Conditional Scenario Generation with a GVAR Model

by

Xinghao Wang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Actuarial Science

Waterloo, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The stress-testing method formed an integral part of the practice of risk management. However, the underlying models for scenarios generation have not been much studied so far. In past practice, the users typically did not model risk factors for portfolios of moderate size endogenously due to the presence of "curse of dimensionality" problem. Moreover, it is almost impossible to impose the expert views for a future outcome of macroeconomy on the scenario generator without making ad-hoc adjustments.

In this thesis we propose a GVAR-based framework which allows an efficient simulation of risk factors for a complex multi-currency portfolio of various classes of assets conditioning on economic scenarios. Given reasonable sets of economic forecasts, the GVAR model anticipates the trend and codependency of the future path of portfolio risk factors and supports the production of meaningful results from risk analytics.

Acknowledgements

I want to take this opportunity to thank my supervisors, Professor David Saunders and Professor Tony Wirjanto for their kind help and support in completion of this thesis. I would like to thank Ian Buckley and Carlos Martinez Amaya for their suggestions on data.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 A Conditional Scenario Generation Problem

In stress testing, practitioners conduct risk analytics based on a portfolio P&L evaluated under a range of scenarios which are relatively adverse but probable. The quality of these scenarios are therefore key determinants of whether our risk analytics can provide a sound basis for an investment decision-making. Over the past decades scenarios have primarily been drawn from statistical models estimated based on using historical data of portfolio risk factors. Although with the help of advanced dynamic models we could generate scenarios based only on a marginal model of risk factors, we lack a formal framework to generate scenarios conditioning on available forecasts of macro-economy. Nowadays' risk managers are becoming more interested in how risk metrics of a portfolio P&L would evolve under different economic scenarios. In the past, however, perspectives regarding future evolution of macro-economic conditions are not sufficiently utilized in the generation of a forwarding-looking scenarios. Alternatively, we focus on modeling cross-sectional and hierarchical interdependencies between economic factors and portfolio risk factors. The estimated models are then used to produce forecasts of risk factors conditioning on the relevant information obtained from analyst reports.

Suppose that all vectors are column vectors unless otherwise specified. We hereby formulate a general set up of the problem as follows. For the set of integer index $\mathbb{Z}$, denote time series of the portfolio risk factors by $\mathcal{X} = \{X_t\}_{t\in\mathbb{Z}}$, where X_t are $k \times 1$ vectors, and that of the

economic factors by $\mathcal{Y} = \{Y_t\}_{t\in\mathbb{Z}}$, which is $m \times 1$ dimensional. Portfolio risk factors are those which have a direct impact on the P&L of the portfolio in question. In other words, suppose that we have a multi-asset portfolio P of which the value is V, then we have

$$V := V(X_t).$$

If P consists of equities, government bonds, corporate bonds and CDS denominated in multiple currencies, then typical portfolio risk factors include various stock indexes, points on the term structure of interest rates (e.g. government bond and swap curve), points on term structure of credit spreads (difference between the yield curves of corporate bonds and government bonds) and exchange rates.

On the other hand, economic factors influence the portfolio value indirectly through their inter-dependencies with risk factors. For example, stock indexes seem to have a positive codependency with real GDP of the country as real GDP is an aggregate measure of firms' performance over time. According to the Purchasing Power Parity theory, inflation, as an economic factor, is negatively correlated with exchange rate of the domestic currency. Typically, for the above portfolio in multiple currencies, one would want to include real GDP, inflation, the unemployment rate, the short and long term interest rates and some commodity prices in the set of economic factors.

We separate the past and future time indexes by setting current time point to T, namely, for $t \leq T$, we have full information on $\mathcal{X}$ and $\mathcal{Y}$, whereas for $t > T$ we only possess partial information on $\mathcal{Y}$ which is acquired from economic forecast reports. For convenience of illustrating conditional expectation, we denote past and future information sets with $\mathcal{P}_T$ and $\mathcal{I}_{T,h}$ respectively. In particular, let

$$\mathcal{P}_T = \{(x_t, y_t) | t \in \mathbb{Z}, 1 \leq t \leq T\}$$
$$\mathcal{I}_{T,h} = \{y_t | \text{ for some } t \in \mathbb{Z}, T < t \leq T + h\},$$

where $h \geq 1$ denotes a maximum forecast horizon. Note that in practice it is not possible to track any factor back to the origin of universe, thus we truncate $\mathcal{P}_T$ for $t < 1$ to reflect the historical data available to us. Moreover, we restrict the focus of this thesis on generating scenarios conditioning on a fixed future path of economic factors, which is the reason why

we specify elements in $\mathcal{I}_{T,h}$ with lower case letters. Also note that we do not require the future information regarding Y_t for all t up to time $T+h$ if we treat the economic variables endogenously (see Section 2.5 for details).

Let A' denote a transpose of matrix/vector A. With the notation defined so far, we formulate a conditional scenario generation problem as sampling from a joint distribution of the random vector $(X'_{T+1}, \ldots, X'_{T+h})'$, conditioning on $\mathcal{P}_T$ and $\mathcal{I}_{T,h}$. Specifically, we need to find a statistical model for

$$F(X'_{T+1}, \ldots, X'_{T+h} | \mathcal{P}_T, \mathcal{I}_{T,h}) \tag{1.1}$$

with $F(\cdot|\cdot)$ denoting the conditional joint distribution function.

For the purpose of risk management, eventually we would like to measure the risk of our portfolio with risk metrics. Given a certain future economic scenario $\mathcal{I}_{T,h}$, we can obtain the portfolio P&L over time horizon s, $1 \leq s \leq h$, for a realization $(x'_{T+1}, \ldots, x'_{T+h})'$, which is drawn out of (1.1), as

$$\Delta_s V = V(x_{T+s} | \mathcal{P}_T, \mathcal{I}_{T,h}) - V(x_T | \mathcal{P}_T),$$

which has the following conditional distribution

$$F_{\Delta_s V | X,Y}(v) = \mathbb{P}\{\Delta_s V \leq v | \mathcal{P}_T, \mathcal{I}_{T,h}\}. \tag{1.2}$$

If $V(\cdot|\cdot)$ is a linear function or has a simple form, we can determine the distribution in (1.2) analytically, with which the risk metrics can be calculated. Otherwise, the empirical distribution can be obtained by means of a Monte Carlo simulation.

1.2 A Joint Factor Model

A key to the conditional scenario generation problem is to determine an appropriate underlying statistical model for (1.1). In spite of the negligible serial correlation for high-frequency financial time series, when it comes to modeling the evolution of risk factors over relatively long time horizons, the autocorrelation becomes persistent enough for us to consider an autoregressive (AR) time series model. Further, to model the cross-sectional dependencies

within and between $\mathcal{X}$ and $\mathcal{Y}$, we apply a vector autoregressive (VAR) model as our joint factor evolution model.

Let $\mathcal{Z} = \{Z_t\}_{t\in\mathbb{Z}}$ be time series of vectors that collect $\mathcal{X}$ and $\mathcal{Y}$ such that $Z_t = (X_t', Y_t')'$. The AR model for $\mathcal{Z}$ can be written as

$$Z_t = \sum_{l=1}^{p} \mathbf{F}_l Z_{t-l} + \mathbf{C}\varepsilon_t, \tag{1.3}$$

where $\mathbf{F}_l$, for $l = 1, \ldots, p$, is the lag l coefficient matrix, $\varepsilon_t \sim N(0, \mathbf{I}_{k+m})$ is a Gaussian white noise process with $\mathbf{I}_\bullet$ denoting the identity matrix and $\mathbf{C}$ specifies the covariance matrix of innovations for each element of Z_t.

The VAR model in (1.3) explicitly allows for potential auto-correlation and co-integrating (long term) relations between time series. By modeling concatenated X_t and Y_t we assume all economic factors are endogenous to our portfolio risk factors. From the perspective of computational efficiency, this specification significantly increases the cost of estimation since the number of parameters in VAR grows quadratically with the dimension of the time series. Alternatively, we can reduce the scale of our coefficient matrix by modeling economic factors as common variables, i.e.

$$X_t = \sum_{l=1}^{p} \mathbf{F}_l X_{t-l} + \sum_{l=0}^{s} \mathbf{H}_l Y_{t-l} + \mathbf{C}\varepsilon_t, \tag{1.4}$$

where $\mathbf{H}_l$, $l = 0, \ldots, s$, is the lag l factor loading matrix. In this case, we assume strong exogeneity (detailed in Section 2.2.3) of the economic factors. Moreover, if we possess complete future path information of $\mathcal{Y}$ over the forecast horizon, we do not need to construct a marginal model for $\mathcal{Y}$ for the purpose of conditional scenario generation.

Even though the number of parameters has been reduced for VAR as specified in (1.4), we still suffer from the "curse of dimensionality" (i.e. a proliferation of parameters as the system dimension grows) since there are usually dozens of risk factors that need to be modeled for the analytics of a multi-asset portfolio. To make this model computationally tractable, we will apply a modified version of VAR, known as the global vector autoregressive approach, which is suitable for modeling high-dimensional systems.

1.3 Literature Review

1.3.1 Relations Between Economic and Portfolio Risk Factors

For the idea of conditional scenario generation to work, there must at least exist some relations between the economic and portfolio risk factors, so that the contemporaneous or lagged values of the economic variables can have predictive power to the portfolio risk variables.

Fortunately, there is a rich literature on this topic. For instance, Tripathi and Seth (2014) examine the causal and co-integration relationships between macroeconomic factors and performance of the Indian stock market, which is represented by BSE Sensex index, employing the Granger causality test, Johansen's co-integration test and other statistical techniques. They select a representative set of macroeconomic variables including inflation, a interest rate, money supply, oil price, exchange rates and an Indian industrial production index and conduct the aforementioned tests on monthly time series spanning the period from July 1997 to June 2011. In terms of causal relations in the short run, their results of the Granger causality test indicate that the exchange rate, money supply and oil price Granger cause Sensex. However, there is bi-directional causality between the oil price and Sensex. Literature reviewed in their paper also provide supporting results. For example, Mukherjee and Naka (1995) found that a co-integration relation exists among Japanese stock market performance and six macroeconomic variables and a positive relation was found between the Japanese industrial production and stock return; Abdalla and Murinde (1997) investigate interactions between exchange rates and stock prices and found that the causality is uni-directional from exchange rates to stock prices.

Hanousek and Filer (2000) also did a research on causal relationship between economic factors and stock returns. However, they provide a new perspective on this type of research by interpreting the results of the causality tests in terms of semi-strong market efficiency [1]. If the market is semi-strongly efficient, the stock prices at any time should reflect all publicly available information. Thus, the contemporaneous value of economic factors should have predictive power in stock returns. The lagged values, however, should not Granger cause the stock returns since the historical information has already been accounted for. In

[1] Although they are not the first one in this class.

mathematical terms, they estimate the equation (1.4) and test the Granger and instantaneous causality. If Granger causality fails to exist while instantaneous causality is found to be statistically significant, the semi-strong efficient market hypothesis holds. By conducting analysis on monthly data from 1993 to 1999 of four central-European countries, namely Hungary, Poland, Czech Republic and Slovakia, they conclude that none of the four countries had developed an efficient market. However, the performance of the more sophisticated two stock markets, namely the Budapest Stock Exchange and the Warsaw Stock Exchange, are more closely related to the macro economy.

A comprehensive survey conducted by Kirkulak and Ezzat (2014) examines the relationship between market efficiency and level of economic development. In their paper, the weak-form of the efficient market hypothesis is investigated by an array of statistical tests including a serial correlation test, runs test and variance ratio test using a GARCH-M$(1, 1)$ model. Daily time series, with the sampling period from 2005 through 2013, of market returns from 17 developed and 12 emerging countries are considered. The emerging countries are further divided into classes of 'Developed', 'Advanced Emerging', 'Secondary Emerging', 'Watch List' and 'Stand Alone' according to the classification of the FTSE Group. They reached the conclusion that market efficiency is associated with a high economic development level. The developed countries generally exhibit greater evidence of market efficiency. Among emerging markets, while the so called Advanced Emerging and Secondary Emerging Markets are found to be weak-form efficient, the other two less developed groups are more predictable with historical data. Although the weak-form efficiency is not of primary concern to our study, such a relationship between market sophistication and efficiency indirectly justifies the assumption of semi-strong form efficiency for more developed markets such as US, GB, JP and EU stock exchanges.

The behavior of spreads on credit default swaps is more complicated, since a CDS is essentially a derivative. There is an extensive literature regarding determinants of CDS spreads. In the papers reviewed by Kim, Park, and Park (2013), researchers found that leverage ratio, implied volatility of the stock option of the firm, realized volatility from high frequency equity prices and the risk-free rate can explain most of the variations in CDS premia. Whereas the previous work mostly concentrates on studying firm-specific variables, Kim, Park, and

Park (2013) did research on a number of macroeconomic variables (e.g. expected market risk premium, S&P500 index, implied volatility of S&P500 index options, Goldman Sachs financial conditions index, industrial price index etc.) in addition to the firm-specific variables studied previously. The sampling period is from 2004 to 2012. The most interesting finding in this paper is that, while the expected market risk premium, financial conditions index, and industrial price index are significant in explaining CDS variation during pre-crisis and post-crisis periods, the significance of these variables are weak during the crisis period. Based on this finding, they concluded that the factor of macroeconomic conditions play a critical role in pricing CDS when the underlying asset value of the CDS is likely to be farther from the default barrier.

1.3.2 Scenario Generation for Stress Testing

Whether the risk management analytics are able to assist informed decision-making depends on the quality of the scenarios studied in the stress-testing task. The "good" scenarios are expected to be relevant and forward-looking and should be able to reflect the various views the management has on the future outcome of the economy. Formally, Meucci (2008) establishes a general theoretical framework of stress-testing. Assume that the portfolio is driven by an N-dimensional vector of risk factors, X. Then at any time t, the portfolio value is a deterministic function of the realization of X, which is similar to our setting in (1.1). In the setting of Meucci (2008), X can contain variables that are not directly fed into the pricing function, e.g. those influence P&L statistically through correlation, and thus is analogous to Z_t of (1.3). The existence of a model for X is assumed and the model is represented by a probability density function

$$X \sim f_X. \tag{1.5}$$

The model in (1.5) without any constraints is analogous to the marginal model in (1.3), which can be seen as a special case of (1.5). Further, we impose "views" on the marginal model (1.5) to obtain the conditional model of X under stressed scenarios. In the most general case, the "views" are defined as constraints on any aspect of generic functions of X. In particular,

for generic functions g_i, $i = 1, \ldots, K$,

$$V := (g_1(X), \ldots, g_K(X))' \sim f_V \tag{1.6}$$

forms a new vector of random variables, upon which the constraints are imposed. f_V can be inferred completely from the g_i's and f_X, although the analytical form is not necessarily available. In the most detailed specification of views, we obtain the full subjective distribution for V,

$$V \sim \tilde{f}_V \neq f_V. \tag{1.7}$$

If the $\tilde{f}_X$ is uniquely solvable from $\tilde{f}_V$, the task of constructing the conditional model is accomplished and what follows is to draw simulations from $\tilde{f}_X$ and conduct portfolio analytics. In a more common practical setting, however, we only impose constraints on a certain aspect of the distribution of V, such as moments, percentiles and dependence structure. See Meucci (2008) for a partial list of possible constraints. When the constraints are looser than a fully specified distribution $\tilde{f}_V$, we are left with a non-singleton set of conditional distributions, among which we are to select an optimal candidate as $\tilde{f}_X$. Meucci (2008) proposed the following optimality criterion

$$\tilde{f}_X \equiv \arg\min_{f \in \mathbb{V}} \{\mathcal{E}(f, f_X)\}, \tag{1.8}$$

where $f \in \mathbb{V}$ stands for all candidate distributions that are consistent with the views imposed and the relative entropy, $\mathcal{E}(\tilde{f}_X, f_X)$, between a generic distribution $\tilde{f}_X$ and a marginal distribution f_X, is defined as

$$\mathcal{E}(\tilde{f}_X, f_X) := \int \tilde{f}_X(x) \left[\ln \tilde{f}_X(x) - \ln f_X(x)\right] \mathrm{d}x. \tag{1.9}$$

As the relative entropy $\mathcal{E}(\tilde{f}_X, f_X)$ quantifies the difference between $\tilde{f}_X$ and f_X, the criterion (1.8) argues that one should select among all candidates the distribution that is "closest" to the marginal model.

Among all possible models for X conditioning on "views", Meucci (2013) proposed a semi-parametric approach where probability is assigned to each observation of the time series of X based on a mixture of exponential smoothing and kernels. In particular, let $\mathcal{X} =$

$\{x_1, \ldots, x_t\}_{t=1}^{\bar{t}}$ denote the past time series of X defined in (1.5). We omit the subscript on X to emphasize the assumption that X are approximately independent and identically distributed[2]. Thus X at any future time is represented as

$$X \sim \{x_t, p_t\}_{t=1}^{\bar{t}}. \tag{1.10}$$

The probabilities p_t, which are the weights assigned to each historical scenario, x_t, are then determined semi-parametrically. Intuitively, recent observations should be weighted more heavily as they carry a higher predictive power. This view is reflected through the so-called exponential smoothing. Suppose that we are at time $\bar{t}$. Conditioning on the observations up until $\bar{t}$, the probabilities, denoted by $p_t|\bar{t}$ for $t = 1, \ldots, \bar{t}$, are determined as follows

$$p_t|\bar{t} \equiv p_t^{exp} \propto e^{-\frac{\ln 2}{\tau}|t-\bar{t}|}, \tag{1.11}$$

where $\tau > 0$ is the half-time[3] of the exponential decay and $\propto$ means "proportional to" with the proportionality constant set in the way that the probabilities sum up to one.

On the other hand, it is desirable that simulations of X can be drawn from model (1.10) with a target on certain elements of the vector X set. That is, the scenarios are generated conditioning on certain levels of market indicators such as the volatility index (VIX) of the Chicago Board Options Exchange (CBOE), real GDP, a five year swap rate and etc. Denote one such market indicator by Z_t, which is univariate, and its conditioned value by z^*. A simple way to implement this is through the so-called "crisp" probabilities

$$p_t|z^* \equiv p^{crisp} \propto \begin{cases} 1 & \text{if } z_t \in \mathcal{R}(z^*) \\ 0 & \text{otherwise} \end{cases}, \tag{1.12}$$

where $\mathcal{R}(z^*)$ denotes a range around the target z^*. The range is set with respect to a predetermined probability α, symmetrically around z^*. Formally, $\int_{\underline{z}}^{z^*} \hat{f}_Z(z)\mathrm{d}z = \frac{\alpha}{2} = \int_{z^*}^{\overline{z}} \hat{f}_Z(z)\mathrm{d}z$, where $\underline{z}$ and $\overline{z}$ denote an upper and lower bound of the range respectively and $\hat{f}_Z$ denotes the empirical distribution of $Z_t \sim \{z_t, p_t\}_{t=1}^{\bar{t}}$ with p_t set either non-parametrically or through an exponential smoothing, crisp or kernel-based probabilities described below. If z^* is in the upper (lower) $\alpha/2$ quantile, we can override it with $\max z_t$ ($\min z_t$).

[2]The author loosens the exact iid assumption for his semi-parametric model.

[3]If $\tau = \frac{\bar{t}}{2}$, then $p_\tau = \frac{1}{2}p_{\bar{t}}$.

A smoother implementation than crisp probability is kernel-based probability, which is given as

$$p_t|z^* \equiv p_t^{ker} \propto e^{-|z_t - z^*|^\gamma / h}, \tag{1.13}$$

where h is the kernel band-width and γ governs the kernel tails. Note that $\gamma = 2$ leads to Gaussian kernel and $\gamma = 1$ is an exponential kernel.

We apply the entropy pooling approach in Meucci (2008) to combine exponential smoothing probabilities with market-conditioned probabilities. We start by computing the moments of a market indicator conditioning on target z^*

$$\mu|z^* = \sum_{t=1}^{\bar{t}} z_t p_t^{crisp}, \qquad \sigma|z^* = (\sum_{t=1}^{\bar{t}} z_t^2 p_t^{crisp} - (\mu|z^*)^2)^{1/2}. \tag{1.14}$$

Next, all probability sets p_t which match the above moments are collected into a collection $\mathcal{V}|z^*$, namely

$$\mathcal{V}|z^* : \begin{cases} \sum_{t=1}^{\bar{t}} p_t z_t = \mu|z^* \\ \sum_{t=1}^{\bar{t}} p_t z_t^2 \leq (\mu|z^*)^2 + (\sigma|z^*)^2 \end{cases} . \tag{1.15}$$

Taking the p_t^{exp} as the marginal model as in formula (1.9), we obtain the optimal probabilities as

$$\mathbf{p}|z^* = \underset{\mathbf{p} \in \mathcal{V}|z^*}{\arg\min}\, \mathcal{E}(\mathbf{p}, \mathbf{p}^{exp}) \tag{1.16}$$

with $\mathcal{E}$ defined in formula (1.9).

Note that entropy pooling techniques allows only for a univariate conditioner. In order to condition on multiple variables $Z_{1,t}, \ldots, Z_{\bar{q},t}$ with their respective targets $z_1^*, \ldots, z_{\bar{q}}^*$, we first compute mixed probabilities for each conditioner individually using entropy pooling techniques as in (1.16)

$$\mathbf{p}_q \equiv \mathbf{p}|z_p^*, \qquad q = 1, \ldots, \bar{q}. \tag{1.17}$$

Each probability set $\mathbf{p}_q$ is then weighted with the following weighting scheme, denoted by $\mathbf{w} = \{w_1, \ldots, w_{\bar{q}}\}$. For $\mathbf{p}_q$, compute the effective number of scenarios as

$$\mathcal{T}_q = e^{-\sum_{t=1}^{\bar{t}} p_{q,t} \ln p_{q,t}}. \tag{1.18}$$

Then for any pair of probability sets $(\mathbf{p}_q, \mathbf{p}_r)$, define $b_{q,r} := \sum_{t=1}^{\bar{t}} (p_{q,t} p_{r,t})^{1/2}$. Then we compute the Hellinger distance $d_{q,r} = \sqrt{1 - b_{q,r}}$ between these two sets of probabilities. The measure of diversity of $\mathbf{p}_q$ from all other set of probabilities can then be computed as

$$\mathcal{D}_q = \frac{1}{\bar{q} - 1} \sum_{r \neq q} d_{q,r}. \tag{1.19}$$

The argument made by Meucci (2013) is that the probability set which is more distant from other sets and has a higher effective number of scenarios should be weighted more heavily, namely

$$w_q = \frac{\mathcal{T}_q \mathcal{D}_q}{\sum_{r=1}^{\bar{q}} \mathcal{T}_r \mathcal{D}_r}. \tag{1.20}$$

Two aggregation methods are proposed to obtain the final ensemble probability set $\mathbf{p}_{ce}$. On the one hand, the simple weighted average, which is given as

$$\mathbf{p}_{ce} = \sum_{r=1}^{\bar{q}} w_r \mathbf{p}_r \tag{1.21}$$

can be applied. Or, the log-linear weighted scheme defined as

$$\ln \mathbf{p}_{ce} \overset{1}{\leftarrow} \sum_{r=1}^{\bar{q}} w_r \ln \mathbf{p}_r \tag{1.22}$$

can be used, where $\overset{1}{\leftarrow}$ denotes that rescaling is needed to make sure that elements of $\mathbf{p}_{ce}$ sum up to one.

