhat_euro[104:(R-k),hh]+Yhat[104:(R-k),e+((hh-1)*NN)]
    }
    Yhateuro[104:(R-k),hh]<-Yhat_euro[104:(R-k),hh]/8
  }
```

```

######################################## Forecast evaluation
if(S==1){
  ti1=(1+tst)
  ti2=104
}
if(S==2){
  ti1=104
  ti2=R
}
if(S==3){
  ti1=(1+tst)
  ti2=R
}
pred_error_factor<-matrix(0,dim(mydata)[1],NN)
pred_error_rw<-matrix(0,dim(mydata)[1],NN)
SPE_SPEAdj<-matrix(NA,dim(mydata)[1],NN)
for(o in 1:NN){
  mydatape[1:tst,o]<-as.matrix(mydata[1:tst,o])-mean(as.matrix(
     mydata[1:tst,o]))
  for(t in tst:tnd){
    C<-as.matrix(mydata[1:t,o])-mean(as.matrix(mydata[1:t,o]))
    mydatape[t,o]<-C[t]
  }

  if(S==2 && o>9){
    pred_error_factor[ti1:ti2,o]<-(Lag(mydatape[ti1:ti2,o],shift =
        -k)-mydatape[ti1:ti2,o])-Yhateuro[ti1:ti2,hh]
  }else{
    pred_error_factor[ti1:ti2,o]<-(Lag(mydatape[ti1:ti2,o],shift =
        -k)-mydatape[ti1:ti2,o])-Yhat[ti1:ti2,o+((hh-1)*NN)]
  }
  pred_error_rw[ti1:ti2,o]<-Lag(mydatape[ti1:ti2,o],shift = -k)-
     mydatape[ti1:ti2,o]

  pred_error_factor_all[,(((hh-1)*NN)+1):(hh*NN)]<-pred_error_
     factor
  pred_error_rw_all[,(((hh-1)*NN)+1):(hh*NN)]<-pred_error_rw

  SPE_Factor<-pred_error_factor[,o]*pred_error_factor[,o]
  SPE_rw<-pred_error_rw[,o]*pred_error_rw[,o]

  if(S==2 && o>9){
    SPE_SPEAdj[ti1:ti2,o]<-(SPE_rw[ti1:ti2]-SPE_Factor[ti1:ti2])
```

```
      +Lag(Yhateuro[ti1:ti2,hh],shift = -k)*Lag(Yhateuro[ti1:ti2,hh
         ],shift = -k)
    }else{
      SPE_SPEAdj[ti1:ti2,o]<-(SPE_rw[ti1:ti2]-SPE_Factor[ti1:ti2])
      +Lag(Yhat[ti1:ti2,o+((hh-1)*NN)],shift = -k)*Lag(Yhat[ti1:ti2,
         o+((hh-1)*NN)],shift = -k)
    }

    MSPEadj[o]<-mean(SPE_SPEAdj[,o],na.rm=TRUE)

    TheilU_CW_statistic[o,hh]<-(mean(SPE_Factor,na.rm=TRUE)/mean(SPE
       _rw,na.rm=TRUE))^0.5
  }
  #Univariate case: Standard errors and CW stats
  if(S==3){
    P1=P-k+1
    P2=P-(2*(k-1))
    t_1=1+tst
    t_2=R-k+1
  }
  if(S==1){
    P1=P-k+1
    P2=P1
    t_1=1+tst
    t_2=105
  }
  if(S==2){
    P1=P-k+1
    P2=P-(2*(k-1))
    t_1=105
    t_2=R-k+1
  }

  Yhatrec<-matrix(0,dim(mydata)[1],NN)
  dist_adj<-matrix(NA,dim(mydata)[1],NN)
  mean_dist<-c(rep(0,NN))
  sq_dist_adj<-c(rep(0,NN))
  CW_statistic<-c(rep(0,NN))
  mean_dist_cent<-matrix(NA,dim(mydata)[1],NN)

  for(jj in 1:NN){
    for(g in 1:k){
      if(S==2 && jj>9){
        Yhatrec[(t_1:t_2),jj]<-Yhatrec[(t_1:t_2),jj]+Lag(Yhateuro[t_
           1:t_2,hh],shift = g)
```

```
      }else{
        Yhatrec[(t_1:t_2),jj]<-Yhatrec[(t_1:t_2),jj]+Lag(Yhat[(t_1:t_2),jj+((hh-1)*NN)],shift = g)
      }
    }
    dist_adj[(t_1:t_2),jj]<-2*(mydatape[(t_1:t_2),jj]-Lag(mydatape[(t_1:t_2),jj],shift = 1))*Yhatrec[(t_1:t_2),jj]
    mean_dist[jj]<-mean(dist_adj[,jj],na.rm=TRUE)
    mean_dist_cent[(t_1:t_2),jj]<-dist_adj[(t_1:t_2),jj]-mean_dist[jj]
    sq_dist_adj[jj]<-(1/P2)*sum(mean_dist_cent[,jj]^2,na.rm = TRUE)

    #Univariate Clark-West stats
    CW_statistic[jj]<-sqrt(P1)*(MSPEadj[jj]/sqrt(sq_dist_adj[jj]))
    TheilU_CW_statistic[jj,hh+lh]=CW_statistic[jj]
  }
}
```