A more prevalent class of model for scenario generation is a fully parametric time series model. Examples include an ARMA-GARCH model, a vector autoregressive (VAR) model and a regime switching model. The global vector autoregressive (GVAR) model, as a restricted case of a VAR model applicable to a high dimensional data setting, is studied extensively in later chapters.

Another interesting paper is Rosen and Saunders (2015), where the authors proposed an easy-to-implement framework to obtain a conditional expectation, or more generally the full conditional distribution, of all portfolio risk factors, conditioning on the outcome of the economic factors from a pre-computed simulation using a least squares regression, referred to as Least Square Stress Testing (LSST). The process can be summarized as follows. First

we need to construct a model for the joint distribution of portfolio risk factors and economic variables, $(X', Y')'$. A large number of simulations are produced from the joint distribution, denoted by $\mathcal{S} = \{(x_i', y_i')'\}_{i=1}^{N}$. We then run a generalized linear model regressing Y on X, with the simulations

$$Y = BX + U, \tag{1.23}$$

where B is the matrix of regression coefficients and U is the vector error with zero mean, and is assumed to be independent of X. The linear model is generalized in the sense that we can replace Y with certain non-linear functions of the economic variables, or we can append X with non-linear functions of its elements. Such setting makes the regression-based approach less restrictive. Denote the estimated coefficient matrix by $\hat{B}$. Then the conditional expectation of portfolio risk factors, given a scenario of economic outcome $X = x_0$, is

$$\hat{y}_0 = \mathbb{E}[Y|X = x_0] = \hat{B}x_0. \tag{1.24}$$

A further construction of the full conditional distribution of $F_{Y|X}(y|x_0)$ depends on the treatment of the regression residuals, $\hat{u}_i$

$$\hat{u}_i = y_i - \hat{B}x_i. \tag{1.25}$$

While the assumption of normal or other parametric distributions on $\hat{u}_i$ makes further calculation analytically tractable, LSST framework does not preclude a non-parametric approach. For example, we may assume that

$$P[U = \hat{u}_i] = \frac{1}{N}, \qquad \text{for } i = 1, \ldots, N \tag{1.26}$$

and consequently specify the conditional distribution of portfolio risk factors as

$$P[Y = \hat{u}_i + \hat{y}_0] = \frac{1}{N}, \qquad \text{for } i = 1, \ldots, N. \tag{1.27}$$

1.3.3 Solutions to Curse of Dimensionality Problem

A recurrent problem in modeling high-dimensional time series is that the number of variables is large relative to the available time dimensions. This problem is usually referred to as a

"curse of dimensionality". Other than the GVAR approach applied in our study, possible approaches to modeling of high-dimensional time series in recent literature are broadly divided into two categories: (i) data shrinkage and (ii) parameter space shrinkage.

Applied in Bernanke, Boivin, and Eliasz (2004) to model a monetary policy shock, the so-called factor-augmented VAR model is a classic approach of data shrinkage. The FAVAR model is devised to solve the "price puzzle", namely the finding in the literature of the VAR model for monetary policy that a contractionary monetary policy shock is followed by an increase in the price level, rather than a decrease as standard economic theory would predict. As pointed out in Sims (1992), this may be caused by absence of data that would have been contributed to an explanation of the policy shock. However, the entire informational time series usually consist of hundreds of variables, making the inclusion of the whole information set impractical. The FAVAR model thus assumes that the economy is driven by an $M \times 1$ vector of observable economic variables and a $K \times 1$ vector of unobservable factors, which jointly follow a transition equation:

$$\begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \mathbf{\Phi}(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + v_t, \tag{1.28}$$

where $\mathbf{\Phi}(L)$ is the lag polynomial of finite p order and the error term v_t is zero mean with covariance matrix $\mathbf{Q}$. Note that this equation does not prevent the imposition of restrictions in the lag polynomial $\mathbf{\Phi}(L)$. Also, if the terms in $\mathbf{\Phi}(L)$ that relate Y_t to F_t are all zeroes, the equation (1.28) reduces to a standard VAR model, otherwise the VAR model for Y_t is augmented by unobservable factors F_t.

Further, let an $N \times 1$ vector, X_t, represents all of the observable variables that reflect activities in the economy. In practice, the dimension N is much larger than the time dimension T of the sampling period. It is also assumed that $N \gg K + M$ is required for estimation. As a result of the assumption that $(F_t', Y_t')'$ collectively determines the economy, all of the variables in X_t are functions of $(F_t', Y_t')'$. Assuming a linear relationship under the FAVAR framework, this leads to an observation equation of the form,

$$X_t = \Lambda^f F_t + \Lambda^y Y_t + e_t, \tag{1.29}$$

where Λ^f and Λ^y are $N \times K$ and $N \times M$ matrices of factor loadings. The error term, e_t, has

a zero mean. Depending on the method of estimation, the error terms will be assumed to be normal and uncorrelated or display only a small amount of correlation. The specification of equation (1.29) that X_t depends only on contemporaneous values of $(F_t', Y_t')'$ is by no means restrictive since F_t can contain arbitrary lags of fundamental factors.

Due to the unobservable factors, the equations (1.28) and (1.29) can not be estimated directly. Bernanke, Boivin, and Eliasz (2004) thus proposed two methods of estimation, namely, a two-step semi-parametric approach and a one-step maximum likelihood approach. The two-step semi-parametric approach is based on a principle component analysis. As specified in equation (1.28), each constituent of X_t is a linear combination of $(F_t', Y_t')'$. Thus when N is sufficiently large (as assumed previously), the first $K + M$ principle components, which are the eigenvectors corresponding to the $K + M$ largest eigenvalues of XX', where $X = [X_1, \ldots, X_T]$, recover the space spanned by $(F_t', Y_t')'$ consistently[4]. The principle components found above, denoted by $\hat{C}(F_t, Y_t)$, are arbitrary linear combinations of $(F_t', Y_t')'$. Once we obtain the estimates of terms in $\hat{C}(F_t, Y_t)$ which involve only Y_t, if they are not known, we can subtract them from $\hat{C}(F_t, Y_t)$ to obtain linear combinations of F_t. $\hat{F}_t$ are solved based on some structural assumptions and an identification schemes. Equation (1.28) is then estimated as a standard VAR model upon substitution in the $\hat{F}_t$ and observations of Y_t. However, the estimation of the linear combination of Y_t in $\hat{C}(F_t, Y_t)$ relies further on assumptions that allows the estimation of the linear combinations of only F_t. The details of this can be found in Bernanke, Boivin, and Eliasz (2004). A more computationally intensive method of estimation, where equations (1.28) and (1.29) are jointly estimated by maximum likelihood with an independent normal assumption for the error terms e_t, is also considered. In the context of high-dimensional time series, the joint estimation is enabled by applying the so-called likelihood-based Gibbs sampling techniques (see Kim and Nelson (1999) for a survey of this technique).

Unlike the FAVAR approach, which reduces the dimensionality of the data, a Bayesian VAR shrinks the parameter space by imposing a prior belief on the coefficients of the VAR model. Based on the results discussed in Litterman (1986), Kadiyala and Karlsson (1997)

[4]For $N \to \infty$ and $T \to \infty$, the bases of the first $K + M$ principle components are equal to the bases of space spanned by $(F_t', Y_t')'$ with probability 1 (see Stock and Watson (2002) for details).

and Sims and Zha, 1998, Bańbura, Giannone, and Reichlin (2010) developed a procedure of applying a Bayesian approach to a potentially large n-dimensional VAR(p) model for $Y_t = (Y_{1,t}, Y_{2,t}, \ldots, Y_{n,t})'$, defined as

$$Y_t = c + \sum_{l=1}^{p} \mathbf{A}_l Y_{t-l} + u_t, \tag{1.30}$$

where u_t is the n-dimensional Gaussian white noise process with a covariance matrix $\mathbf{\Psi}$, c is a n-dimensional vector of constants and $\mathbf{A}_l$'s are VAR coefficient matrices. The prior belief that all equations are centered around a random walk with drift,

$$Y_t = c + Y_{t-1} + u_t \tag{1.31}$$

is first imposed on the VAR model in (1.30). The implications for the prior are that more recent lags and the variables' own lags should provide more information than other lags. Thus the coefficients alone the diagonal of $\mathbf{A}_1$ are shrunk to one while the others are shrunk towards zero. Mathematically, such prior beliefs are equivalent to setting the mean and variance of the prior distribution for coefficients as follows:

$$\mathbb{E}[(\mathbf{A}_l)_{ij}] = \begin{cases} \delta_i, & j = i, l = 1 \\ 0, & \text{otherwise} \end{cases}, \qquad \mathbb{V}[(\mathbf{A}_l)_{ij}] = \begin{cases} \frac{\lambda^2}{l^2}, & j = i \\ \vartheta \frac{\lambda^2}{l^2} \frac{\sigma_i^2}{\sigma_j^2}, & \text{otherwise} \end{cases}. \tag{1.32}$$

The parameters, δ_i's, are determined according to the prior beliefs on the time series being modeled. For the time series whose recent shocks are believed to be highly persistent, we set $\delta_i = 1$, while for the time series which display mean reversion, we set $\delta_i = 0$, which implies a white noise assumption for that variable. The scale parameters σ_i are set to the estimated variance of the residuals of the univariate AR(p) model for the variable $Y_{i,t}$. As stated in Bańbura, Giannone, and Reichlin (2010), this contradicts the principles of the Bayesian approach but is acceptable in practice. λ controls the overall tightness of the prior distribution around the random walk or white noise and its value should decline as the system of Y_t gets larger to avoid over-fitting the system. $\vartheta \in (0, 1)$ governs the extent to which lags of other variables are less important than their own lags. However, in order to obtain a non-diagonal covariance $\mathbf{\Psi}$ for the residual u_t, to make it possible to model the correlations among residuals of different equations, ϑ is required to be set to one. Then we collect all coefficients

of the VAR(p) model of (1.30) in one matrix $\mathbf{B}$ by re-writing (1.30) in its companion form:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{U}, \tag{1.33}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_T)'$, $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_T)'$ with $\mathbf{X}_t = (Y'_{t-1}, \ldots, Y'_{t-p}, 1)'$, $\mathbf{U} = (u_1, \ldots, u_T)'$ and $\mathbf{B} = (\mathbf{A}_1, \ldots, \mathbf{A}_p, c)'$. We posit the normal inverse Wishart prior on the joint system of $\mathbf{B}$ and $\mathbf{\Psi}$, which has the form

$$\text{vec}(\mathbf{B}) \sim N(\text{vec}(\mathbf{B}_0), \mathbf{\Psi} \otimes \mathbf{\Omega}_0) \qquad \text{and} \qquad \mathbf{\Psi} \sim iW(\mathbf{S}_0, \alpha_0), \tag{1.34}$$

where the prior parameters are chosen such that the prior expectation and variance of $\mathbf{B}$ match those implied by equations (1.32). To this end, Bańbura, Giannone, and Reichlin (2010) append $\mathbf{Y}$ with the following T_d dummy observations

$$\mathbf{Y}_d = \begin{bmatrix} \text{diag}(\delta_1\sigma_1, \ldots, \delta_n\sigma_n)/\lambda \\ \mathbf{0}_{n(p-1)\times n} \\ \cdots \\ \text{diag}(\sigma_1, \ldots, \sigma_n) \\ \cdots \\ \mathbf{0}_{1\times n} \end{bmatrix} \qquad \mathbf{X}_d = \begin{bmatrix} \mathbf{J}_p \otimes \text{diag}(\sigma_1, \ldots, \sigma_n)/\lambda & \mathbf{0}_{np\times 1} \\ \mathbf{0}_{n\times np} & \mathbf{0}_{n\times 1} \\ \mathbf{0}_{1\times np} & \epsilon \end{bmatrix}, \tag{1.35}$$

where $\mathbf{J}_p = \text{diag}(1, 2, \ldots, p)$ and ϵ is a small number that reflects the uninformative prior for the intercept. Augmenting the regression in (1.33) with $\mathbf{Y}_* = (\mathbf{Y}', \mathbf{Y}'_d)'$, $\mathbf{X}_* = (\mathbf{X}', \mathbf{X}'_d)'$ and $\mathbf{U}_* = (\mathbf{U}', \mathbf{U}'_d)'$, we have

$$\mathbf{Y}_* = \mathbf{X}_*\mathbf{B} + \mathbf{U}_*. \tag{1.36}$$

Then the posterior has the form

$$\text{vec}(\mathbf{B})|\mathbf{\Psi}, \mathbf{Y} \sim N(\text{vec}(\tilde{\mathbf{B}}), \mathbf{\Psi} \otimes (\mathbf{X}'_*\mathbf{X}_*)^{-1}) \qquad \text{and} \qquad \mathbf{\Psi}|\mathbf{Y} \sim iW(\tilde{\Sigma}, T_d + 2 + T - k) \tag{1.37}$$

with $\tilde{\mathbf{B}} = (\mathbf{X}'_*\mathbf{X}_*)^{-1}\mathbf{X}'_*\mathbf{Y}_*$ and $\tilde{\Sigma} = (\mathbf{Y}_* - \mathbf{X}_*\tilde{\mathbf{B}})'(\mathbf{Y}_* - \mathbf{X}_*\tilde{\mathbf{B}})$. The posterior expectation of the coefficients coincides with the OLS estimates of the regression of $\mathbf{Y}_*$ on $\mathbf{X}_*$ and also with the posterior mean of the Minnesota setup in Litterman (1986). The computational feasibility of the estimation is ensured as it only requires the inversion of an $(np+1) \times (np+1)$ square matrix.

Chapter 2

Global Vector Autoregressive Model

Originally proposed by Pesaran, Schuermann, and Weiner (2004), the global vector autoregressive (GVAR) approach provides a simple yet effective way of modeling interdependencies in a complex high-dimensional system. The methodological contribution of GVAR lays in dealing with the curse of dimensionality in a theoretically coherent and statistically consistent manner. Also, among all global models, GVAR is currently the only one that presents a complete and closed system, which is required for simulation (see Chudik and Pesaran, 2014).

The two merits of the GVAR approach mentioned above motivate the application of this model to the conditional scenario generation problem. Firstly, we need an easy-to-estimate model to facilitate the modeling of interactions between portfolio risk factors and economic variables, both of which are of high dimension. Secondly, for the purpose of risk analytics, we require the joint factor model to be simple-to-simulate, as in most of the cases, the P&L of the portfolio are not linear with respect to the underlying risk factors, which makes obtaining close-form formulas for risk metrics particularly difficult, and necessitates the use of Monte Carlo methods.

The GVAR method tackles the curse of dimensionality problem by using a simple data shrinkage approach. For a high-dimensional system, we first divide its elements into segments of a smaller scale. Each segment is then modeled with an augmented vector autoregressive (VAR), denoted as VARX*, which features both domestic variables of the segment and weighted cross-section averages of foreign variables (aka. star variables) in the system. The

VARX* model not only shrinks the size of the coefficient matrix but also allows for the cross-sectional interaction between segments through the star variables. Once the VARX* models have been estimated, we can stack them and solve the simultaneous equations as one large global VAR model as in (1.3) and (1.4). Together with the normality assumption for the innovations, we can then use the solved GVAR model to construct the joint factor model defined in (1.1).

2.1 Augmented Vector Autoregressive Model

Suppose that we divide the global vector into N cross-sectional units, X_{it} for $i = 1, \ldots, N$, such that $X_t = (X_{1t}', \ldots, X_{Nt}')'$ with some appropriate rearrangement of elements in X_t. Assume that X_{it} has a dimension $k_i \times 1$, then clearly $k = \sum_{i=1}^{N} k_i$. The first step of constructing a GVAR model for the global vector X_t is to establish and estimate a VARX* model for X_{it} of each of the cross-sectional units. Note that the feature of VARX* model is the presence of star variables which are collected in the $k^* \times 1$ vector X_{it}^* defined as

$$X_{it}^* = \tilde{\mathbf{W}}_i' X_t, \tag{2.1}$$

for $i = 1, \ldots, N$, where $\tilde{\mathbf{W}}_i$ is the $k \times k^*$ weight matrix for the ith unit. In general, the dimension of X_{it}^* does not need to be the same for all units. In recent applications of the GVAR approach to modeling of international macroeconomics, weight matrices are constructed using foreign trade or capital flows (see e.g. Bussière, Chudik, and Sestieri, 2009). To ensure the sufficiency of the parameter reduction, both k_i and k^* should be sufficiently small (typically between 4 and 6).

The star variables and their lagged values are then included in the VAR model representing each unit given as

$$X_{it} = \sum_{l=1}^{p_i} \mathbf{\Phi}_{il} X_{i,t-l} + \mathbf{\Lambda}_{i0} X_{it}^* + \sum_{l=1}^{q_i} \mathbf{\Lambda}_{il} X_{i,t-l}^* + \mathbf{C}_i \varepsilon_{it}, \tag{2.2}$$

for $i = 1, \ldots, N$, where $\mathbf{\Phi}_{il}$ for $l = 1, \ldots, p_i$, $\mathbf{\Lambda}_{il}$ for $l = 0, \ldots, q_i$, are unknown $k_i \times k_i$ and $k_i \times k^*$ coefficient matrices, $\mathbf{C}_i$ specifies the covariance matrix of the innovations of the ith

cross-sectional unit. Under certain granularity conditions of the weight matrices, namely

$$||\mathbf{W}_i|| = O(N^{-\frac{1}{2}}), \tag{2.3}$$

and

$$\frac{||\mathbf{W}_{ij}||}{||\mathbf{W}_i||} = O(N^{-\frac{1}{2}}), \tag{2.4}$$

,where the norm is the so-called spectral norm[1], as $N, T \to \infty$ and $T/N \to 0$[2], the star variables can be seen as an approximation to global common factors and thus be treated as exogenous for the purpose of estimating the unknown coefficients. This means ideally that if we have a large number of cross-sectional units and the granularity conditions specified in (2.3) and (2.4) are satisfied, the actual value of the weights are of secondary importance (see Dees et al., 2007; Chudik and Pesaran, 2011).

If one decides to model economic factors endogenously as the representation in (1.3), then Y_t can be modeled in the same manner as X_t. In this case we allow for the feedback effect from portfolio risk factors to economic factors through the star variables. Otherwise, Y_t can be included in (2.2) as the common factors for all sub-sectional VARX* models as

$$X_{it} = \sum_{l=1}^{p_i} \mathbf{\Phi}_{il} X_{i,t-l} + \mathbf{\Lambda}_{i0} X_{it}^* + \sum_{l=1}^{q_i} \mathbf{\Lambda}_{il} X_{i,t-l}^* + \mathbf{D}_{i0} Y_t + \sum_{l=1}^{s_i} \mathbf{D}_{il} Y_{t-l} + \mathbf{C}_i \varepsilon_{it}, \tag{2.5}$$

where $\mathbf{D}_{il}$ for $l = 0, \ldots, q_i$, are $k_i \times m$ unknown factor loading matrices.

2.2 Causality and Exogeneity

Both the assumed forms of the VARX* model in (2.2) and (2.5) imply that the star variables are treated as exogenous. This helps us further in achieving a more parsimonious model, even though it is not a required treatment under the GVAR framework. On the other hand, the common factors, Y_t, are also assumed to be exogenous as in (2.5). In our application, economic variables are likely to be taken as the common factors in the VARX* model. Intuitively, we

[1]The spectral norm of matrix $\mathbf{A}$ is the largest singular value of $\mathbf{A}$.

[2]$T/N \to 0$ is required in case of variables integrated of order one (meaning they are non-stationary but can be rendered to be weakly stationary by first differencing), while $T/N \to \kappa$ for some $0 \leq \kappa < \infty$ is sufficient in case of weakly stationary variables.

expect that the economic variables are at a higher level of economic hierarchy and thus are causal for the risk factors of interest but do not take any feedbacks from them (see e.g. Wilkie, 1984). The assumption of uni-directional causality can be tested through hypothesis tests constructed in Chapter 3. The weak form of exogeneity, however, has to be obtained by construction. In the following subsections, we first discuss the definitions of Granger causality and instantaneous causality, based on which the concept of exogeneity is then defined in the following subsection.

2.2.1 Granger Causality and Instantaneous Causality

Granger (1969) proposed a simple and intuitive definition of a causal relationship which can be easily applied to the context of a VARX* model. Upon assuming that the future can not cause the past in the predictive sense, Granger causality between the vectors of stochastic processes, X_t and Y_t, are defined with respect to the effect that (lack of) past information of one variable vector has on the current value of another variable. The predictor used in Granger (1969) is the optimal (in the sense of minimum mean square error) predictor defined on the set of all past information available in the universe up to time t, $\mathcal{U}_t$, instead of the information set $\mathcal{P}_t$ defined previously in Section 1.1, which consists of only the past information of variables of interest. Originally the definitions are proposed for a bivariate model. Following the description in Lütkepohl (2005), we extend the definition to VAR models of an arbitrary finite dimension. Let $Z_t(h|\mathcal{U}_t)$ denote the optimum h-step predictor of Z_{t+h} at time t and $\Sigma_Z(h|\mathcal{U}_t)$ denote the corresponding mean square error (MSE). We have the following definitions.

Definition 2.1 (Granger Causality for a VAR Model)**.** *The vector stochastic process X_t is said to be Granger-causal for Y_t if $\exists h = 1, 2, \ldots$ such that*

$$\Sigma_Y(h|\mathcal{U}_t) \neq \Sigma_Y(h|\mathcal{U}_t \backslash \{X_s | s \leq t\}), \tag{2.6}$$

where $\mathcal{U}_t \backslash \{X_s | s \leq t\}$ is the set containing all relevant information except that of X_t, and

$$\Sigma_Y(h|\mathcal{U}_t) \leq \Sigma_Y(h|\mathcal{U}_t \backslash \{X_s | s \leq t\}), \tag{2.7}$$

where the inequality is defined in the sense that the matrix forming by subtracting the term on the left hand side from that on the right hand side is positive semidefinite.

Note that when extending the concept of Granger causality to a VAR model, we require both (2.6) and (2.7) to hold for the existence of a causal relationship. This is only to exclude the case of equality in (2.7) since a zero matrix is also positive semidefinite. Despite the elusive notation, this definition simply tells us that adding information on causal variables helps improve the prediction efficiency. When both X_t and Y_t mutually Granger-cause each other, we say that they form a *feedback system*.

In some cases, contemporaneous information regarding one (set of) variable(s) is relevant in making prediction for the other variables in the system. This is the so-called instantaneous causality, defined as follows.

Definition 2.2 (Instantaneous Causality)**.** *There exists instantaneous causality between X_t and Y_t if*

$$\Sigma_Y(1|\mathcal{U}_t) \neq \Sigma_Y(1|\mathcal{U}_t \cup \{X_{t+1}\}), \tag{2.8}$$

and

$$\Sigma_Y(1|\mathcal{U}_t) \leq \Sigma_Y(1|\mathcal{U}_t \cup \{X_{t+1}\}) \tag{2.9}$$

hold.

Note that unlike the case of Granger causality, where the causal relationship can be either one way or mutual, the instantaneous causality must be symmetric. In the following discussion we will see that this is because instantaneous causality exists only when the innovation processes of X_t and Y_t have contemporaneous correlation.

In practice, however, it is not possible to acquire all relevant information. Therefore, regardless of the problems considered, we still use $\mathcal{P}_t$ in place of $\mathcal{U}_t$. Comments on issues caused by such loss of information can also be found in Granger (1969). Moreover, since it is not easy to obtain the optimal predictor among all possible functional forms, including both linear and nonlinear ones, we use the optimal linear predictor instead for a characterization of Granger causality. The reason for using the optimal linear predictor is to ensure

that Granger-noncausality results in the same predictor and MSE for the information sets in comparison[3], so as to simplify the characterization of causality in the context of a VAR model.

2.2.2 Characterizing Causality for VAR Models

In this section we discuss how Granger causality and instantaneous causality are characterized within the context of a VAR model. For convenience, we collect both risk and economic factors in the vector Z_t, such that $Z_t = (X_t', Y_t')'$ and assume a stable VAR(p) model without a constant term

$$Z_t = \begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \sum_{l=1}^{p} \mathbf{\Psi}_l Z_t + \varepsilon_t \tag{2.10}$$

$$:= \sum_{l=1}^{p} \begin{bmatrix} \mathbf{\Psi}_{11,l} & \mathbf{\Psi}_{12,l} \\ \mathbf{\Psi}_{21,l} & \mathbf{\Psi}_{22,l} \end{bmatrix} \begin{bmatrix} X_{t-l} \\ Y_{t-l} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}, \tag{2.11}$$

where ε_t is the innovation process with a nonsingular covariance matrix Σ_ε. In this case both X_t and Y_t are assumed to have zero means. Since (2.11) is stable, it has an MA representation as

$$\begin{aligned} Z_t &= \sum_{i=0}^{\infty} \mathbf{\Theta}_i \varepsilon_{t-i} \\ &:= \sum_{i=0}^{\infty} \begin{bmatrix} \mathbf{\Theta}_{11,i} & \mathbf{\Theta}_{12,i} \\ \mathbf{\Theta}_{21,i} & \mathbf{\Theta}_{22,i} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-i} \\ \varepsilon_{2,t-i} \end{bmatrix}, \end{aligned} \tag{2.12}$$

where the coefficient matrices $\mathbf{\Theta}_i$ are defined using the following recursion formula

$$\begin{aligned} &\mathbf{\Theta}_0 = \mathbf{I}_{k+m}, \\ &\mathbf{\Theta}_i = \sum_{l=1}^{i} \mathbf{\Theta}_{i-l} \mathbf{\Psi}_l, \quad i = 1, 2, \ldots. \end{aligned} \tag{2.13}$$

[3]In our application, the information sets in comparison refer to $\mathcal{P}_t$ and $\mathcal{P}_t \backslash \{X_s | s \leq t\}$

Then the h-step optimal (linear) predictor of Y_t based on (2.12) is given by (see Lütkepohl, 2005, pp. 37 - 38)

$$\begin{aligned} Y_t(h|\mathcal{P}_t) &= [\mathbf{0}\ \mathbf{I}_m] Z_t(h|\mathcal{P}_t) \\ &= \sum_{i=0}^{\infty} \mathbf{\Theta}_{21,i+h} \varepsilon_{1,t-i} + \sum_{i=0}^{\infty} \mathbf{\Theta}_{22,i+h} \varepsilon_{2,t-i}, \end{aligned} \tag{2.14}$$

where $\mathbf{0}$ is zero matrix of a conformable size. Using (2.12) and (2.14), the forecast error is hence

$$Y_{t+h} - Y_t(h|\mathcal{P}_t) = \sum_{i=0}^{h-1} \mathbf{\Theta}_{21,i} \varepsilon_{1,t+h-i} + \sum_{i=0}^{h-1} \mathbf{\Theta}_{22,i} \varepsilon_{2,t+h-i}. \tag{2.15}$$

Since any subprocess of a stable VAR process also admits an MA representation (see Section 2.1.3 of Lütkepohl (2005)), Y_t can be expressed in terms of another independent white noise ξ_t as

$$Y_t = \sum_{i=0}^{\infty} \mathbf{\Xi}_i \xi_{t-i}. \tag{2.16}$$

By the same token, the forecast error of Y_t based on its own information set is given by

$$Y_{t+h} - Y_t(h|\mathcal{P}_t \backslash \{X_s | s \leq t\}) = \sum_{i=0}^{h-1} \mathbf{\Xi}_i \xi_{t+h-i}. \tag{2.17}$$

The definition of Granger-noncausality requires the prediction MSE based on $\mathcal{P}_t$ and $\mathcal{P}_t \backslash \{X_s | s \leq t\}$ to be the same for all h. Thus, we first compare (2.15) with (2.17) for $h = 1$. Since $\mathbf{\Theta}_{21,0} = \mathbf{0}$, we need $\varepsilon_{2,t+1} = \xi_{t+1}$ for non-causality to hold. In this case, the uniqueness of the canonical MA representation (An MA representation of the form in (2.12)) then implies $\mathbf{\Xi}_i = \mathbf{\Theta}_{22,i}$ and $\mathbf{\Theta}_{21,i} = \mathbf{0}$ for all $i = 0, 1, 2, \ldots$. Further it is easy to see that the conditions required for (2.6) and (2.7) to hold for $h = 1$ imply that the same hold for all h. Then the recursion in (2.13) shows that $\mathbf{\Psi}_{21,l} = 0$ for all l. Through the above derivation, we obtain the necessary and sufficient condition for Granger-noncausality defined using the optimal linear predictor, which is summarized in the following proposition (see Corollary 2.2.1 of Lütkepohl (2005)).

Proposition 2.1 (Characterization of Granger-noncausality)**.** *If Z_t is a stable VAR(p) process as in* (2.11) *with a nonsingular covariance matrix, then*

$$\begin{aligned} & Y_t(h|\mathcal{P}_t) = Y_t(h|\mathcal{P}_t \backslash \{X_s | s \leq t\}), h = 1, 2, \dots \\ \Leftrightarrow \quad & \mathbf{\Psi}_{21,l} = \mathbf{0}, l = 1, 2, \dots . \end{aligned} \tag{2.18}$$

The above proposition states that as long as X_t is not Granger-causal for Y_t, there is no need for us to specify parameters of X_t in the model for the process Y_t. Therefore the concept of exogeneity defined later helps reduce the number of parameters to be estimated in the system.

Next, we are to characterize instantaneous causality. Even though ε_t is assumed to be an independent innovation process, there may potentially be non-zero correlation between its components. To work with residual vectors with uncorrelated components, we first apply a Cholesky decomposition to the positive definite covariance matrix Σ_ε to obtain $\Sigma_\varepsilon = CC'$, where C is an upper triangular nonsingular matrix. Then (2.12) becomes

$$Z_t = \sum_{i=0}^{\infty} \mathbf{\Theta}_i \varepsilon_{t-i} = \sum_{i=0}^{\infty} \mathbf{\Theta}_i C C^{-1} \varepsilon_{t-i} = \sum_{i=0}^{\infty} \mathbf{P}_i \omega_{t-i}, \tag{2.19}$$

where $\mathbf{P}_i = \mathbf{\Theta}_i C$ and $\omega_t = C^{-1}\varepsilon_t$. ω_t is a white noise process with a covariance matrix $\Sigma_\omega = \mathbf{I}_{k+m}$. The lower triangular form of C allows us to partition Z_t as

$$Z_t = \begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \sum_{i=0}^{\infty} \begin{bmatrix} \mathbf{P}_{11,i} & \mathbf{0} \\ \mathbf{P}_{21,i} & \mathbf{P}_{22,i} \end{bmatrix} \begin{bmatrix} \omega_{1,t-i} \\ \omega_{2,t-i} \end{bmatrix}. \tag{2.20}$$

From (2.20) we can derive the 1-step optimal predictor of Y_t taking into account the information of X_{t+1} as

$$\begin{aligned} Y_t(1|\mathcal{P}_t \cup \{X_{t+1}\}) &= Y_t(1|\mathcal{P}_t \cup \{\omega_{1,t+1}\}) \\ &= \mathbf{P}_{21,0}\omega_{1,t+1} + \sum_{i=1}^{\infty} \mathbf{P}_{21,i}\omega_{1,t+1-i} + \sum_{i=1}^{\infty} \mathbf{P}_{22,i}\omega_{2,t+1-i}, \end{aligned} \tag{2.21}$$

while its counterpart based solely on $\mathcal{P}$ is

$$Y_t(1|\mathcal{P}_t) = \sum_{i=1}^{\infty} \mathbf{P}_{21,i}\omega_{1,t+1-i} + \sum_{i=1}^{\infty} \mathbf{P}_{22,i}\omega_{2,t+1-i}. \tag{2.22}$$

To equate (2.21) and (2.22), we require $\mathbf{P}_{21,0} = \mathbf{0}$ and this equality implies that there is no instantaneous causality between X_t and Y_t. We also summarize this result in the following proposition (see Proposition 2.3 of Lütkepohl (2005)).

Proposition 2.2 (Characterization of Instantaneous Noncausality)**.** *If Z_t is a stable VAR(p) process as in* (2.11) *with a nonsingular covariance matrix, then the absence of instantaneous causality between X_t and Y_t is equivalent to*

$$\mathbb{E}(\varepsilon_{1t}\varepsilon_{2t}') = \mathbf{0}. \tag{2.23}$$

Proposition 2.2 shows that instantaneous causality is equivalent to a contemporaneous correlation between the white noises of the two VAR processes in question. Thus instantaneous causality must be symmetric.

Regardless of the criticisms levied against these definitions of a causal relationship, at the very least they formalize a way to analyze whether the past information of a certain set of variables will be helpful in making predictions about the other variables. Therefore, Granger-causality analysis, together with the definition of exogeneity discussed below, provide a theoretical support for the parsimonious modeling assumption posited in the VARX* specification.

2.2.3 Exogeneity

In Section 2.1 we mentioned that under certain granularity conditions, the weak exogeneity of star variables holds asymptotically and thus we can ignore the marginal distribution of these variables in estimation. Moreover, the GVAR approach provides us with the flexibility to model economic factors as (strongly) exogenous variables. As such, the prediction can then be carried out step-wise, where the economic variables are first projected forward with separate marginal models and the forecasts of portfolio risk factors are then conducted by conditioning on the prediction of economic variables, which are assumed to be strongly exogenous in this case.

We conclude this section by discussing respective definitions for weak and strong exogeneity. Weak exogeneity is concerned mainly with estimation and statistical inference on parameters of interest using conditional distributions only, while incurring no loss of information. Thus it is important to define which parameters are of interest. Parameters may be of interest

because they are directly related to economic theories under study or are presented in the final model of our interest. For example, in the context of conditional scenario generation, the ultimate objective is to obtain the conditional distribution for portfolio risk factors as in (1.1). Thus if the true underlying joint system is the VAR model given in (2.11), then the parameters of interest are in λ_1 as defined in (2.29) below.

Using the notation defined previously, the definition for weak exogeneity given in Engle, Hendry, and Richard (1983) is restated below.

Definition 2.3 (Weak Exogeneity)**.** *Let $D(\bullet;\bullet|\bullet)$ denote the conditional density function. Then Y_t is weakly exogenous with respect to parameters of interest over the sample period if and only if the distribution function of the random vector Z_t can be decomposed into a conditional distribution for X_t and a marginal distribution for Y_t, i.e.*

$$D(Z_t;\lambda|\mathcal{P}_{t-1}) = D(X_t;\lambda_1|Y_t,\mathcal{P}_{t-1})D(Y_t;\lambda_2|\mathcal{P}_{t-1}) \tag{2.24}$$

where the parameter vector λ is partitioned into a variation-free sub-vector (λ_1,λ_2), i.e. $(\lambda_1,\lambda_2)\in\Lambda_1\times\Lambda_2$; and the vector for parameters of interest is a function of λ_1.

Note that condition (2.24) ensures that the likelihood function $L(\lambda;\mathcal{P}_T)$ can be decomposed as

$$L(\lambda;\mathcal{P}_T) = \prod_{t=1}^{T} D(X_t;\lambda_1|Y_t,\mathcal{P}_{t-1})D(Y_t;\lambda_2|\mathcal{P}_{t-1}) =: L_1(\lambda_1;\mathcal{P}_T)L_2(\lambda_2;\mathcal{P}_T). \tag{2.25}$$

Together with the variation-free condition, the aforementioned decomposition facilitates an independent analysis for the two partial likelihoods L_1 and L_2. Further, if the parameters in λ_1 provide all information regarding parameters of interest, inference can be conducted based solely on L_1 and we can thus exclude L_2 from our analysis, which substantially improves computational efficiency.

However, whether Y_t can be treated as weakly exogenous depends on the selection of parameters of interest. For the VAR model in (2.11), partition Σ_ε conformably with $Z_t = (X_t', Y_t')'$ as

$$\Sigma_\varepsilon = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}. \tag{2.26}$$

Then by assuming normality for ε_t, the equation system for X_t given Y_t can be written as

$$X_t = \mathbb{A}Y_t + \sum_{l=1}^{p} \mathbb{B}_l X_{t-l} + \sum_{l=1}^{p} \mathbb{C}_l Y_{t-l} + \xi_t, \tag{2.27}$$

where $\mathbb{A} = \Sigma_{12}\Sigma_{22}^{-1}$, $\mathbb{B}_l = \Psi_{11,l} - \Sigma_{12}\Sigma_{22}^{-1}\Psi_{21,l}$, $\mathbb{C}_l = \Psi_{12,l} - \Sigma_{12}\Sigma_{22}^{-1}\Psi_{22,l}$ and ξ_t follow $N(0, \Sigma_\xi)$, where $\Sigma_\xi = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. And the marginal distribution of Y_t is then

$$Y_t = \sum_{l=1}^{p} \mathbf{\Psi}_{21,l} X_{t-l} + \sum_{l=1}^{p} \mathbf{\Psi}_{22,l} Y_{t-l} + \varepsilon_{2t}, \qquad \varepsilon_{2t} \sim N(0, \Sigma_{22}). \tag{2.28}$$

Note that ε_{2t} and ξ_t are independent by construction, which ensures the separation specified in (2.24) to have a conditional density for X_t of the form

$$D(X_t; \lambda_1 | Y_t, \mathcal{P}_{t-1}) = N\left(\mathbb{A}Y_t + \sum_{l=1}^{p} \mathbb{B}_l X_{t-l} + \sum_{l=1}^{p} \mathbb{C}_l Y_{t-l}, \Sigma_\xi\right), \tag{2.29}$$

where $\lambda_1 = (\mathbb{A}, \mathbb{B}_l, \mathbb{C}_l, \Sigma_\xi), l = 1, 2, \ldots, p$. The parameter vector for the marginal density, $D(Y_t; \lambda_2 | \mathcal{P}_{t-1})$, is then $\lambda_2 = (\Psi_{21,l}, \Psi_{22,l}, \Sigma_{22}), l = 1, 2, \ldots, p$. It is easy to see that λ_1, λ_2 are variation-free. Thus if we are only interested in λ_1, Y_t is by construction weakly exogenous, even though bi-directional Granger causality exists between X_t and Y_t. In our application, weak exogeneity holds true by construction for economic factors in the VARX* model. But weak exogeneity of star variables depends on certain granularity conditions (e.g. (2.3) and (2.4)) and holds only asymptotically.

The concept of strong exogeneity is built upon its weak counterpart and further requires Granger noncausality from conditioning variables to conditioned variables, which ensures that the forecast of conditioned variables can be carried out independently of the conditioning variables. Here we restate the definition for strong exogeneity in Engle, Hendry, and Richard (1983).

Definition 2.4 (Strong Exogeneity)**.** *Y_t is strongly exogenous with respect to the parameters of interest over the sample period if and only if Y_t is weakly exogenous and Y_t is not Granger-caused by X_t.*

As shown in Proposition 2.1, the additional condition required for strong exogeneity ensures that the past realization of X_t does not affect the future evolution of Y_t, which lays the foundation for forecasting Y_t independently first before carrying out the conditional forecast of X_t.

2.3 Estimation of VARX*

We restrict our attention to the stable VAR processes, for which the prevalent estimation approaches include least squares, the Yule-Walker equation and maximum likelihood. In this study, we focus on applying an ordinary least squares (OLS) approach, which, due to the result in Zellner (1962), is known as the generalized least squares estimator. The OLS approach is applied to mean-adjusted data to avoid estimating the intercept jointly with other parameters, so as to reduce computational complexity. In this case, the variables X_t, X_t^* and Y_t in the VARX* model are implicitly assumed to have zero mean (or mean-adjusted) and thus the model specified below does not contain an intercept term. Since both X_t^* and Y_t are treated as weakly exogenous, we collect them in a single Y_t vector and focus our attention on estimation of the model of the following form

$$X_t = \sum_{l=1}^{p} \mathbf{\Phi}_l X_{t-l} + \sum_{l=0}^{s} \mathbf{D}_l Y_{t-l} + u_t, \tag{2.30}$$

where X_t are mean-adjusted and u_t is a white noise process, $p \geq 1$ and $s \geq 0$. The subscripts, i, for each VARX* are omitted for expositional convenience.

For simplicity, we assume that presample values x_{-p+1}, x_{-p+2}, ..., x_0 and y_{-s+1}, y_{-s+2}, ..., y_0 are available. This assumption is not at all restrictive since so long as we have a sample of appropriate length, it is always possible to select the datum as a starting point such that the required length of presample is available. Further, the following notations are defined so

that we can re-write the VARX* model in (2.30) in a VAR(1) form as,

$$
\begin{aligned}
&\mathbf{X} := (x_1, \ldots, x_T) && (k \times T),\\
&\mathbf{B} := (\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_p, \mathbf{D}_0, \ldots, \mathbf{D}_s) && (k \times (kp + ms)),\\
&Z_t := \begin{bmatrix} x_t \\ \vdots \\ x_{t-p+1} \\ y_{t+1} \\ \vdots \\ y_{t-s+1} \end{bmatrix} && ((kp + ms) \times 1),\\
&\mathbf{Z} := (Z_0, \ldots, Z_{T-1}) && ((kp + ms) \times T),\\
&\mathbf{U} := (u_1, \ldots, u_T) && (k \times T),\\
&\mathbf{x} := \operatorname{vec}(X) && (kT \times 1),\\
&\beta := \operatorname{vec}(B) && (k(kp + ms) \times 1),\\
&\mathbf{u} := \operatorname{vec}(\mathbf{U}) && (kT \times 1).
\end{aligned}
\tag{2.31}
$$

With the notation defined previously, the VAR(1) form of VARX* can be expressed as

$$\mathbf{X} = \mathbf{BZ} + \mathbf{U} \tag{2.32}$$

or

$$
\begin{aligned}
\operatorname{vec}(\mathbf{X}) &= \operatorname{vec}(\mathbf{BZ}) + \operatorname{vec}(\mathbf{U})\\
&= (\mathbf{Z}' \otimes \mathbf{I}_k) \operatorname{vec}(B) + \operatorname{vec}(\mathbf{U}),
\end{aligned}
$$

which, in our notation, is

$$\mathbf{x} = (\mathbf{Z}' \otimes \mathbf{I}_k)\beta + \mathbf{u}. \tag{2.33}$$

The definition and properties of Kronecker product ($\otimes$) and vectorization operators can be found in Appendix A.1 and Appendix A.2.

2.3.1 Estimation of Process Mean

In general the mean of a process is unknown and thus needs to be estimated in order to transform the data so that (2.30) can be fitted to it. Let $\tilde{X}_t := X_t + \mu_X$ and $\tilde{Y}_t := Y_t + \mu_Y$

denote the original processes. Further, define the following quantities:

$$\bar{X} = \frac{1}{T}\sum_{t=1}^{T} \tilde{X}_t, \tag{2.34}$$

$$\bar{Y} = \frac{1}{T}\sum_{t=1}^{T} \tilde{Y}_t. \tag{2.35}$$

Using (2.30) and (2.34), we have

$$\begin{aligned}\bar{X} &= \frac{1}{T}\sum_{t=1}^{T}\left(\mu_X + \sum_{l=1}^{p}\mathbf{\Phi}_l\left(\tilde{X}_{t-l} - \mu_X\right) + \sum_{l=0}^{s}\mathbf{D}_l\left(\tilde{Y}_{t-l} - \mu_Y\right) + u_t\right) \\ &= \mu_X + \sum_{l=1}^{p}\mathbf{\Phi}_l\bar{X} + \frac{1}{T}\sum_{l=1}^{p}\mathbf{\Phi}_l\sum_{j=0}^{l-1}\left(\tilde{X}_{-j} - \tilde{X}_{T-j}\right) - \sum_{l=1}^{p}\mathbf{\Phi}_l\mu_X \\ &\quad + \sum_{l=0}^{s}D_l\bar{Y} + \sum_{l=1}^{s}\mathbf{D}_l\sum_{j=0}^{l-1}\left(Y_{-j} - Y_{T-j}\right) - \sum_{l=0}^{s}\mathbf{D}_l\mu_Y + \frac{1}{T}\sum_{t=1}^{T}u_t.\end{aligned}$$

Hence,

$$\mathbf{\Phi}(\bar{X} - \mu_X) = \frac{1}{T}(S_T + R_T) + \mathbf{D}(\bar{Y} - \mu_Y) + \frac{1}{T}\sum_{t=1}^{T}u_t, \tag{2.36}$$

where $\mathbf{\Phi} = \mathbf{I}_k - \mathbf{\Phi}_1 - \ldots - \mathbf{\Phi}_p$, $\mathbf{D} = \sum_{i=0}^{s}\mathbf{D}_i$, $S_T = \sum_{l=1}^{p}\mathbf{\Phi}_l\sum_{j=0}^{l-1}\left(\tilde{X}_{-j} - \tilde{X}_{T-j}\right)$ and $R_T = \sum_{l=1}^{s}\mathbf{D}_l\sum_{j=0}^{l-1}\left(Y_{-j} - Y_{T-j}\right)$. Under the assumptions made so far, we have

$$\mathbb{E}\left(\frac{1}{\sqrt{T}}S_T\right) = 0, \tag{2.37}$$

and

$$\lim_{T\to\infty}\mathrm{Var}\left(\frac{1}{\sqrt{T}}S_T\right) = \lim_{T\to\infty}\frac{1}{T}\mathrm{Var}(S_T) = 0 \tag{2.38}$$

for X_t; additionally, due to the central limit theorem (see e.g. Fuller, 2009)

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t \xrightarrow{d} N(0, \Sigma_u). \tag{2.39}$$

On the other hand, assume that $\tilde{Y}_t$ follows

$$\tilde{Y}_t - \mu_Y = \sum_{l=1}^{q}\mathbf{A}_l(\tilde{Y}_t - \mu_Y) + v_t, \tag{2.40}$$

where v_t is a white noise process with a covariance matrix Σ_v. Moreover, $(u_t', v_t')'$ has a covariance matrix of Σ. Presample observations, y_{-q+1}, y_{-q+2}, ..., y_0, some or all of which may overlap the presample observations required previously, are also assumed to be available. We further have

$$\mathbb{E}\left(\frac{1}{\sqrt{T}}R_T\right) = 0, \tag{2.41}$$

and

$$\lim_{T\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{T}}R_T\right) = \lim_{T\to\infty} \frac{1}{T}\mathrm{Var}(R_T) = 0. \tag{2.42}$$

Following the same derivation as before we obtain

$$\mathbf{A}(\bar{Y} - \mu_Y) = \frac{1}{T}P_T + \frac{1}{T}\sum_{t=1}^{T} v_t, \tag{2.43}$$

where $\mathbf{A} = \mathbf{I}_k - \mathbf{A}_1 - \ldots - \mathbf{A}_p$ and $P_T = \sum_{l=1}^{q} \mathbf{A}_l \sum_{j=0}^{l-1} (Y_{-j} - Y_{T-j})$. Using the same rationale, we have

$$\mathbb{E}\left(\frac{1}{\sqrt{T}}P_T\right) = 0, \tag{2.44}$$

and

$$\lim_{T\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{T}}P_T\right) = \lim_{T\to\infty} \frac{1}{T}\mathrm{Var}(P_T) = 0. \tag{2.45}$$

Assuming invertibility of $\mathbf{A}$, by (2.36) and (2.43), we have

$$\sqrt{T}\mathbf{\Phi}(\bar{X} - \mu_X) = \frac{1}{\sqrt{T}}(S_T + R_T + \mathbf{D}\mathbf{A}^{-1}P_T) + \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} u_t + \sum_{t=1}^{T} w_t\right), \tag{2.46}$$

where $w_t = \mathbf{D}\mathbf{A}^{-1}v_t$ is white noise with a covariance matrix of $\Sigma_w := \mathbf{D}\mathbf{A}^{-1}\Sigma_v(\mathbf{D}\mathbf{A}^{-1})'$, which is independent of u_t. Thus terms in the first bracket converge in probability to 0. Let $\mathbf{E} = [\mathbf{I}_k\ \mathbf{D}\mathbf{A}^{-1}]$, due to the central limit theorem,

$$\frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} u_t + \sum_{t=1}^{T} w_t\right) \xrightarrow{d} N\left(0, \mathbf{E}\Sigma\mathbf{E}'\right). \tag{2.47}$$

Hence we summarize the asymptotic distribution of the mean estimator in the following proposition.

Proposition 2.3 (Asymptotic Property of Sample Mean)**.** *For $\tilde{X}_t$ in the VARX* model with common variables, if the (strongly) exogenous variables $\tilde{Y}_t$ follow a stable VAR process defined in* (2.40)*, then*

$$\sqrt{T}(\bar{X} - \mu_X) \xrightarrow{d} N(0, \Sigma_{\bar{X}}), \tag{2.48}$$

where

$$\Sigma_{\bar{X}} = \mathbf{\Phi}^{-1}\mathbf{E}\Sigma\mathbf{E}'(\mathbf{\Phi}^{-1})'. \tag{2.49}$$

The above proposition results from (2.46), (2.47) and properties of stochastic convergence listed in Appendix B. The consistency of $\bar{X}$ is also confirmed by the Proposition 2.3.

2.3.2 Ordinary Least Squares Estimator

For all of the derivations in this section, we assume that variables in both X_t and Y_t are independent of u_{t+i} for $i > 0$. Using the expression in (2.33), the OLS estimator is obtained by minimizing:

$$\begin{aligned} S(\beta) = \mathbf{u}'\mathbf{u} &= [\mathbf{x} - (\mathbf{Z}' \otimes \mathbf{I}_k)\beta]'[\mathbf{x} - (\mathbf{Z}' \otimes \mathbf{I}_k)\beta] \\ &= \mathbf{x}'\mathbf{x} - 2\beta'(\mathbf{Z} \otimes \mathbf{I}_k)\mathbf{x} + \beta'(\mathbf{Z}\mathbf{Z}' \otimes \mathbf{I}_k)\beta. \end{aligned} \tag{2.50}$$

Taking the first order derivative of $S(\beta)$ gives

$$\frac{\partial S(\beta)}{\partial \beta} = 2(\mathbf{Z}\mathbf{Z}' \otimes \mathbf{I}_k)\beta - 2(\mathbf{Z} \otimes \mathbf{I}_k)\mathbf{x}. \tag{2.51}$$

Setting it to zero and solving for β yields the estimator

$$\hat{\beta} = ((\mathbf{Z}\mathbf{Z}')^{-1}\mathbf{Z} \otimes \mathbf{I}_k)\mathbf{x}. \tag{2.52}$$

Then the Hessian matrix

$$\frac{\partial^2 S(\beta)}{\partial \beta \partial \beta'} = 2(\mathbf{Z}\mathbf{Z}' \otimes \mathbf{I}_k) \tag{2.53}$$

is positive definite, which confirms that $\hat{\beta}$ is indeed the estimator.

Substituting (2.33) into (2.52) yields an alternative expression for the estimator, i.e.

$$\hat{\beta} = \beta + ((\mathbf{Z}\mathbf{Z}')^{-1}\mathbf{Z} \otimes \mathbf{I}_k)\mathbf{u}. \tag{2.54}$$

Based on (2.54), we reproduce the derivation in Section 3.2.2 of Lütkepohl (2005) for the asymptotic distribution of $\hat{\beta}$. The prerequisites for the asymptotic distribution to exist include:

$$\Gamma := \operatorname{plim} \frac{\mathbf{Z}\mathbf{Z}'}{T} \text{ exists and is nonsingular,} \tag{2.55}$$

and

$$\frac{1}{\sqrt{T}}(\mathbf{Z} \otimes \mathbf{I}_k)\mathbf{u} \xrightarrow[T\to\infty]{d} N(0, \Gamma \otimes \Sigma_u). \tag{2.56}$$

Due to Lemma 3.1 of Lütkepohl (2005), one sufficient condition for (2.55) and (2.56) to hold for a stable VAR(p) defined in (2.30) is that its residual is a *standard white noise* as defined in the following definition.

Definition 2.5 (Standard White Noise)**.** *A white noise process $u_t = (u_{1t}, \ldots, u_{kt})$ is standard if u_t is a continuous random vector satisfying $\mathbb{E}(u_t) = 0$, $\mathbb{E}(u_t u_t') = \Sigma_u$, which is nonsingular, u_t, u_s are independent for $t \neq s$, and, for some finite constant c,*

$$\mathbb{E}|u_{it}u_{jt}u_{ht}u_{gt}| \leq c \qquad \text{for } i, j, h, g = 1, \ldots, k, \text{ and all } t. \tag{2.57}$$

All of the conditions in Definition 2.5 are familiar except the last one, which requires that all fourth moments exist and are bounded. A typical white noise that is standard is an independent Gaussian process, i.e. $u_t \sim N(0, \Sigma_u)$.

Note that in the context of VARX* as in (2.30), so long as the weak exogeneity of Y_t holds, the VARX* model can be written as a VAR model. Thus conditions (2.55) and (2.56) hold for VARX* with a standard white noise, by using (2.54), we have

$$\begin{aligned}\sqrt{T}(\hat{\beta} - \beta) &= \sqrt{T}((\mathbf{Z}\mathbf{Z}')^{-1}\mathbf{Z} \otimes \mathbf{I}_k)\mathbf{u} \\ &= \left(\left(\frac{1}{T}\mathbf{Z}\mathbf{Z}'\right)^{-1} \otimes \mathbf{I}_k\right) \frac{1}{\sqrt{T}}(\mathbf{Z} \otimes \mathbf{I}_k)\mathbf{u}.\end{aligned} \tag{2.58}$$

Thus $\sqrt{T}(\hat{\beta} - \beta)$ has the same asymptotic distribution as

$$\left(\operatorname{plim}\left(\frac{1}{T}\mathbf{Z}\mathbf{Z}'\right)^{-1} \otimes \mathbf{I}_k\right) \frac{1}{\sqrt{T}}(\mathbf{Z} \otimes \mathbf{I}_k)\mathbf{u} = \left(\Gamma^{-1} \otimes \mathbf{I}_k\right) \frac{1}{\sqrt{T}}(\mathbf{Z} \otimes \mathbf{I}_k)\mathbf{u}. \tag{2.59}$$

Then (2.59) and condition (2.56) show that

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow[T\to\infty]{d} N(0, \Sigma_\beta), \tag{2.60}$$

where

$$\Sigma_\beta = (\Gamma^{-1} \otimes \mathbf{I}_k)(\Gamma \otimes \Sigma_u)(\Gamma^{-1} \otimes \mathbf{I}_k) = \Gamma^{-1} \otimes \Sigma_u. \tag{2.61}$$

The above result is summarized below.

Proposition 2.4 (Asymptotic Distribution of OLS Estimator)**.** *For a stable VARX* model defined in* (2.30) *with standard white noise residuals, and with conditions specified in* (2.55) *and* (2.56)*, the OLS estimator*

$$\hat{\beta} = ((\mathbf{Z}\mathbf{Z}')^{-1}\mathbf{Z} \otimes \mathbf{I}_k)\mathbf{x} \tag{2.62}$$

has the asymptotic distribution

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow[T\to\infty]{d} N(0, \Sigma_\beta). \tag{2.63}$$

Note that Γ and Σ_u need to be estimated before we can carry out statistical tests. If condition (2.55) holds, Γ can be estimated consistently with

$$\hat{\Gamma} = \frac{\mathbf{Z}\mathbf{Z}'}{T}. \tag{2.64}$$

For simplicity of derivation we first give the matrix representation of the OLS estimator, $\hat{\mathbf{B}}$, i.e.

$$\begin{aligned}
\operatorname{vec}(\hat{\mathbf{B}}) = \hat{\beta} &= ((\mathbf{Z}\mathbf{Z}')^{-1}\mathbf{Z} \otimes \mathbf{I}_k)\mathbf{x} \\
&= ((\mathbf{Z}\mathbf{Z}')^{-1}\mathbf{Z} \otimes \mathbf{I}_k)\operatorname{vec}(\mathbf{x}) \\
&= \operatorname{vec}(\mathbf{X}\mathbf{Z}'(\mathbf{Z}\mathbf{Z}')^{-1}).
\end{aligned}$$

Thus,

$$\begin{aligned}
\hat{\mathbf{B}} &= \mathbf{X}\mathbf{Z}'(\mathbf{Z}\mathbf{Z}')^{-1} \\
&= (\mathbf{B}\mathbf{Z} + \mathbf{U})\mathbf{Z}'(\mathbf{Z}\mathbf{Z}')^{-1} \\
&= \mathbf{B} + \mathbf{U}\mathbf{Z}'(\mathbf{Z}\mathbf{Z}')^{-1}.
\end{aligned}$$

Since $\Sigma_u = \mathbb{E}(u_t u_t')$, a plausible estimator is seen to be

$$\tilde{\Sigma}_u = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t \hat{u}_t' = \frac{1}{T}\hat{\mathbf{U}}\hat{\mathbf{U}}' = \frac{1}{T}(\mathbf{X} - \hat{\mathbf{B}}\mathbf{Z})(\mathbf{X} - \hat{\mathbf{B}}\mathbf{Z})'. \tag{2.65}$$

Since we treat all of the regressors as predetermined, an adjustment for degrees of freedom is needed in order to obtain an unbiased estimator, namely

$$\hat{\Sigma}_u = \frac{T}{T - kp - 1}\tilde{\Sigma}_u. \tag{2.66}$$

Evidently the two estimators, $\hat{\Sigma}_u$ and $\tilde{\Sigma}_u$, are asymptotically equivalent. The consistency of these two estimators is demonstrated in Corollary 3.2.1 of Lütkepohl (2005).

2.4 Stacking VARX*

Once the unknown parameters in each VARX* have been estimated, we can then stack these models to form one large GVAR model for statistical inference, impulse response analysis, simulation and forecasting. First we define a $(k_i + k^*) \times k$ link matrix, $\mathbf{W}_i := (\mathbf{E}_i', \tilde{\mathbf{W}}_i')'$, where $\mathbf{E}_i'$ select X_{it} out of X_t, namely $X_{it} = \mathbf{E}_i' X_t$. Then variables of the ith unit and corresponding star variables can be linked to the global vector in the following way

$$U_{it} := (X_{it}', X_{it}^{*\prime})' = \mathbf{W}_i X_t.$$

Define $p = \max(p_i, q_i)$ for all $i = 1, 2, \ldots, N$ and

$$\mathbf{A}_{i0} = (\mathbf{I}_{k_i}, -\mathbf{\Lambda}_{i0}) \tag{2.67}$$

$$\mathbf{A}_{il} = (\mathbf{\Phi}_{il}, \mathbf{\Lambda}_{il}), \tag{2.68}$$

for $l = 1, \ldots p$, where $\Phi_{il} = \mathbf{0}$ for $l > p_i$ and $\mathbf{\Lambda}_{il} = \mathbf{0}$ for $l > q_i$. Rearranging contemporaneous star variables in (2.5) we obtain

$$\mathbf{A}_{i0}\mathbf{W}_i X_t = \sum_{l=1}^{p} \mathbf{A}_{il}\mathbf{W}_i X_{t-l} + \mathbf{C}_i \varepsilon_{it}. \tag{2.69}$$

Stacking all units in the representation of (2.69) yields

$$\mathbf{G}_0 X_t = \sum_{l=1}^{p} \mathbf{G}_l X_{t-l} + \mathbf{C}\varepsilon_t \tag{2.70}$$

where $\varepsilon_t = (\varepsilon'_{1t}, \ldots, \varepsilon'_{Nt})'$, and

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{C}_N \end{bmatrix} \qquad (k \times k), \tag{2.71}$$

$$\mathbf{G}_l = \begin{bmatrix} \mathbf{A}_{1l}\mathbf{W}_1 \\ \mathbf{A}_{2l}\mathbf{W}_2 \\ \vdots \\ \mathbf{A}_{Nl}\mathbf{W}_N \end{bmatrix} \qquad (k \times k). \tag{2.72}$$

Note that if we model the evolution of economic factors and market factors using a model of the form in (1.3), then Y_t is in fact modeled as one (or more) cross-sectional unit(s) and is contained in X_t in (2.70). Otherwise if economic factors are modeled as exogenous, we simply append (2.70) to include Y_t as a common factor, i.e.

$$\mathbf{G}_0 X_t = \sum_{l=1}^{p} \mathbf{G}_l X_{t-l} + \mathbf{D}_0 Y_t + \sum_{l=1}^{s} \mathbf{D}_l Y_{t-l} + \mathbf{C}\varepsilon_t \tag{2.73}$$

where $s = \max_i(s_i)$ and

$$\mathbf{D}_l = \begin{bmatrix} \mathbf{D}_{1l} \\ \mathbf{D}_{2l} \\ \vdots \\ \mathbf{D}_{Nl} \end{bmatrix} \qquad (k \times m), \tag{2.74}$$

where similarly for $l > s_i$, $D_{il} = \mathbf{0}$.

Since our ultimate task is to conduct risk analytics with this model, it is very important that the matrix $\mathbf{G}_0$ is invertible in both (2.70) and (2.73) so that we can solve the system simultaneously. Otherwise one must resort to more advanced techniques for simulation (see e.g. Bańbura, Giannone, and Lenza, 2015). Given that $\mathbf{G}_0$ is invertible, we have

$$X_t = \sum_{l=1}^{p} \mathbf{F}_l X_{t-l} + \mathbf{Q}\varepsilon_t \tag{2.75}$$

$$X_t = \sum_{l=1}^{p} \mathbf{F}_l X_{t-l} + \mathbf{H}_0 Y_t + \sum_{l=1}^{s} \mathbf{H}_l Y_{t-l} + \mathbf{Q}\varepsilon_t \tag{2.76}$$

where $\mathbf{F}_l = \mathbf{G}_0^{-1}\mathbf{G}_l$, $\mathbf{H}_l = \mathbf{G}_0^{-1}\mathbf{D}_l$ and $\mathbf{Q} = \mathbf{G}_0^{-1}\mathbf{C}$.

In the case that $\mathbf{G}_0$ is not invertible, $\text{rank}(\mathbf{G}_0) = k - m < k$ for some $m > 0$ and the systems in (2.70) and (2.73) are undetermined. As suggested by Chudik, Grossman, and Pesaran (2016), m additional equations are needed to complete the system. It is recommended that the equations of cross-sectional averages in a VAR form should be good candidates to consider.

2.5 Conditional Forecasting and Simulation

The solved GVAR systems in (2.75) and (2.76) are essentially VAR models. Conditional forecasting using VAR models based on fixed (referred to as hard condition) or a range (referred to as soft condition) of future values of variables has been thoroughly studied in Waggoner and Zha (1999). Subsequently a singular value decomposition based approach is proposed in Jarociński (2010) to improve the computational efficiency in the case where the number of conditioning path is small.

For a GVAR model with common variables, the approach is summarized as follows.

Re-write the augmented GVAR model defined by (2.76) in its companion form

$$\tilde{X}_t = \mathbf{F}\tilde{X}_{t-1} + \mathbf{H}\tilde{Y}_t + \tilde{\mathbf{Q}}\tilde{\varepsilon}_t \tag{2.77}$$

with the following newly defined matrices

$$\tilde{X}_t = \begin{bmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{bmatrix} \qquad (pk \times 1),$$

$$\tilde{Y}_t = \begin{bmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-s} \end{bmatrix} \qquad ((s+1)m \times 1),$$

$$\tilde{\varepsilon}_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (pk \times 1),$$

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_1 & \dots & \mathbf{F}_p \\ \mathbf{I} & & \mathbf{0} \end{bmatrix} \qquad (pk \times pk),$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_0 & \dots & \mathbf{H}_s \\ \mathbf{0} & & \mathbf{0} \end{bmatrix} \qquad (pk \times (s+1)m),$$

$$\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \qquad (pk \times pk).$$

For $s \leq h$ the s-step forecast of $\tilde{X}$ is thus

$$\tilde{X}_{T+s} = \mathbf{F}^s \tilde{X}_T + \sum_{l=0}^{s-1} \mathbf{F}^l \mathbf{H} \tilde{Y}_{T+s-l} + \sum_{l=0}^{s-1} \mathbf{F}^l \tilde{\mathbf{Q}} \tilde{\varepsilon}_{T+s-l}. \tag{2.78}$$

Collecting first k rows of (2.78) gives

$$X_{T+s} = \mathbf{F}^s_{(1:k,\cdot)} \tilde{X}_T + \sum_{l=0}^{s-1} \mathbf{\Psi}_l \tilde{Y}_{T+s-l} + \sum_{l=0}^{s-1} \mathbf{\Theta}_l \varepsilon_{T+s-l}, \tag{2.79}$$

where $\mathbf{F}^s_{(1:k,\cdot)}$ denotes the first k rows of $\mathbf{F}^s$, $\mathbf{\Psi}_l$ denotes the first k rows of $\mathbf{F}^l\mathbf{H}$ and $\mathbf{\Theta}_l$ denotes the upper left $k \times k$ of $\mathbf{F}^l\tilde{\mathbf{Q}}$. Suppose that we have information of a complete future path of

$\mathcal{Y}$, then the conditional forecast is

$$\mathbb{E}[X_{T+s}|\mathcal{P}_T, \mathcal{I}_{T,h}] = \mathbf{F}^s_{(1:k,\cdot)}\tilde{x}_T + \sum_{l=0}^{s-1} \mathbf{\Psi}_l \tilde{y}_{T+h-l}. \tag{2.80}$$

If we only have partial information on the future path of $\mathcal{Y}$, some terms in the summation of Y_t need to be replaced with the corresponding forecasts, which means a separate VAR model for $\mathcal{Y}$ is required.

From (2.80), it is easy to see that the conditional joint distribution function in (1.1) is $N(\mathbb{E}[X_{T+s}|\mathcal{P}_T, \mathcal{I}_{T,h}], \sum_{l=0}^{s-1} \mathbf{\Theta}_l (\sum_{l=0}^{s-1} \mathbf{\Theta}_l)')$. However, from the perspective of computational efficiency, it is recommended to first sample from i.i.d normal distributions and then transform the realization using $\mathbf{\Theta}_l$'s as in (2.79).

The more challenging task is to conduct forecasting with VAR model as in (2.75), where we condition it on a subset of variables in X_t. This time we start from (2.79). Since Y_t has been included in X_t, we can leave out the Y_t terms to obtain

$$X_{T+s} = \mathbf{F}^s_{(1:k,\cdot)}\tilde{X}_T + \sum_{l=0}^{s-1} \mathbf{\Theta}_l \varepsilon_{T+s-l}. \tag{2.81}$$

Stack (2.81) for all $s = 1, \ldots, h$,

$$X = \mathcal{F}\tilde{X}_T + \mathbf{R}\varepsilon \tag{2.82}$$

with the following newly defined matrices

$$X = \begin{bmatrix} X_{T+1} \\ \vdots \\ X_{T+h} \end{bmatrix} \qquad (hk \times 1),$$

$$\varepsilon = \begin{bmatrix} \varepsilon_{T+1} \\ \vdots \\ \varepsilon_{T+h} \end{bmatrix} \qquad (hk \times 1),$$

$$\mathcal{F} = \begin{bmatrix} \mathbf{F}_{(1:k,\cdot)} \\ \vdots \\ \mathbf{F}^h_{(1:k,\cdot)} \end{bmatrix} \qquad (hk \times pk),$$

$$\mathbf{R} = \begin{bmatrix} \mathbf{\Theta}_0 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{\Theta}_1 & \mathbf{\Theta}_0 & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{\Theta}_{h-1} & \mathbf{\Theta}_{h-2} & \dots & \mathbf{\Theta}_0 \end{bmatrix} \qquad (hk \times hk).$$

Assume that the index set of the elements of X which are in $\mathcal{I}_{T,h}$ is κ. We then substitute the known values into (2.82) and extract these rows to form a system of equations as

$$\tilde{\mathbf{R}}\varepsilon = r \tag{2.83}$$

where $\tilde{\mathbf{R}} = \mathbf{R}_{(\kappa,\cdot)}$ (with a MATLAB style of matrix indexing), $r = X_{(\kappa)} - \mathcal{F}_{(\kappa,\cdot)}\tilde{X}_T$. With a little abuse of notation we denote the distribution of the random vector $(\varepsilon', (\tilde{\mathbf{R}}\varepsilon)')'$ as

$$\begin{bmatrix} \varepsilon \\ \tilde{\mathbf{R}}\varepsilon \end{bmatrix} = N\left[0, \begin{bmatrix} I & \tilde{\mathbf{R}}' \\ \tilde{\mathbf{R}} & \tilde{\mathbf{R}}\tilde{\mathbf{R}}' \end{bmatrix}\right]. \tag{2.84}$$

Clearly the covariance matrix of this normal distribution is singular as some elements of the joint normal random vector are a linear combination of the others. However, this does not prevent us from using the results of a conditional multi-variate normal distribution as studied in Alexander, Graybill, and Duane (1974) to obtain

$$F(\varepsilon|\tilde{\mathbf{R}}\varepsilon = r) = N\left[\tilde{\mathbf{R}}'(\tilde{\mathbf{R}}\tilde{\mathbf{R}}')^{-1}r, \mathbf{I} - \tilde{\mathbf{R}}'(\tilde{\mathbf{R}}\tilde{\mathbf{R}}')^{-1}\tilde{\mathbf{R}}\right]. \tag{2.85}$$

Then the conditional forecasting for all $1 \leq s \leq h$ is

$$\mathbb{E}[X|\mathcal{P}_T, \mathcal{I}_{T,h}] = \mathcal{F}\tilde{x}_T + \mathbf{R}\mathbb{E}[\varepsilon|\mathcal{P}_T, \mathcal{I}_{T,h}], \tag{2.86}$$

where the conditional expectation of ε is given in (2.85). The simulation can be done by sampling from (2.85) as well.

2.6 Interpolating Times Series with Linear Constraints

In this section we discuss an easy-to-implement, nonstandard solution to a practical problem one may encounter when applying the GVAR approach to a conditional scenario generation task. In modeling the joint distribution of portfolio risk factors and economic factors with the GVAR approach, one limitation inherited from the construction of the GVAR model is that both factors have to be modeled with the same frequency. As most of the economic forecasts are conducted with annual frequency, a problem may arise when the evolution of the risk factors over a more granular time interval is of interest. A piece of good news is that, if we model economic factors as endogenous variables and we have the information of these variables at certain future times but with frequency which is different from that of the model, we can still obtain the conditional expectation of the risk factors and conduct simulations with the help of the techniques introduced in Section 2.5. An even worse case, however, is that we only possess information about linear combinations of the future path for the economic factors. For example, we construct a GVAR model, which models the difference of economic factors with quarterly frequency but we can only obtain a one year forecast value at level for them. In this case, by converting the level forecast to difference data, the sum of four quarterly difference forecasts can be obtained. But this practice entails loss of low frequency information. Without further information, there is no way to determine the exact value of these variables and thus neither the conditional forecast nor simulation for portfolio risk factors can be done.

To solve the aforementioned problem, we need to interpolate (or to forecast) the time series for the economic factors over a number of future time intervals under certain linear constraints. Collect the relevant past data and future values to be interpolated in vectors Y_p and Y_f

respectively. Particularly, if we model the economic factors, Y_t, with a VAR model of order p, and we are to interpolate over h time periods, then

$$Y_p = \begin{bmatrix} Y_{T-p+1} \\ Y_{T-p+2} \\ \vdots \\ Y_T \end{bmatrix} \qquad (pm \times 1),$$

$$Y_f = \begin{bmatrix} Y_{T+1} \\ Y_{T+2} \\ \vdots \\ Y_{T+h} \end{bmatrix} \qquad (hm \times 1).$$

Let $f_{f|p}(Y_f|Y_p)$ be the joint distribution of Y_f conditioning on Y_p. Under the framework of maximum likelihood estimation, the interpolation problem can be formulated as determination of

$$\begin{aligned} \hat{Y}_f &= \arg\max_{Y_f} f_{f|p}(Y_f|Y_p) \\ &\text{subject to} \qquad \mathbf{A}Y_f = \mathbf{b}, \end{aligned} \tag{2.87}$$

where $\mathbf{A}$ and $\mathbf{b}$ are $c \times hm$ matrix and $c \times 1$ vector respectively, which specify the linear constraints on Y_f. The matrix $\mathbf{A}$ is expected to be of rank c such that all constraints are linearly independent.

Suppose that the random vector $(Y_p', Y_f')'$ has a multivariate normal distribution with mean $(\mathbb{E}(Y_p'), \mathbb{E}(Y_f'))'$ and a covariance matrix Σ, such that it can be divided conformably as

$$\Sigma = \begin{bmatrix} \Sigma_{pp} & \Sigma_{pf} \\ \Sigma_{pf} & \Sigma_{ff} \end{bmatrix}. \tag{2.88}$$

Then $f_{f|p}(Y_f|Y_p)$ is a multivariate normal density. In particular,

$$f_{f|p}(Y_f|Y_p) = \frac{1}{\sqrt{(2\pi)^{hm} \det \Sigma_{f|p}}} \exp\left\{ -\frac{1}{2}(Y_f - \mathbb{E}(Y_f|Y_p))' \Sigma_{f|p}^{-1} (Y_f - \mathbb{E}(Y_f|Y_p)) \right\}, \tag{2.89}$$

where $\Sigma_{f|p} = \Sigma_{ff} - \Sigma_{pf}\Sigma_{pp}^{-1}\Sigma_{pf}$. Without the constraints of (2.87), it is easy to see that (2.89) is maximized at $\hat{Y}_f = \mathbb{E}(Y_f|Y_p)$, which results in the optimal linear predictor of Y_f

given Y_p. It thus implies that the inclusion of the constraints leads to a loss of estimation efficiency to some extent, as the estimates obtained under constraints are neither unbiased nor consistent.

Further, suppose that Y_t follows a VAR(p) model

$$Y_t = \mathbf{\Phi}_1 Y_{t-1} + \cdots + \mathbf{\Phi}_p Y_{t-p} + \varepsilon_t, \tag{2.90}$$

where $p > 0$ and the residual, ε_t, follows $N(0, \Sigma_\varepsilon)$ and is a white noise process. Suppose that the observations of $Y_T, \ldots, Y_{T-p+1}$ are available. Then

$$\begin{aligned} f_{f|p}(Y_f|Y_p) &= \prod_{t=1}^{h} f(Y_{T+t}|Y_{T+t-1}, \ldots, Y_p) \\ &= \prod_{t=1}^{h} \frac{1}{\sqrt{(2\pi)^m \det \Sigma_\varepsilon}} \\ &\qquad \exp\left\{ -\frac{1}{2} \left(Y_{T+h} - \sum_{l=1}^{p} \mathbf{\Phi}_l Y_{T+h-l} \right)' \Sigma_\varepsilon^{-1} \left(Y_{T+h} - \sum_{l=1}^{p} \mathbf{\Phi}_l Y_{T+h-l} \right) \right\}. \end{aligned}$$

The log-likelihood function is thus given as

$$\begin{aligned} l(Y_f|Y_p) = &- \frac{hm}{2} \log(2\pi) - \frac{h}{2} \log(\det(\Sigma_\varepsilon)) \\ &- \frac{1}{2} \sum_{t=1}^{h} \left(Y_{T+h} - \sum_{l=1}^{p} \mathbf{\Phi}_l Y_{T+h-l} \right)' \Sigma_\varepsilon^{-1} \left(Y_{T+h} - \sum_{l=1}^{p} \mathbf{\Phi}_l Y_{T+h-l} \right). \end{aligned}$$

Ignoring all constants, the maximization of $l(Y_f|Y_p)$ is equivalent to minimization of the sum of squared errors, $S(X_f|X_p)$,

$$\begin{aligned} S(X_f|X_p) &= \sum_{t=1}^{h} \left(Y_{T+h} - \sum_{l=1}^{p} \mathbf{\Phi}_l Y_{T+h-l} \right)' \left(Y_{T+h} - \sum_{l=1}^{p} \mathbf{\Phi}_l Y_{T+h-l} \right) \\ &= \sum_{t=1}^{h} \left(\mathbf{C}_t Y_f - D_t Y_p \right)' \left(\mathbf{C}_t Y_f - D_t Y_p \right) \\ &= Y_f' \sum_{t=1}^{h} \mathbf{C}_t' \mathbf{C}_t Y_f - 2 \sum_{t=1}^{h} Y_p' \mathbf{D}_t' \mathbf{C}_t Y_f + \mathbf{E}, \end{aligned}$$

where

$$\mathbf{C}_t = \begin{bmatrix} -\mathbf{\Phi}_{t-1} & -\mathbf{\Phi}_{t-2} & \cdots & -\mathbf{\Phi}_{t-h} \end{bmatrix}, \tag{2.91}$$

$$\mathbf{D}_t = \begin{bmatrix} -\mathbf{\Phi}_{t+p-1} & -\mathbf{\Phi}_{t+p-2} & \cdots & -\mathbf{\Phi}_{t} \end{bmatrix}, \tag{2.92}$$

$\mathbf{\Phi}_0 = \mathbf{I}_m$, $\mathbf{\Phi}_l = \mathbf{0}$ for $l > p$ or $l < 0$. For a VAR(p) model, the interpolation problem posited in (2.87) can thus be solved as a quadratic programming problem

$$\begin{aligned} \underset{Y_f}{\text{minimize}} \quad & Y_f' \sum_{t=1}^{h} \mathbf{C}_t' \mathbf{C}_t Y_f - 2 \sum_{t=1}^{h} Y_p' \mathbf{D}_t' \mathbf{C}_t Y_f \\ \text{subject to} \quad & \mathbf{A} Y_f = \mathbf{b}. \end{aligned}$$

As stated previously, due to the linear constraints, we lose the efficiency of maximum likelihood estimation to some extent. Moreover, the coefficients of the VAR(p) model are not available to us in practice. Therefore, the coefficient matrices $\mathbf{\Phi}_l$ in $\mathbf{C}_t$ and $\mathbf{D}_t$ need to be replaced by the corresponding estimates, which introduces further uncertainty. As an intermediate step to overcome the limitation of the GVAR model, however, as shown by the numerical example given in Section 4.2.3, this approach gives reasonable interpolation results and is very easy to implement.

Chapter 3

Statistical Inference for Vector Autoregressive Model

In this chapter, we introduce statistical tests for causality and assumptions on the residuals (whiteness and normality) for the VAR model. The causality test is derived based on the characteristics of Granger causality for the VAR model and can be applied to determine whether a variable is to be modeled exogenously or endogenously. On the other hand, the whiteness and normality tests are based on the estimated residuals of the fitted VAR model. We note that these tests can be easily extended to apply to the GVAR model since the GVAR model can be seen as a restricted version of the VAR model.

The rest of this chapter is organized as follows. In Section 3.1, test statistics for Granger and instantaneous causality are derived, with a simple example given to illustrate the test procedure. Residuals based tests for whiteness and normality are discussed in Section 3.2 first for the VAR model, and then with the extensions to the GVAR model.

3.1 Testing for Causality

In modeling the economic factors as common variables, implicitly we assume that they are strongly exogenous. One way to test this assumption in the context of the VAR model is to examine the significance of Granger causality, as stated in Definition 2.4. The VARX* model also provides the flexibility of including the contemporaneous values of common variables.

This can be viewed as a reflection of the correlation between the residual processes of common variables and those of the domestic variables if normality is assumed for the distribution of the residual process. This is due to the effect of instantaneous causality which can be seen from (2.27) and Proposition 2.2. Thus, an alternative way to justify the inclusion of contemporaneous variables in the VARX* system, other than testing the significance of the estimated parameter, is to test for the instantaneous causality.

Both of the following causality tests are derived from the VAR(p) system as in (2.11), which collects risk factors and economic factors in a single vector Z_t. Let $K := k + m$. With some minor modification to the notation in (2.31) as follows:

$$\begin{aligned}
&\mathbf{X} := (z_1, \ldots, z_T) && (K \times T),\\
&\mathbf{B} := (\mathbf{\Psi}_1, \ldots, \mathbf{\Psi}_p) && (K \times Kp),\\
&Z_t := \begin{bmatrix} z_t \\ \vdots \\ z_{t-p+1} \end{bmatrix} && (Kp \times 1),\\
&\mathbf{Z} := (Z_0, \ldots, Z_{T-1}) && ((Kp) \times T),\\
&\mathbf{U} := (\varepsilon_1, \ldots, \varepsilon_T) && (K \times T),\\
&\mathbf{x} := \mathrm{vec}(X) && (KT \times 1),\\
&\beta := \mathrm{vec}(B) && (K^2 p \times 1),\\
&\mathbf{u} := \mathrm{vec}(\mathbf{U}) && (KT \times 1),
\end{aligned} \tag{3.1}$$

the OLS estimator of β is given as

$$\hat{\beta} = ((\mathbf{Z}\mathbf{Z}')^{-1}\mathbf{Z} \otimes \mathbf{I}_k)\mathbf{x}. \tag{3.2}$$

Then as stated in Proposition 3.1 of Lütkepohl (2005), we have

$$\sqrt{T}(\hat{\beta} - \beta) - \xrightarrow[d]{T\to\infty} N(0, \Sigma_\varepsilon \otimes \Gamma^{-1}), \tag{3.3}$$

where $\Gamma = \operatorname{plim} \mathbf{Z}\mathbf{Z}'/T$.

The asymptotic distribution for the estimator of Σ_ε is required for testing instantaneous causality. We restate Proposition 3.4 of Lütkepohl (2005) for the asymptotic distribution of the maximum likelihood (ML) estimator of σ.

Proposition 3.1 (Asymptotic Distribution of ML Estimator of σ)**.** *For a stationary, stable Gaussian VAR(p) process defined in* (2.11)*, its ML estimator* $\tilde{\sigma}$ *for* $\sigma := \text{vech}(\Sigma_\varepsilon)$ *is consistent and*

$$\sqrt{T}(\tilde{\sigma} - \sigma) \xrightarrow{d} N(0, 2\mathbb{D}_k^+(\Sigma_\varepsilon \otimes \Sigma_\varepsilon)\mathbb{D}_k^{+\prime}), \tag{3.4}$$

where $\mathbb{D}_k$ *is the* $(k^2 \times \frac{1}{2}k(k+1))$ *matrix such that*

$$\text{vec}(\Sigma_\varepsilon) = \mathbb{D}_k \,\text{vech}(\Sigma_\varepsilon).$$

and $\mathbb{D}_k^+ := (\mathbb{D}_k'\mathbb{D}_k)^{-1}\mathbb{D}_k'$*, which is the Moore-Penrose inverse of* $\mathbb{D}_k$*.*

3.1.1 Testing for Granger Causality

For a VAR(p) model defined in (2.11), we characterize the Granger-noncausality from X_t to Y_t in Proposition 2.1 as a test of zero constraints on some parameters in the coefficient matrices $\mathbf{\Psi}_l$. Moreover, it is established in Section 2.3.2 that the parameter estimator $\hat{\beta}$ follows an asymptotic normal distribution (see also B.2). Based on the aforementioned facts, we can construct test statistics to examine the validity of the zero constraints.

If the sub-matrix $\mathbf{\Psi}_{21,l}$ is of dimension $(m \times k)$ for all $l = 1, \ldots, p$, the total number of constrained parameters is $n_c = mkp$. Let $\mathbf{C}$ be a collection matrix which selects all the parameters out of the β and c be a $(n_c \times 1)$-dimensional zero vector. For example, if we are to conduct the zero restriction test,

$$H_0 : \psi_{12,1} = \psi_{13,1} = \psi_{12,2} = \psi_{13,2} = 0,$$

with respect to a VAR(2) model, where $\psi_{ij,l}$ are elements of the coefficient matrix $\mathbf{\Psi}_l$, then

$$\mathbf{C} = \begin{bmatrix} 0\,0\,0 & 0\,0\,0 & 1\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 1\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 1\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 1\,0\,0 \end{bmatrix},$$

since β as defined in (2.31) is a column vector collecting all parameters. Correspondingly, we have $c = (0, 0, 0, 0)$.

Even though $\mathbf{C}$ takes the form of a selection matrix and c is a zero vector for the purpose of testing Granger causality, the derivation of the test statistic can be done in a more general setting. For testing

$$H_0 : \mathbf{C}\beta = c \qquad \text{against} \qquad H_1 : \mathbf{C}\beta \neq c \tag{3.5}$$

with respect to an arbitrary $(n_c \times K^2p)$ matrix $\mathbf{C}$ of rank n_c and an $(n_c \times 1)$ vector c, Proposition 2.4 states that the OLS estimator of β has an asymptotic normal distribution, i.e.

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Sigma_\beta). \tag{3.6}$$

Then we have

$$\sqrt{T}(\mathbf{C}\hat{\beta} - \mathbf{C}\beta) \xrightarrow{d} N(0, \mathbf{C}\Sigma_\beta\mathbf{C}') \tag{3.7}$$

due to Proposition B.3. Hence

$$T(\mathbf{C}\hat{\beta} - c)'[\mathbf{C}\Sigma_\beta\mathbf{C}']^{-1}(\mathbf{C}\hat{\beta} - c) \xrightarrow{d} \chi^2(n_c). \tag{3.8}$$

Further replacing Σ_β with its estimator, $\hat{\Sigma}_\beta = \hat{\Gamma}^{-1} \otimes \hat{\Sigma}_u = T(\mathbf{Z}\mathbf{Z}')^{-1} \otimes \hat{\Sigma}_u$, which is consistent if Z_t satisfies the conditions in Proposition 2.4, we obtain our test statistic as

$$\lambda_W = (\mathbf{C}\hat{\beta} - c)'[\mathbf{C}((\mathbf{Z}\mathbf{Z}')^{-1} \otimes \hat{\Sigma}_u)\mathbf{C}']^{-1}(\mathbf{C}\hat{\beta} - c), \tag{3.9}$$

which is a *Wald statistic*. λ_W in 3.9 follows an asymptotic χ^2 distribution with n_c degrees of freedom. The result is summarized in the following Proposition, which is a restatement of Proposition 3.5 of Lütkepohl (2005).

Proposition 3.2 (Wald Statistic)**.** *Under the conditions in Proposition 2.4, suppose that* $\hat{\Sigma}_\beta = T(\mathbf{Z}\mathbf{Z}')^{-1} \otimes \hat{\Sigma}_u$ *is nonsingular and* $H_0 : \mathbf{C}\beta = c$ *is true, with* $\mathbf{C}$ *being an* $(n_c \times (k^2p+k))$ *matrix of rank* n_c *and* c *being an arbitrary* n_c*-dimensional vector. Then*

$$\lambda_W = (\mathbf{C}\hat{\beta} - c)'[\mathbf{C}((\mathbf{Z}\mathbf{Z}')^{-1} \otimes \hat{\Sigma}_u)\mathbf{C}']^{-1}(\mathbf{C}\hat{\beta} - c). \tag{3.10}$$

To compensate for the fact that Σ_β is unknown in practice, the Wald statistic is usually adjusted in the following way

$$\lambda_F = \lambda_W / n_c, \tag{3.11}$$

and compared to the critical value of an F-distribution. This adjustment is due to the fact that

$$n_c F(n_c, T) \xrightarrow[T\to\infty]{d} \chi^2(n_c) \tag{3.12}$$

Since an $F(n_c, T)$ distribution has a fatter tail than the $\chi^2(n_c)$ distribution, the adjustment takes into account the uncertainty of parameter estimation.

Another question to ask regarding the adjusted statistic is what degrees of freedom should be used for the numerator and denominator of this statistic, respectively. From the previous discussion, n_c qualifies as the numerator degrees of freedom. For the denominator, the idea is to use the number of observation minus the total number of parameters, as is applied to the residual degrees of freedom in the context of a regression. In the case of a VAR model, one can compute the degree of freedom with respect to either the entire VAR model, i.e. $kT - k^2p - k$, or each individual equation of the system, i.e. $T - kp - 1$. Either case should result in a similar outcome if the sample size is large enough.

3.1.2 Testing for Instantaneous Causality

Due to Proposition 2.2, the absence of instantaneous causality for VAR(p) in (2.11) is characterized as a zero constraint on some parameters in $\sigma := \text{vech}(\Sigma_\varepsilon)$. Thus we can derive the asymptotic test for instantaneous causality with the same rationale underlying the derivation of the Granger causality test.

Let n_c be the number of zero constraints and $\mathbf{C}$ denote a $(n_c \times (m+k)(m+k+1)/2)$ matrix of rank n_c, Then, for testing

$$H_0 : \mathbf{C}\sigma = 0 \qquad \text{against} \qquad H_1 : \mathbf{C}\sigma \neq 0, \tag{3.13}$$

again by using Proposition B.3, the test statistic is given as

$$\lambda = T\tilde{\sigma}'\mathbf{C}'[2\mathbf{C}\mathbb{D}_k^+(\Sigma_u \otimes \Sigma_u)\mathbb{D}_k^{+'}\mathbf{C}']\mathbf{C}\tilde{\sigma}, \tag{3.14}$$

where $\tilde{\sigma} = \text{vech}(\tilde{\Sigma}_\varepsilon)$ which is the vectorized ML estimator of Σ_ε and λ follows the $\chi^2(n_c)$ distribution asymptotically.

Note that the estimator we derived in Section 2.3.2 is the OLS estimator instead of the ML estimator. However, the form of the ML estimator coincides with that of (2.65), which is the same as the OLS estimator up to a constant and thus they are asymptotically equivalent. Also, thanks to the consistency of the OLS estimator, replacing all of the unknown quantities in (3.14) by its OLS estimator, we have the following test statistic under the OLS framework

$$\lambda_W = T\hat{\sigma}'\mathbf{C}'[2\mathbf{C}\mathbb{D}_k^+(\hat{\Sigma}_u \otimes \hat{\Sigma}_u)\mathbb{D}_k^{+'}\mathbf{C}']\mathbf{C}\hat{\sigma}, \tag{3.15}$$

which is asymptotically equivalent to λ.

3.2 Residual Based Test

3.2.1 Test for Whiteness of Residuals

Due to its simplicity in computation, the *portmanteau test* is one of the most popular tests for serial correlation of residuals. Let R_i denote the autocorrelation matrix of the residual process at lag i for $i = 0, 1, \ldots$. This test is designed to test the hypothesis

$$H_0 : \mathbf{R}_h = (R_1, \ldots, R_h) = 0 \qquad \text{against} \qquad H_1 : \mathbf{R}_h \neq 0. \tag{3.16}$$

The autocorrelation matrix is defined as a standardized autocovariance matrix

$$R_i = D^{-1}C_iD^{-1}, \tag{3.17}$$

where for $i = 0, 1, \ldots$, C_i is defined as

$$C_i = \mathbb{E}(u_t u'_{t-i}) \tag{3.18}$$

and D is a $(k \times k)$ diagonal matrix with the square roots of the diagonal elements of Σ_u on its main diagonal. Further, $\mathbf{C}_h := (C_1, \ldots, C_h)$.

In practice, the aforementioned quantities are all unknown and need to be estimated. The respective consistent estimators are given as

$$\hat{C}_i := \frac{1}{T}\sum_{t=i+1}^{T} \hat{u}_t\hat{u}'_{t-i}, \tag{3.19}$$

$$\hat{\mathbf{C}}_h := (\hat{C}_1, \ldots, \hat{C}_h), \tag{3.20}$$

$$\hat{R}_i := \hat{D}^{-1}\hat{C}_i\hat{D}^{-1}, \tag{3.21}$$

where $\hat{D}$ is a $(k \times k)$ diagonal matrix with the square roots of the diagonal elements of $\hat{C}_0$ on its main diagonal.

Following the illustration of Section 4.4.3 in Lütkepohl (2005), we present formulas of the test statistic and derive a simpler form of it,

$$\begin{aligned} Q_h &:= T \sum_{i=1}^{h} \operatorname{tr}\left(\hat{R}_i' \hat{R}_0^{-1} \hat{R}_i \hat{R}_0^{-1}\right) \\ &= T \sum_{i=1}^{h} \operatorname{tr}\left(\hat{R}_i' \hat{R}_0^{-1} \hat{R}_i \hat{R}_0^{-1} \hat{D}^{-1} \hat{D}\right) \\ &= T \sum_{i=1}^{h} \operatorname{tr}\left(\hat{D} \hat{R}_i' \hat{D} (\hat{D} \hat{R}_0 \hat{D})^{-1} \hat{D} \hat{R}_i \hat{D} (\hat{D} \hat{R}_0 \hat{D})^{-1} \hat{D}^{-1} \hat{D}\right) \\ &= T \sum_{i=1}^{h} \operatorname{tr}\left(\hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1}\right) \end{aligned} \tag{3.22}$$

We cite the Proposition 4.7 of Lütkepohl (2005) for the asymptotic distribution of Q_h.

Proposition 3.3 (Asymptotic Distribution of Portmanteau Statistic)**.** *Under the conditions in Proposition 2.4, we have, approximately, for large T and h,*

$$Q_h = T \sum_{i=1}^{h} \operatorname{tr}\left(\hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1}\right) \tag{3.23}$$

$$= T \operatorname{vec}(\hat{\mathbf{C}}_h)' (I_h \otimes \hat{C}_0^{-1} \otimes \hat{C}_0^{-1}) \operatorname{vec}(\hat{\mathbf{C}}_h) \approx \chi^2(k^2(h-p)). \tag{3.24}$$

The detailed proof for Proposition 3.3 can be found in Ahn (1988). Nevertheless, it is important to point out the asymptotic χ^2 distribution is only an approximation to the true asymptotic distribution. The approximation becomes exact when h tends to infinity. Thus, in practice we need to bear it in mind that even if the sample size is large enough, a small h can produce misleading test results. On the other hand, because in practice the sample size is always finite, we are interested in whether the asymptotic test is under or over rejecting the null hypothesis relative to its nominal size. Using Monte Carlo techniques, researchers found that with small samples, the test rejects the null hypothesis less often than indicated by its size (see e.g. Davies, Triggs, and Newbold, 1977; Ljung and Box, 1978; Hosking, 1980). To address this problem, it is proposed to use a modified test statistic of the form

$$\bar{Q}_h := T^2 \sum_{i=1}^{h} (T-i)^{-1} \operatorname{tr}(\hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1}), \tag{3.25}$$

which accounts for the number of terms in the summation of

$$\hat{C}_i = \frac{1}{T} \sum_{t=i+1}^{T} \hat{u}_t \hat{u}'_{t-i}. \tag{3.26}$$

Clearly, as $T \to \infty$, $T/[T^2(T-i)] \to 1$. Therefore $\bar{Q}_h$ is asymptotically equivalent to Q_h, which leads to the following approximate asymptotic distribution of $\bar{Q}_h$

$$\bar{Q}_h \approx \chi^2(K^2(h-p)). \tag{3.27}$$

When applying this test to a GVAR model, we need to keep in mind that GVAR is essentially a VAR model with linear constraints on the parameters. For example, when we assign different numbers of lags to each of the sub-sectional VARX*, we are in fact assuming that certain blocks of parameter matrices of the final GVAR model are zero. Moreover, because the coefficients that link the domestic variables of a VARX* and variables in other sub-sections are solved through estimated parameters of star variables and the predetermined weights, this can also be deemed as a restriction. As shown in Sections 4.4.2 and 4.4.3 of Lütkepohl (2005), the derivation of an approximate, asymptotic distribution of the Portmanteau statistic is based on an unrestricted estimator. In the case of a restricted VAR model, this test may be more aggressive in rejecting the null hypothesis since the restrictions reduce the number of parameters we can use to model the correlation.

In fact, this test can be adapted to the situations where we impose linear constraints on coefficients of a VAR model. In the context of a general VAR model as defined in (2.11), one way to formulate linear constraints is given as follows:

$$\beta = \mathbf{R}\gamma + r, \tag{3.28}$$

where β is defined in (3.1), $\mathbf{R}$ is a known $(K^2p \times M)$ matrix of rank M, γ is a $(M \times 1)$ vector of unconstrained parameters and r is a $(K^2p \times M)$ vector of known constants. This specification of linear constraints seems to be different from the more familiar form

$$\mathbf{C}\beta = c, \tag{3.29}$$

where $\mathbf{C}$ is a known $(N \times (K^2p))$ matrix of rank N and c is a known $(K^2p \times 1)$ vector. However, it is easy to show that (3.29) is equivalent to (3.28).

All of the implicit constraints that we impose on a GVAR model can be specified using (3.28). On the one hand, for a VARX* model that includes less lags, the constraint is seen as equating certain elements in β to zero. On the other hand, the estimated coefficients of star variables can be seen as elements in γ with the predetermined weights being elements of $\mathbf{R}$.

Under the above settings, we reproduce Proposition 5.8 of Lütkepohl (2005) for the approximated asymptotic distribution for the portmanteau statistic.

Proposition 3.4 (Asymptotic distribution of Portmanteau Statistic Under VAR with Linear Constraints)**.** *Suppose that x_t as defined in* (2.11) *is a stable, stationary, K-dimensional VAR(p) process with identically distributed standard white noise u_t and the parameter vector β satisfies the restrictions* (3.28)*. If the coefficients of y_t are subject to linear constraints specified as in* (3.28)*, then the portmanteau test statistic, Q_h has an approximate limiting χ^2-distribution with $K^2h - rank(\mathbf{R})$ degrees of freedom for large h.*

Proposition 3.4 states that, when applying a portmanteau test to restricted VAR, instead of subtracting K^2p from K^2h, we need to subtract the total number of unconstrained parameters, which equals the rank of matrix $\mathbf{R}$. For a GVAR model, it is easy to see that the unconstrained parameters are coefficients of the main variables and star variables in each of its underlying VARX* models. The coefficients of common variables, if included, are also unconstrained, and thus will also be subtracted in the determination of degrees of freedom, even though common variables are assumed to be strongly exogenous.

3.2.2 Test for Normality of Residuals

Unless a rich enough dataset is available to support the use of a nonparametric approach in the scenario generation exercise, a certain parametric model is required for the simulated scenarios to cover a sufficiently wide range of possible future outcomes, so as to ensure the adequacy of our risk analytics. Moreover, for the asymptotic analysis we derived previously to hold, the residuals process of the VAR model is required to be standard white noise. Except for the zero-mean and nonsingular covariance matrix requirements, confirmation of the standard white noise assumption is two fold, namely independence and existence of

moments up to fourth order. We apply the portmanteau test to determine whether there is significant autocorrelation for the estimated residuals in Section 3.2.1. In this section, we are to extend the Jarque-Bera Normality test for univariate random variables to multivariate random variables. If the null hypothesis for both tests can not be rejected, at least we have some degree of confidence that our GVAR model is a good representation of historical data.

The univariate Jarque-Bera test (see e.g. Jarque and Bera, 1987; Lomnicki, 1961) is constructed based on the respective estimators for the third and fourth moment of the random variable in question, i.e. for $x \sim N(0,1)$, we have $\mathbb{E}[x^3] = 0$ and $\mathbb{E}[x^4] = 3$. Further, let b_1 and b_2 be the estimator of third and fourth moment respectively, it can be shown that b_1 and $b_2 - 3$ follow a joint normal distribution and test statistic is then constructed accordingly. Following the derivation in Section 4.5 of Lütkepohl (2005), we first present the asymptotic result for a white noise process. We then reproduce Proposition 4.10 for the asymptotic results of estimated residuals of a VAR process.

Let u_t be a K-dimensional Gaussian white noise process such that $u_t \sim N(\mu_u, \Sigma_u)$. Let P be a matrix such that $PP' = \Sigma_u$, which can be obtained from a Cholesky decomposition. Then

$$w_t = (w_{1t}, \ldots, w_{Kt})' := P^{-1}(u_t - \mu_t) \sim N(0, \mathbf{I}_K), \tag{3.30}$$

which is a vector of independent standard normal variables. Hence,

$$\mathbb{E}\begin{bmatrix} w_{1t}^3 \\ \vdots \\ w_{Kt}^3 \end{bmatrix} = 0 \qquad \text{and} \qquad \mathbb{E}\begin{bmatrix} w_{1t}^4 \\ \vdots \\ w_{Kt}^4 \end{bmatrix} = \begin{bmatrix} 3 \\ \vdots \\ 3 \end{bmatrix} =: \mathbf{3}_K. \tag{3.31}$$

The same rationale is used to construct the normality test for a white noise process. Suppose that we have observations $u_1, \ldots, u_T$ and define

$$\bar{u} := \frac{1}{T}\sum_{t=1}^{T} u_t, \qquad S_u := \frac{1}{T-1}\sum_{t=1}^{T}(u_t - \bar{u})(u_t - \bar{u})', \tag{3.32}$$

and P_s is a matrix such that $P_s P_s' = S_u$. Then P_s is a consistent estimator of P. Further

define

$$v_t := (v_{1t}, \ldots, v_{Kt})' = P_s^{-1}(u_t - \bar{u}),\ t = 1, \ldots, T, \tag{3.33}$$

$$b_1 := (b_{11}, \ldots, b_{K1})' \text{ with } b_{k1} = \frac{1}{T}\sum_{t=1}^{T} v_{kt}^3,\ k = 1, \ldots, K, \tag{3.34}$$

$$b_2 = (b_{12}, \ldots, b_{K2})' \text{ with } b_{k2} = \frac{1}{T}\sum_{t=1}^{T} v_{kt}^4,\ k = 1, \ldots, K, \tag{3.35}$$

where b_1 and b_2 are the estimators of the moments in (3.31). With the notations defined above, the following proposition, which is a reproduction of Proposition 4.9 of Lütkepohl (2005), states the asymptotic distribution of these estimators.

Proposition 3.5 (Asymptotic Distribution for Third and Fourth Moment Estimators)**.** *If u_t is a Gaussian white noise such that $u_t \sim N(\mu_u, \Sigma_u)$, where Σ_u is a nonsingular matrix, then*

$$\sqrt{T}\begin{bmatrix} b_1 \\ b_2 - \mathbf{3}_K \end{bmatrix} \xrightarrow{d} N\left(0, \begin{bmatrix} 6\mathbf{I}_K & 0 \\ 0 & 24\mathbf{I}_K \end{bmatrix}\right). \tag{3.36}$$

With this proposition, the overall test statistic for the third moment can be constructed as

$$\lambda_1 := T b_1' b_1/6 \xrightarrow{d} \chi^2(K), \tag{3.37}$$

which is used to test

$$H_0 : \mathbb{E}\begin{bmatrix} w_{1t}^3 \\ \vdots \\ w_{Kt}^3 \end{bmatrix} = 0 \qquad \text{against} \qquad H_1 : \mathbb{E}\begin{bmatrix} w_{1t}^3 \\ \vdots \\ w_{Kt}^3 \end{bmatrix} \neq 0. \tag{3.38}$$

Similarly, for the fourth moment, the hypotheses are set up as

$$H_0 : \mathbb{E}\begin{bmatrix} w_{1t}^4 \\ \vdots \\ w_{Kt}^4 \end{bmatrix} = \mathbf{3}_K \qquad \text{against} \qquad H_1 : \mathbb{E}\begin{bmatrix} w_{1t}^4 \\ \vdots \\ w_{Kt}^4 \end{bmatrix} \neq \mathbf{3}_K, \tag{3.39}$$

which is tested using

$$\lambda_2 := T(b_2 - \mathbf{3}_K)'(b_2 - \mathbf{3}_K)/24 \xrightarrow{d} \chi^2(K). \tag{3.40}$$

One can also test both moments jointly using

$$\lambda := \lambda_1 + \lambda_2 \xrightarrow{d} \chi^2(2K), \tag{3.41}$$

for which the null hypothesis will be the combination of H_0 in (3.38) and (3.39).

As discussed in White and MacDonald (1980) and Jarque and Bera (1987), the test generally has poor power for a typical small sample size. Moreover, because these tests are based on estimators of the third and fourth moments, we can not expect them to have power to distinguish non-normal distributions that have the same third and fourth moments as those of normal distributions.

For a general K-dimensional VAR(p) process

$$X_t = \sum_{l=1}^{p} \mathbf{\Phi}_l X_{t-l} + u_t, \tag{3.42}$$

the estimated residuals are obtained by

$$\hat{u}_t := X_t - \sum_{l=1}^{p} \hat{\mathbf{\Phi}}_l X_{t-l}, \tag{3.43}$$

where $\hat{\mathbf{\Phi}}_l$ is the consistent and asymptotically normally distributed estimator of $\mathbf{\Phi}_l$. By using the estimated residuals, the covariance matrix of the innovation process is estimated as

$$\hat{\Sigma}_u := \frac{1}{T - Kp - 1}. \tag{3.44}$$

Let $\hat{P}$ be a matrix satisfying $\hat{P}\hat{P}' = \hat{\Sigma}_u$, so that it is a consistent estimator of P. Then we define

$$\hat{w}_t := (\hat{w}_{1t}, \ldots, \hat{w}_{Kt})' = \hat{P}^{-1}\hat{u}_t,\ t = 1, \ldots, T, \tag{3.45}$$

$$\hat{b}_1 := (\hat{b}_{11}, \ldots, \hat{b}_{K1})' \text{ with } \hat{b}_{k1} = \frac{1}{T}\sum_{t=1}^{T} \hat{w}_{kt}^3,\ k = 1, \ldots, K, \tag{3.46}$$

$$\hat{b}_2 = (\hat{b}_{12}, \ldots, \hat{b}_{K2})' \text{ with } \hat{b}_{k2} = \frac{1}{T}\sum_{t=1}^{T} \hat{w}_{kt}^4,\ k = 1, \ldots, K. \tag{3.47}$$

With the above notations, we restate Proposition 4.10 of Lütkepohl (2005) here.

Proposition 3.6 (Asymptotic Distribution of Residual Third and Fourth Moment Estimators)**.** *Let y_t be a stationary, stable K-dimensional VAR(p) process as in* (3.42) *with a*

Gaussian standard white noise. Then the estimators of residual third and fourth moments as defined in (3.46) *and* (3.47) *converge to a multivariate normal distribution, i.e.*

$$\sqrt{T}\begin{bmatrix}\hat{b}_1\\ \hat{b}_2-\mathbf{3}_K\end{bmatrix}\xrightarrow{d} N\left(0,\begin{bmatrix}6\mathbf{I}_K & 0\\ 0 & 24\mathbf{I}_K\end{bmatrix}\right). \tag{3.48}$$

Similar to (3.37), (3.40) and (3.41), the test statistic based on $\hat{b}_1$ and $\hat{b}_2$, i.e.

$$\hat{\lambda}_1 := T\hat{b}_1'\hat{b}_1/6 \xrightarrow{d} \chi^2(K), \tag{3.49}$$

$$\hat{\lambda}_2 := T(\hat{b}_2-\mathbf{3}_K)'(\hat{b}_2-\mathbf{3}_K)/24 \xrightarrow{d} \chi^2(K), \tag{3.50}$$

$$\hat{\lambda} := \hat{\lambda}_1+\hat{\lambda}_2 \xrightarrow{d} \chi^2(2K), \tag{3.51}$$

can be constructed to test the same hypothesis.

In the context of an estimated GVAR model, we note that the estimated covariance matrix $\hat{\Sigma}_u$ is obtained by stacking the estimated covariance matrices of all VARX* and then solving the structural VAR model. Empirically, the difference between the estimated covariance matrix obtained in such a way and that obtained based on estimated residuals is substantial. Following the standard framework of the GVAR model, we recommend that the stack-and-solved covariance matrix be used, so as to ensure the covariance matrix to be tested is consistent with the one used in simulation.

For an estimated GVAR model without exogenous variables, the test derived can be applied directly because the multivariate Jarque-Bera test does not require the VAR model to be unconstrained. On the other hand, if we are to include exogenous variables, the asymptotic distributions of $\hat{\lambda}$ should not be affected, because the inference can be conducted conditioning on fixed values of exogenous variables (see Section 2.2.3 and Engle, Hendry, and Richard (1983)). Thus we will apply the test to general GVAR models.

Chapter 4

Empirical Analysis

With the help of the GVAR Model, we are now able to construct a joint factor model which endogenously includes all market risk factors that directly impact a complex multi-asset portfolio. Moreover, to obtain a joint factor model to facilitate portfolio analytics conditioning on different perspectives of the future macroeconomic environment, a reasonable subset of economic factors will also be a part of the joint factor model. Note that for a simple portfolio, for which the joint factor model is only required to include a small number of factors, the traditional VAR model is ideal from the perspective of the complexity/efficiency trade-off. What essentially necessitates the use of the GVAR model is the presence of the "curse of dimensionality" problem, in the case where a joint model with dozens or even hundreds of factors is required. In order to illustrate the flexibility for inclusion of exogenous variables, we construct the joint risk and economic factor model as specified in (1.4). The specification is also examined with the causality tests constructed in Section 3.1. Note that, however, if we are to collect a large set of economic variables in practice, we may need to model them endogenously so as to enable parameter estimation for high-dimensional data. Regardless of whether Granger causality is supported by the data or not, this will not cause a problem since weak Granger causality simply results in models with parameters of lower dimension.

Once the joint factor model is estimated, we can then draw simulation of future scenarios. In the case that economic factors are treated exogenously, it is required that the complete paths of evolution for these economic variables are available so that we can conduct simulation with the estimated model. This, however, is not necessary if we instead treat economic variables

endogenously, in which case the techniques introduced in Section 2.5 are applied to enable the simulation. In either case, simulation or forecasting is not feasible when the only information regarding the future outcome of economic factors available are some linear combinations of these variables. In this case the approach introduced in Section 2.6 is available as a quick and easy solution to interpolate the future path of economic variables under these constraints, unless this information can be obtained in a more direct way.

Once we obtained sets of simulated market risk factors conditioned on different economic scenarios, the task for the joint factor model is accomplished. To illustrate the effectiveness of the simulated scenario, we further construct a virtual portfolio which is driven by the set of risk factors we selected. We value our portfolio on simulations of risk factors and compute expected returns and a risk metric (Value-at-Risk) for the returns of the portfolio to analyze the risks of our portfolio under different future macroeconomic conditions.

The rest of this chapter is organized in the following way. Section 4.1 describes the constituents of the virtual portfolio and how each type of asset is evaluated in this thesis. The mark-to-market value of the portfolio as of 31 December 2014 is also given in this section. In Section 4.2, we construct three economic scenarios which respectively reflect favorable, unfavorable and neutral views on future outcomes of macroeconomic conditions. The constituents of economic scenarios are determined based solely on an expert judgment. We also produce an unrealistically bullish economic scenario, which is the forecast based on a statistical model. Section 4.3 demonstrates the procedure with which we construct the joint factor model of the portfolio risk factors and the economic variables. Various tests are conducted on the estimated model to examine the validity of the model specification. We also give some illustration of simulated scenarios of risk factors in this section. Finally, in Section 4.4 we compare the expected performance under each of the economic scenarios by analyzing its expected returns and risk metrics.

4.1 The Portfolio

To demonstrate the ability of the GVAR model to cope with a large joint system of portfolio and risk factors, we construct a fairly complex multi-asset portfolio which consists of equities,

Table 4.1: High-level Summary of the virtual portfolio.

MtM($M)	USD	EUR	GBP	JPY	Total
Equity	46.6	2.1	56.1	0.1	104.9
Bond	51.2	68.9			120.1
Credit	3.3	1.6			4.9
Total	101.1	72.6	56.1	0.1	229.9

government bonds and credit default swaps (CDS) denominated in four different currencies: USD, EUR, GBP and JPY. The portfolio is evaluated in terms of USD. The high-level summary of the entire portfolio as of 31 December 2014 is given in Table 4.1.

4.1.1 Equity Portfolio

Past empirical research on stylized facts of stock prices states that, while serial autocorrelations are negligible for low frequency data, at higher frequency (usually intra-day time scale) their presence is significant (see Cont, 2001, for details). Moreover, Tanaka-Yamawaki (2011) states that cross-correlations between stock prices are present as multiple stocks in the same business sectors move coherently (although these correlations are never static through time). Therefore, it is reasonable to represent a panel time series of equity prices with a vector autoregressive model, although in general we do not expect significant coefficients of autocorrelation.

In our virtual portfolio, we hold a combination of well-diversified long positions in equities traded on USD, EUR, GBP and JPY stock exchanges. For EUR, GBP and JPY equity portfolios, we assume that they track the local market indexes closely enough such that their performance can be approximated with the corresponding market indexes. We model the USD equity portfolio in more detail by using the sector indexes, for example, Consumer Discretionary, Energy, Finance, etc., to represent the stocks in our portfolio. Index values are taken as proxies of sector portfolio prices. As of 31 December 2014, the entire equity portfolio is worth $105M, with 44% in USD, 53% in the GBP market and the rest in the EUR and JPY markets. The composition of USD equity portfolio is summarized in Figure

4.1.

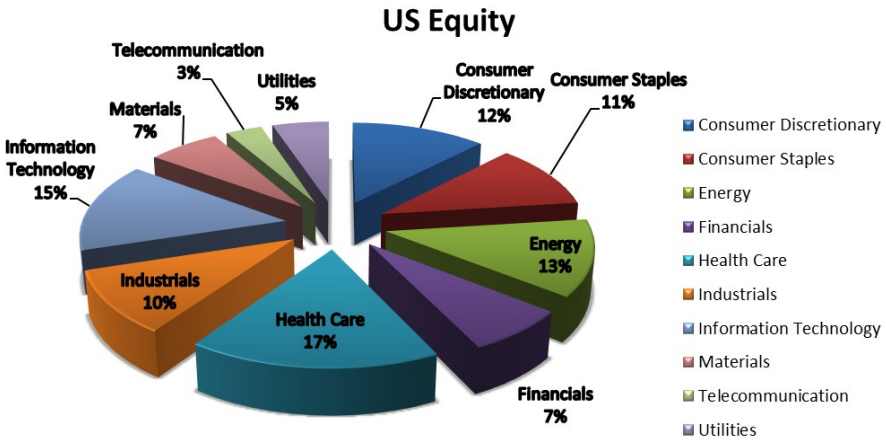


Figure 4.1: US Equity Portfolio Composition.

Before leaving the EQ portfolio, we want to point out that modeling equity portfolios using an index approximation is not necessary. In a recent application of the GVAR model (see e.g. Bussière, Chudik, and Sestieri, 2009), practitioners were able to model nearly a hundred variables endogenously. Thus for an EQ portfolio with a moderate number of constituents, it is possible to model using historical price data for each individual stocks.

4.1.2 Interest Rate Portfolio

Our interest rate portfolio consists of long positions in US and German government bonds with remaining time to maturity ranging from 6 months to 30 years. We assume that all our government bonds whose time to maturity is longer than 1 year will bear 2.5% coupons, which are payable semi-annually. Those expiring within a year are all zero-coupon bonds. Note that we are to project the portfolio one year forward. Thus, a reinvestment strategy may impact the actual portfolio performance since there will be cash inflows due to coupon payments and principal settlement. However, we simplify this part by assuming that the matured bond will be replaced with a new one with the same term to maturity and coupon income will be taken out of the portfolio. In this case, a small fraction of capital gain (or loss) and cash income is ignored, which should not cause a material impact.

Modeling interest rates under a real world measure has always been an extensive area of research. The term structure model that currently prevails in the industry relies on some dimension reduction techniques such as principle component analysis to enable inference for time series of differences or log differences of historical interest rate data. Rebonato et al. (2005) commented on this class of approaches and proposed a semi-parametric model to overcome some of its limitations. Unfortunately, although the simplicity in the implementation of this approach is very appealing, it can not be incorporated into the GVAR framework. Another class of term structure models approximates the panel time series of the term structure at each time point with a Nelson-Siegel (NS) curve and model the evolution of the three parameters of NS curves with a Vector Autoregressive model (see e.g. Diebold and Li, 2006; Xiang and Zhu, 2013). The three parameters can be easily incorporated into the GVAR system as GVAR is essentially VAR with certain parameter restrictions imposed. However, the implementation per se is extremely costly. In this study, we model the panel time series of differences (or logarithmic differences at the long end) of the term structure, in which 6 months, 1 year, 3 years, 5 years, 7 years and 30 years are included directly as part of the GVAR joint factor model, and enjoy the simplicity of implementation. This approach does not impose any restrictions on curvature of simulated term structures. However, the simulations seems reasonable in general. Another problem is modeling the term structure under the current regime of low interest rates. The global capital market post 2008 has gone into an era of low interest rates and this phenomenon seems to be persistent. Observing our data in Figure 4.2, which span the period of 1998 to 2014, we found that the logarithmic difference time series of the short end of the spot curve clearly presents the characteristic of regime switching. Using this data in model estimation and simulation leads to extremely unreasonable spot rate curves. This is mainly due the regime change we observed. Thus, we model the short end of the term structures at difference instead. The difference data is illustrated in Figure 4.3. This approach also has the benefit of allowing negative interest rates to be generated at the short end, which conforms with the current market observation.

The valuation of a bond with an arbitrary term to maturity requires the knowledge of the entire term structure. The model presented by Cairns (2004) allows for such flexibility.

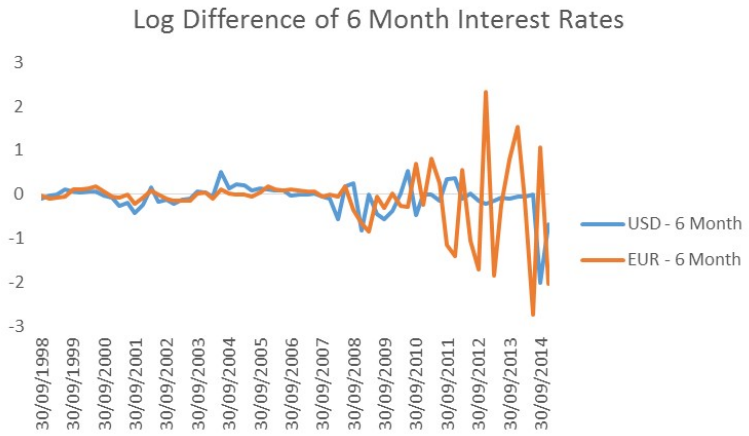


Figure 4.2: This figure shows the log difference of the time series for 6 month interest rates on USD and German government bonds. Data pre-2008 seem to be generated under a regime of low volatility while post-2008 data are much more volatile.

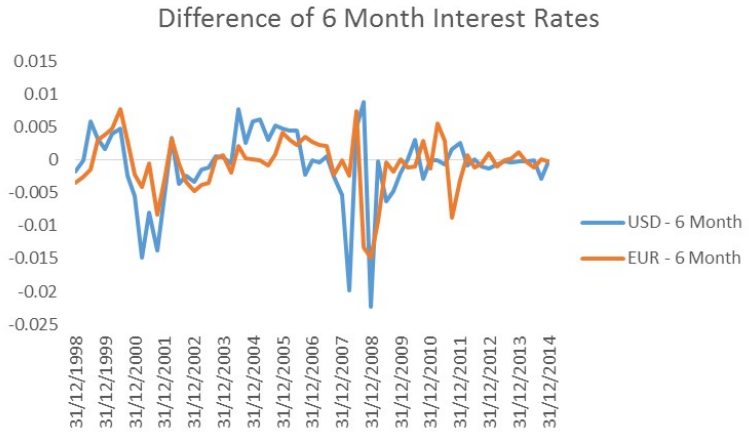


Figure 4.3: This figure displays the difference of the time series for 6 month interest rates on USD and German government bonds.

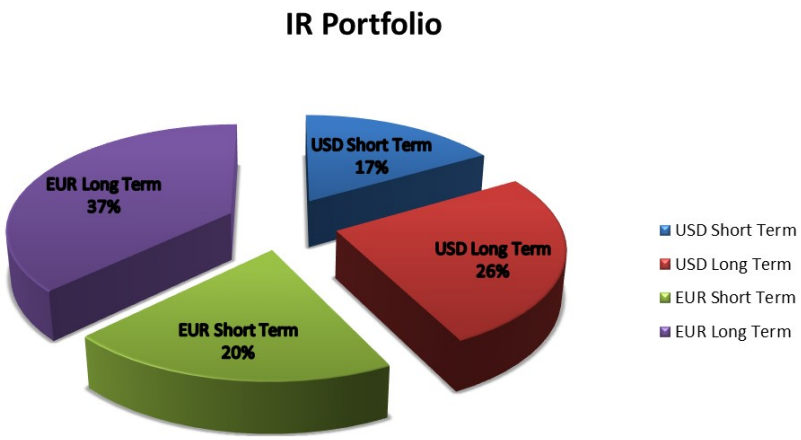


Figure 4.4: Interest Rate Portfolio Composition.

However, as a limitation of the approach we choose for this study, only values at reference points that we include in the model are known exactly. As an approximation, we interpolate between these points linearly. Such a treatment may leave some space for arbitrage but it is not a material problem in this application. Finally, as of 31 December 2014, the entire interest rate portfolio is worth $120M. Its composition is summarized in Figure 4.4.

4.1.3 Credit Portfolio

A small portion of the virtual portfolio is formed of long positions in credit derivatives such as credit default swaps (CDS) and CDS indexes (CDX.NA.IG and CDX.NA.HY). A CDS is a contract between two parties where an array of credit events with respect to a certain issuer, which is usually not involved in the contract, are specified. Until the trigger of any credit events, the buyer (long position) of the CDS pays a periodic fee specified in the contract (usually payable quarterly) to the seller (short position). Upon the triggering of a credit event, the seller has the obligation to pay the buyer the notional amount specified in the contract in return for the defaulted bond with the same amount of face value.

The accurate valuation of CDS is rather complicated. The industry standard is the JPMorgan CDSW calculator, available on the Bloomberg terminal. As the focus of this thesis is on the construction of the GVAR joint factor model instead of an asset valuation, we follow the CDS

valuation approach presented in Appendix I of Beinstein and Scott (2006) which is devised based on certain simplifying assumptions. Here we briefly describe this approach.

Assume that the credit curve is flat, namely the CDS spread for the same entity will be the same through all terms to maturity. In particular, we will use the current market spread of the CDS to infer the default probability, PD, for all future cash flows in valuation. Also assume that no accrual CDS fee is payable[1]. Then the risk neutral PD, between time (in integer years, from commencement of the contract) n and time $n+1$, given that the underlying entity has not defaulted until time x, is approximately current spread/(1 − recorvery rate), where the recovery rate is the proportion of the notional amount that can be recovered by selling the defaulted bond and is assumed to be 40%. Now, to interpolate between integer years, assume that the time until default between any adjacent integer years, given that the entity has not defaulted up to the beginning of this year, follows an exponential distribution, then the probability of surviving beyond any time x (not necessarily an integer) from commencement of the CDS contract is approximately $1/(1+PD)^x$. Then we will use this survival probability to adjust the net cash flows from unwinding the current CDS contract by entering an offsetting one. The present value of the risk-adjusted cash flows at the risk-free rate of interest will be the mark-to-market value of the current CDS. Here we use the term structure of interest rates on government bonds of the same denominated currency as the risk-free rates. The difference between a contract spread and a current market spread, multiplied by the notional amount of the contract, is the unadjusted cash flows from unwinding the contract.

We also take positions in CDS indexes such as investment grade (IG) and high yield (HY) Dow Jones CDX. CDX indexes give investors the opportunity to take exposure to a basket of credits of similar ratings. The CDX index is constructed in a way that is very similar to a single-name CDS contract, except that the underlying credit is a portfolio of reference entities with equal weights. Like a CDS contract, the CDX indexes have fixed maturities and the seller will receive periodic payments specified by a deal spread, with the current spread determined by market demand and supply. Since the indexes are designed to reflect performance of a portfolio of entities, default of any one entity does not lead to termination of the indexes but only a pro-rated reduction of notional. However, in the projection of our

[1]In practice, if the reference entity defaults between coupon dates, a pro-rated fee is payable.

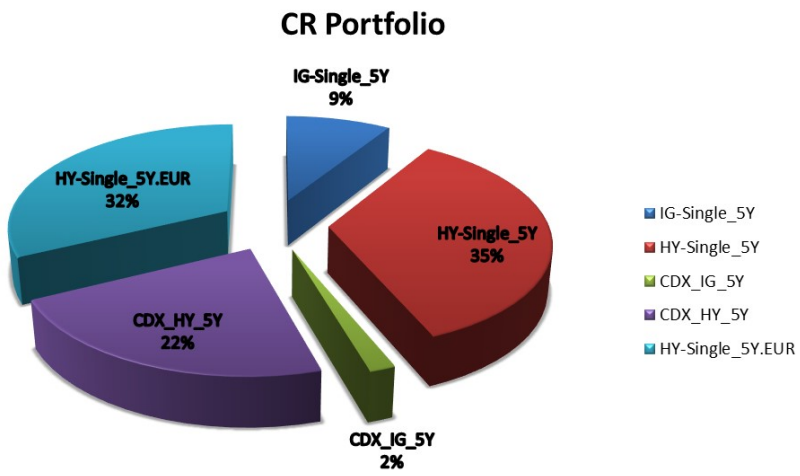


Figure 4.5: Credit Portfolio Composition.

portfolio, we do not quantify this partial default, as the impact of partial default tends to be immaterial over the first year of the life time of the CDX indexes.

In the virtual portfolio we hold only long positions in CDS and CDX contracts with 5 years to maturity. We assume that the deal spreads are either 1% or 0.5%. The composition of our credit (CR) portfolio, valued at $4.9M as of 31 December 2014, is illustrated in Figure 4.5.

4.1.4 Market Data

The sampling period for all data of market variables ranges from 30, June, 1998 to 31, December, 2014. Thus our quarterly time series contains 67 longitudinal data points. For each data point there are 13 variables for EQ, 12 variables for IR and 5 variables for CR.

We model the USD denominated EQ portfolio at a more granular level by representing the portfolio with ten S&P 500 sectors indexes, while the GBP, EUR and JPY EQ portfolios are approximated roughly by only one major stock index of the region. The sector indexes selected for our data set are S&P 500 Consumer Discretionary, S&P 500 Consumer Staples, S&P 500 Energy, S&P 500 Financials, S&P 500 Health Care, S&P 500 Industrials, S&P 500 Info Tech, S&P 500 Materials, S&P 500 Telecom Services and S&P 500 Utilities. All components of these indexes belong to the S&P 500 index and are classified according to the Global

Industry Classification Standard. The respective indexes selected to represent the GBP, EUR and JPY EQ portfolios are the FTSE 350, which includes a diversified portfolio of 351 stocks listed on the London Stock Exchange; MSCI EMU, which consists of 240 constituents and covers approximately 85% of the free float-adjusted market capitalization of the European Economic and Monetary Union; and TOPIX 500, which measures the performance of the 500 most liquid stocks with the largest market capitalization that are members of the TOPIX index. All index data are obtained from the Bloomberg terminal.

For the IR portfolios, we model the evolution of the spot rate curves of US and German government bonds. The data used are the zero coupon bond yields up to 30 years of term to maturity (values for longer terms are bootstrapped from coupon bond prices), which are directly available from the Bloomberg terminal. The compounding frequency is annual. Among all available terms, we selected the 3M, 6M, 1Y, 3Y, 5Y, 7Y, 10Y and 30Y as reference points in our term structure model. Note that 3M and 10Y terms are assumed to be the proxy for short-term and long-term policy rates of each sovereign respectively and thus are included in the economic data set instead.

For the CDX in the CR portfolio, we use the time series of theoretical spreads on 5-year CDX.NA.HY and CDX.NA.IG contracts at issuance as proxies. The CDX data are downloaded from the Bloomberg terminal. As stated in Section 4.1.3, the CDS spread varies with time to maturity, even for the same entity. However, each series of CDX has a different basket of credits. Therefore forming a term structure of credit spreads with data from different series is not theoretically sound either. Thus we rely on the assumption of a flat spread curve and model the time series of a single point of the spread curve for each CDX contract. For modeling of the single name CDS contracts, we use the BofA Merrill Lynch US Corporate Master Option-Adjusted Spread, which is a capitalization-weighted average of option-adjusted spreads on corporate bonds with issuers rated as investment grade, as the proxy of a USD denominated investment grade single name CDS spread. BofA Merrill Lynch US High Yield Option-Adjusted Spread, which is the counterpart of the previous index for below-investment grade rated entities, is used as the proxy for a USD denominated high yield single name CDS spread. BofA Merrill Lynch Euro High Yield Index Option-Adjusted Spread is used as the proxy for a EUR denominated high yield single name CDS spread.

We note that the corporate bond spread is generally different from the CDS spread. However, these are the best existing proxies for the single name CDS spread. The corporate bond spread indexes data are constructed by BofA Merrill Lynch and retrieved from FRED, Federal Reserve Bank of St. Louis.

4.2 The Economic Scenarios

Although the set of risk factors can be determined based on the nature of the assets in the portfolio, the economic factors conditioned on which the scenarios are generated have to be chosen with experience and expert judgment. The data availability and modeling requirements are also considerations that need to be taken into account in the selection of the economic factors.

4.2.1 Constructing Economic Scenarios

In this study we generate risk factor scenarios conditioned on point economic scenarios only, namely each economic scenario is a vector of economic variables with fixed values. Referring to the S&P economic research, Credit Week, April 22nd 2015, we construct three economic scenarios expressing favorable, best-estimate and unfavorable views on future macroeconomic conditions respectively. The economic factors selected and the construction of the scenarios can be viewed in Table 4.2.

We select a manageable set of 18 economic variables, which consists of macroeconomic indicators for the regions where we invested in, such as real GDP growth, inflation (CPI growth) and interest rates. Due to our focus on modeling of the USD assets, it is mainly the variation of US economic variables that forms different scenarios. In general, economists expect the global economy to continue to expand in 2015, at a slightly higher pace than 2014. In the UP and BASE scenarios we anticipate to see a growth of 3.4% and 3% respectively, both higher than the 2.4% growth in 2014. Moreover, EU, GB and JP are all expected to expand faster than they did in 2014, when the real growths were 1.2%, 2.7% and -0.1%. On the monetary policy side, EU is expected to continue its quantitative easing program and maintain a level as low as 0.1% for the short-term (3-month) and 0.3% for the long-term

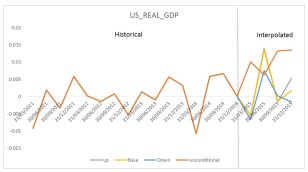
Table 4.2: The three examples of economic scenarios. The FX's stand for exchange rate converted to USD.

	UP	BASE	DOWN	Unconstrained
GDP Growth (GB)	2.8%	2.8%	2.8%	8.7%
GDP Growth (US)	3.4%	3%	2.1%	6.7%
GDP Growth (EU)	1.5%	1.5%	1.5%	5.9%
GDP Growth (JP)	0.8%	0.8%	0.8%	4.9%
Inflation (GB)	0.1%	0.1%	0.1%	-0.6 %
Inflation (US)	-0.3%	-0.3%	-0.3%	2.9%
Inflation (EU)	-0.3%	-0.3%	-0.3%	1.2%
Inflation (JP)	0.4%	0.4%	0.4%	-0.2%
Unemployment (US)	5.4%	5.4%	5.8%	4.0%
IR Short (USD)	0.4%	0.3%	0%	0.4%
IR Short (EUR)	0.1%	0.1%	0.1%	0.8%
IR Long (USD)	2.4%	2.3%	2.1%	3.4%
IR Long (EUR)	0.3%	0.3%	0.3%	1.0%
FX (GBP)	1.428	1.428	1.428	1.941
FX (EUR)	1.111	1.111	1.111	1.287
FX (JPY)	0.008	0.008	0.008	0.007
Oil Price	50.12	50	48.56	87.12
S&P 500 Return	10%	7.6%	8.5%	43.0%

(10-year) interest rates. In the favorable scenario, the Federal Reserve will start to raise the policy rates. However, there will still be a significant degree of uncertainty in the process of normalization of the monetary policy. Thus we expect policy rates to stay at the 2014 level in the BASE scenario and drop if the economic performance turns out to be unfavorable. A counterintuitive point may be the S&P 500 return in the DOWN scenario, which is even higher than that in the BASE scenario. This is because we believe a continued quantitative easing in the DOWN scenario may result in a capital injection into the equity market.

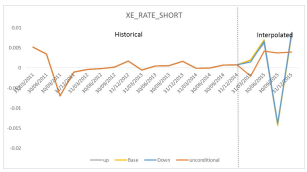
4.2.2 Economic Data

The economic data spans the same period as that of the market data and contains time series for all of the variables listed in Table 4.2. The quarterly real GDP data for all countries are downloaded from the S&P Capital IQ platform, and cross-validated with Bloomberg data. The GDP growth figures are computed as a first-difference of logarithmic GDP to reduce over-dispersion. The inflation data for all countries are derived from the their CPI times series, which are obtained from the same source, as the logarithmic first-difference, which represents an approximate growth rate of the variable. The unemployment rates are reported on a monthly basis on the S&P Capital IQ platform. We take a simple average of the data of the three months in the same quarter to convert monthly data to quarterly frequency. As discussed in Section 4.1.4, the 3M and 1Y spot rates on the sovereign bonds are taken as the short- and long-term rates respectively. The exchange rates are close prices in terms of USD per unit of the currency, which are recored on the last day of each quarter and obtained from the S&P Capital IQ platform. The time series for oil prices are the Crude Oil Prices: West Texas Intermediate (WTI - Cushing), published by the US. Energy Information Administration and retrieved from FRED, Federal Reserve Bank of St. Louis. The oil prices are reported on a daily basis and the missing values are replaced by the price on the previous day. The values on the last day of the quarter are extracted to form the quarterly time series. Finally, the S&P 500 index is selected as the indicator for the US equity market. The time series data of the S&P 500 index are obtained from the Bloomberg terminal.

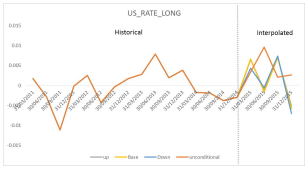


(a) US Real GDP

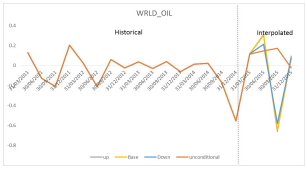
(b) GB Real GDP



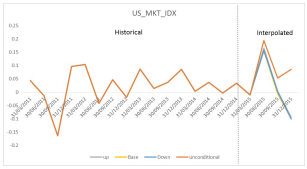
(c) GE 3-month policy rate



(d) US 10-year policy rate



(e) Oil Price



(f) S&P 500 Index

Figure 4.6: Interpolating time series of economic variables over 2015 conditioning on the scenarios given in Table 4.2.

4.2.3 Interpolating Economic Scenarios

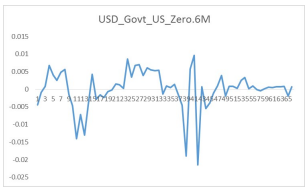
Note that the data in Table 4.2 describe the evolution of economic variables that spans a one-year period, while in our joint factor model, the economic variables evolve on a quarterly basis. The interpolation technique is applied to translate the annual data to a quarterly frequency. We model logarithmic first-differences of the 18 economic variables, except for the unemployment rate and interest rates, which are modeled at first difference. We examine the quarterly time series, which span the period from Q2 1998 to Q4 2014, of the economic variables that we use to estimate the joint factor model by conducting an augmented Dickey-Fuller test on the differenced data for stationarity. The test results are presented in Appendix C.1. We found that after the first-difference/logarithmic first-difference treatment, the null hypothesis of a unit root can be rejected in most series. As a result, we continue with the treatments tabulated in Table C.1.

We first fit a VAR model to the first-differenced panel time series of the economic variables and then apply the technique introduced in Section 2.6 to interpolate the time series under the fitted model, conditioning on the scenarios given in Table 4.2. We compare the interpolation results given by the VAR models with 1, 2 and 3 lags respectively and determine to proceed with the VAR(2) model. We only present the interpolated time series for some variables due to the space limit. Note that the magnitude of the y-axis represents values at first difference.

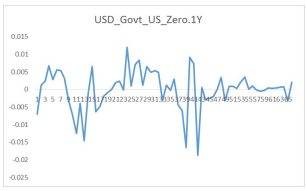
As demonstrated in Figure 4.6, the interpolations under different scenarios generally present a similar cyclical pattern as their historical paths and seem to be reasonable. In Figure 4.6b, we note that, although the forecasts of the UK real GDP growth over 2015 do not vary across scenarios, there are still variations between paths under different economic scenarios due to their correlation with other variables. Such interdependence causes some unexpected interpolation results. For example, the UK real GDP growth is the highest under DOWN scenario among the three towards the end of the interpolation period. Although the total growth during 2015 is still 2.8%, it is possible that the market is projected to be less bullish due to such interpolations. A similar observation is made in Figure 4.6e. As an approximation, interpolation introduces model mis-specification. Since the conditional forecasts of portfolio risk factors are very sensitive to the input economic scenarios, a better solution is to obtain reliable economic forecasts of the same frequency as the model. Finally, note that the unconstrained forecasts of the economic variables, which are generated directly from the fitted VAR(2) model, are much more bullish than the constrained ones. The unrealistic bullishness is expected as a result of forecasting the economy with purely statistical tools. When conditioning on this scenario to simulate portfolio risk factors, we expect this bullishness to have a major impact on the performance of the portfolio.

4.3 The Joint Factor Model

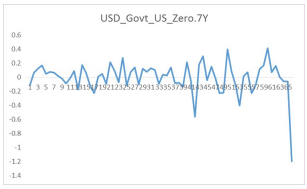
In this section, we construct a joint factor model with the GVAR approach using the quarterly panel time series that spans the period of Q2 1998 to Q4 2014. We model the first-difference or logarithmic first-difference of the panel time series and assume that all of the original time series are of order one ($I(1)$ variables) so that we are not worried about the cointegration relations in the GVAR model. The stationarity assumption is confirmed by using the ADF test in a similar manner as Section 4.2.3. The test results are presented in Appendix C.2. Note that there are three times series that fail to reject the null hypothesis of a unit root even after being first differenced. By observing their plots in Figure 4.7, we reckon that this may result from switches in variance, rather than non-stationary in mean. Therefore, we proceed with the first-difference treatment tabulated in Table C.2 and Table C.3. As mentioned in



(a) USD IR - 6 Month



(b) USD IR - 1 Year



(c) USD IR - 7 Year

Figure 4.7: Plots for first-differenced time series of 6 month, 1 year and 7 year spot rates on US government bonds.

Section 4.1, we include equity indexes (13), CDS/CDX spread (5), and reference points on the spot rate curves of US and German government bonds (12) as endogenous variables of the GVAR model and the 18 economic factors listed in Table 4.2 as common factors.

4.3.1 Sub-sector Model Specification

The first step in constructing a GVAR model is specifying and estimating the VARX* model for each sub-sector. In previous empirical applications of the GVAR approach in global macroeconomic modeling, such a sectioning is intuitive. Usually econometricians assume that all of the economic variables form a large endogenous system and each country acts as a sub-sector. Naturally, the weighted averages of the economic variables of the same type, e.g. GDP, export, import, from all countries, are defined as the star variables.

In this study, we attempt to construct a joint factor model, where the portfolio risk factors are determined endogenously while the economic variables are exogenous to the system. Note that the extent of a parameter shrinkage by using a GVAR approach is related to how granular the sectioning is. The larger the sub-sector compared to the entire system, the less sufficient the parameter shrinkage is. Since our portfolio consists of assets denominated in four currencies, using a currency as a sub-sector rendering a small number of sub-sectors and thus an insufficient shrinkage in the parameter space. The same problem exists when sectioning variables by asset type. To maximize the parameter shrinkage, we make each portfolio risk factor a sector by itself. Then we take the average of variables by asset type as the star variable. Note that the star variables are sector-specific. On the one hand, it is a common practice that the domestic variables of the VARX* model for each sub-sector

are not included in the computation of the corresponding star variables. On the other hand, weights can vary across different sub-sectors. As for our case in particular, the star variables are averages of US equity indexes, world equity indexes, credit derivatives spreads, spot rates of US government bonds and of German government bonds. A financial justification of this is that one would expect the star variables to act as proxies of unobservable common factors. In this case, our star variables represent aggregate movements in different markets.

The choice of weights that one should employ in constructing star variables still remains an open question in the empirical literature. In modeling global trade, possibilities include trade shares and GDP weights, which are intuitive. However, as stated in Section 2.1, the weights are of secondary importance as long as they meet the granularity conditions. Therefore, we assign the same weight to every variable in the calculation of star variables to approach granularity conditions.

Finally, we test for Granger causality from the domestic variable of each VARX* model to the common factors (economic variables) and instantaneous causality. As suggested by the test results tabulated in Appendix C.3, both Granger and instantaneous causality exist between portfolio risk factors and economic factors. Interestingly, however, we observe that, when testing in the bivariate systems formed by a pair of a portfolio risk factor and an economic variable, the Granger causality test becomes insignificant. The test results for instantaneous causality justify the inclusion of the contemporaneous value of the economic variables. Despite the test results for Granger causality, we proceed with the choice of modeling the economic variables exogenously.

4.3.2 Fitting VARX*

In other empirical applications of the GVAR model, significant effort is devoted to a specification of cointegration relations, if the forcing variables are modeled as $I(1)$. In the presence of $I(1)$ variables, we can fit the VARX* through its error-correction form (see Chudik and Pesaran (2014) for details). In our application, we focus on modeling the short-term dynamics of portfolio risk factors and include only stationary time series. Thus, for each VARX* specified in Section 4.3.1, we estimate the parameters in formula (2.5) with mean-adjusted data. Note that the VARX* model allows for the flexibility of different lags for domestic

variables, star variables and common factors. The model selection results with both AIC and BIC are tabulated in Appendix C.4. We take lags suggested by BIC to avoid an over fitting problem and retain higher variability in residual terms. The VARX* models with BIC lags are then estimated using the OLS techniques introduced in Section 2.3.

Once the sub-sector models are established, we then stack and solve them to obtain the final GVAR model, following the steps in Section 2.4. In our application, the system is solvable. Therefore we save the effort of appending the system with equations on star variables to make the system uniquely determined. Note that the solved system of VARX* models essentially forms a VAR model with linear restrictions on the parameters. All of the residual-based tests for the restricted VAR models are therefore applicable. Using the solved GVAR model, we estimate the residuals and apply the tests for whiteness and normality introduced in Section 3.2. Not surprisingly, both tests reject the null hypothesis with a high level of confidence (over 99.9%). If we are only concerned with the mean scenario of portfolio risk factors under each conditioned economic scenario, the assumptions regarding residuals are unimportant. More effort can be invested into a careful modeling of residuals if we are interested in obtaining a full conditional distribution under each economic scenario. The possible solutions include a vector moving average process, parametric distributions other than the assumed normal distribution or even a non-parametric approach. For this study, we will proceed with the assumption of a normal white noise.

4.3.3 Out of Sample Forecast

We leave out the data over the period of Q1 2013 to Q4 2014 and fit a GVAR model with the above specification to the truncated data. We then project the portfolio risk factors by 8 periods forward, conditioning on the realized path of the economic variables and compare the forecasts with the actual evolution of the markets to see how well the GVAR model under our specification can capture the market variation.

In Figure 4.8, we present the out of sample forecast for some portfolio risk factors. The upper and lower bound present a 99% confidence interval. Note that due to our normality assumption, the complete distribution of the forecast can be obtained analytically. Nevertheless, if a more sophisticated approach is applied to model the residuals in practice, it is still possible

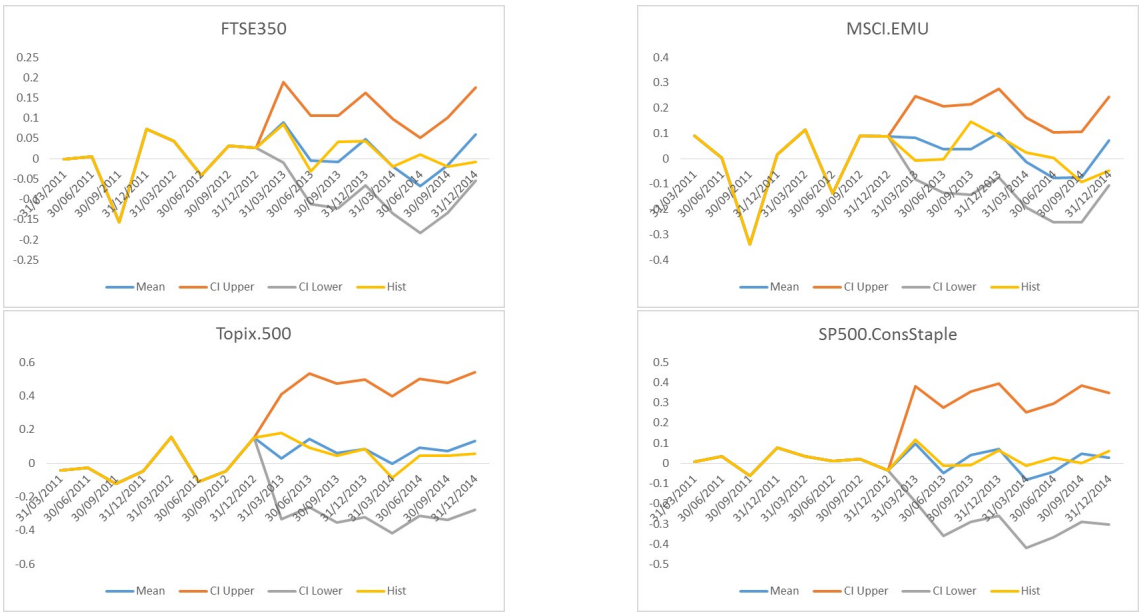


Figure 4.8: Projected portfolio risk factors 2 years ahead out of sample, conditioning on the actual evolution of the economic variables over the projection period.

to approximate the distribution with a Monte-Carlo method. Observing the plots of the forecasts, we find that for most of the portfolio risk factors, the actual path is covered by the 99% interval. This is good news for the GVAR approach in applications to risk management since the model contains the actual outcome with its 99.5% VaR, which is a confidence level frequently adopted by financial institutions. Moreover, some marginal processes even follow the actual paths quite closely, testifying to the effectiveness of the GVAR approach.

In Figure 4.9, we further observe the ability of the GVAR model in modeling the dependence structure of short-term dynamics in the joint system formed by the risk factors and economic variables. It is easy to see that the GVAR model successfully captures the linear correlation within the system, especially for credit and interest rate sectors, where the pattern of dependence has been stable over the sample period. It seems that the dependence structure in the equity sector is much more difficult to capture with a linear model. However, we would attribute this observation to a characteristic of the equity market, rather than to the model limitation.

In sum, the GVAR model has the abilities to capture both the trend and variability in the marginal portfolio risk factor processes and to model the linear dependence between variables

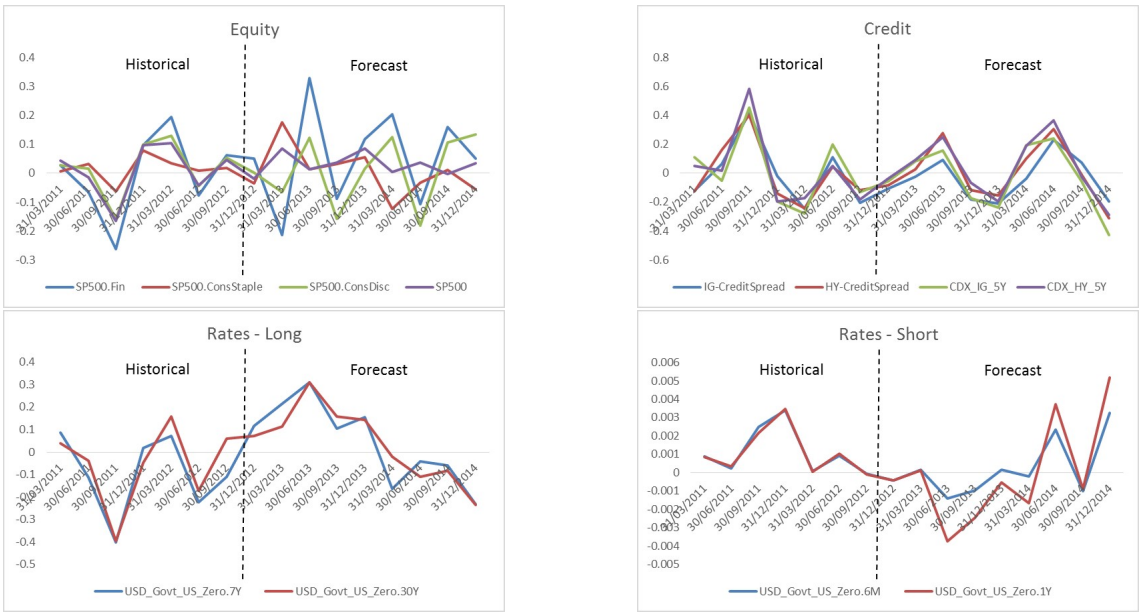


Figure 4.9: The mean path of some portfolio risk factors.

under the given economic forecasts, provided that the system is indeed stable over the sample period.

4.4 Portfolio Analytics

In this section, we simulate our virtual portfolio over scenarios of risk factors generated conditionally with the GVAR model. The model with the specification stated in Section 4.3.1 is estimated with the entire panel time series that spans the period from 1998 Q2 to 2014 Q4. We then simulate the short term dynamics of the joint factor system four periods ahead conditioning on the interpolated economic scenarios described in Section 4.2. For each simulation, the dynamics are added to the Q4 2014 data to obtain the simulations of the Q4 2015 risk factors, which forms the basis for valuation of the virtual portfolio. When the calculation is completed, we will possess the empirical full conditional distribution for the P&L of each asset, from which a huge amount of information can be drawn. Alternatively, we note that the current model specification is equivalent to a joint distribution with a normal marginal distribution for the short-end of the interest rate term structure and log-normal marginal distributions for other variables. Therefore, the full conditional distribution can be obtained

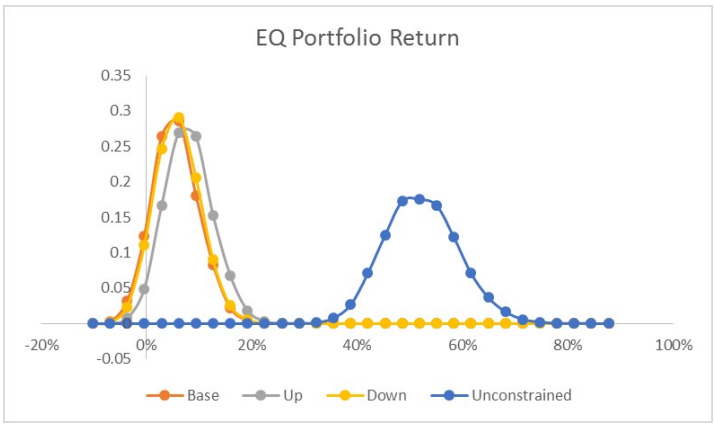


Figure 4.10: Full Conditional Distribution of Equity Portfolio.

Table 4.3: Descriptive Statistics for Return on EQ Portfolio under Each Economic Scenario.

	UP	BASE	DOWN	Unconstrained
Min	-7.6%	-11.1%	-11.4%	29.3%
Max	26.1%	24.7%	23.7%	79.8%
Mean	6.4%	3.9%	4.3%	50.6%
Stdev	4.5%	4.3%	4.3%	7.0%
Skewness	0.2518	0.2138	0.2316	0.2383
Kurtosis[1]	0.1020	0.0609	0.0615	0.0084
99.5% VaR	-4.2%	-6.4%	-5.9%	34.4%

[1]The kurtosis is sample excess kurtosis.

analytically, at least for equities. However, as we do not implement heteroskedasticity and regime-switching features, the variability of risk factors remains the same for all economic scenarios. We focus on analyzing the mean returns under the first three scenarios, since they are conditioned on carefully constructed economic scenarios and are expected to be more meaningful. The simulation results for the Unconstrained scenario are given mainly for the purpose of demonstrating the sensitivity of simulated risk factors to the input economic scenario. The 99.5% VaR of returns will also be tabulated for reference.

The full conditional distributions of the entire equity portfolio under each economic scenario is illustrated in Figure 4.10, with some statistics tabulated in Table 4.3.

As expected, the distributions are right skewed with the standard deviations increasing slightly with the means. Among the three postulated economic scenarios, the UP scenario produces the highest expected return. Although a slightly higher GDP is anticipated in the BASE scenario, the S&P 500 index is much higher in the DOWN scenario, which brings the expected return to a level above that of the BASE scenario. Due to the weak GBP and EUR, both of which are expected to depreciate by over 8% during 2015, the overall return falls behind the S&P 500 index. Under the unrealistically bullish scenario, the portfolio is expected to grow by 50.6% and produce huge profit even in the worst 99.5% case.

A careful look at the sub-portfolio level provides more information. Here we only tabulate in Table 4.4 the mean and 99.5% VaR of the return, as these are sufficient to represent all the information we can draw from the model.

Table 4.4: Expected Return at Sub-portfolio Level

	UP		BASE		DOWN		Unconstrained	
	Return	99.5%VaR	Return	99.5% VaR	Return	99.5% VaR	Return	99.5% VaR
EQ_USD	11.0%	2.5%	8.3%	-0.3%	8.8%	0.6%	42.2%	31.8%
EQ_GBP	12.3%	-8.9%	9.7%	-11.1%	10.0%	-10.7%	71.2%	38.7%
EQ_EUR	6.3%	-20.9%	3.4%	-23.6%	5.4%	-21.9%	86.8%	38.9%
EQ_JPY	39.0%	-18.7%	36.1%	-20.7%	32.0%	-24.2%	41.7%	-16.6%
EQ_ConsDisc	8.1%	-21.3%	4.5%	-24.8%	2.9%	-25.8%	52.8%	9.7%
EQ_ConsStaple	-1.8%	-34.4%	-4.8%	-36.9%	-7.6%	-37.9%	13.6%	-24.1%
EQ_Energy	8.9%	-13.9%	7.4%	-15.0%	8.2%	-13.7%	43.7%	12.5%
EQ_Fin	-0.9%	-34.4%	-3.5%	-37.3%	-2.3%	-36.4%	90.9%	26.4%
EQ_Health	6.1%	-26.4%	3.7%	-27.3%	2.4%	-27.9%	25.3%	-12.1%
EQ_Indust	29.2%	-2.3%	24.7%	-5.1%	24.8%	-5.3%	60.2%	21.5%
EQ_InfoTech	26.7%	-17.5%	22.9%	-19.8%	28.4%	-16.5%	39.0%	-9.0%
EQ_Mat	11.5%	-44.9%	10.1%	-45.5%	8.6%	-45.8%	74.4%	-13.1%
EQ_Telco	2.7%	-36.0%	1.7%	-37.2%	3.4%	-34.8%	22.8%	-24.5%
EQ_Utils	4.1%	-34.2%	4.1%	-34.5%	9.2%	-32.0%	5.3%	-34.3%

The returns in Table 4.4 are computed in the denominated currency. Ignoring the effect of depreciating GBP and EUR, we can conclude that the performances of the GBP and EUR EQ portfolios are not too unsatisfactory. Their variations across different economic scenarios are similar to those of the USD EQ portfolio. We note that the JPY EQ portfolio produces an extraordinary result at a cost of huge risk (large loss in 99.5% case). We reckon that such a high expectation of return may not be reliable for two reasons. Firstly, the volatility measure implied by the data is extremely high, leading to an unreliable estimation of the model parameters. Secondly, we observe that the forecast of the JPY EQ index is highly sensitive to the short-term interest rate on German government bond spot curve and the long-term rate of US government bonds. However, as illustrated in Figures 4.6c and 4.6d, the interpolations are much more volatile than the unconditional ones, which seem to be unreasonable as economic forecasts.

Since the economic scenarios mainly vary in terms of US economic variables, we observe more interesting movements of returns on S&P sector and industry indexes across scenarios. Unlike the returns of the GBP, EUR and JPY market indexes, which are more closely related to the S&P 500 index, the returns of consumer discretionary and consumer staples have higher correlation with GDP and unemployment, which are proxies for household income. Moreover, the consumer discretionary sector moves more drastically with the macroeconomic conditions than the consumer staples sector does. The economic implication behind this observation is that, although both sectors reflect the level of consumption in the economy, the consumer discretionary sector, which includes automobile and apparel industry, relates more closely to the luxury needs of consumers. When economic conditions are favorable, as reflected by high GDP growth and low unemployment rate, consumers tend to spend more in the consumer discretionary sector.

Theoretically, the performance of the energy sector should be positively correlated with the oil price. The simulation results are consistent with this intuition as the mean return of the energy sector under the BASE scenario is lower than that under the UP scenario, in which the oil price is higher. The higher mean return under the DOWN scenario is due to the combined effects of a higher expected S&P 500 index and lower oil price. Movements of return in other sectors are generally at reasonable levels, although the implications are harder

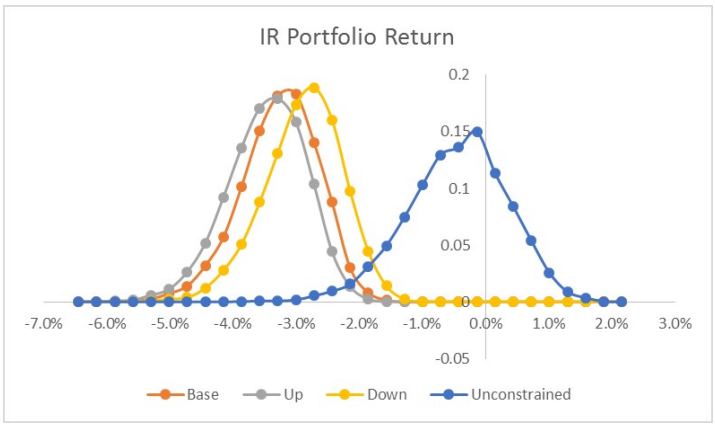


Figure 4.11: Full Conditional Distribution of Return on the Interest Rate Portfolio.

to interpret. On the other hand, the excess return observed in the utility sector under the DOWN scenario over the other two scenarios implies better performance that results from lower operational costs represented by the lower oil price.

Analyzing the IR portfolio is more difficult since the variations come from the entire term structure, rather than a single index. From Figure 4.11, we observe a general trend that the IR portfolio return tends to decrease as macroeconomic conditions become more favorable. In particular, the mean returns under the UP, BASE and DOWN scenarios are -3.6%, -3.4% and -3.0% respectively. This is consistent with the fact that, in recent decades, Federal Reserve increases policy rates to avoid an unhealthy expansion of the economy while reduces the interest rates when economic conditions are unfavorable to stimulate the market. The higher return in the Unconstrained scenario results from a strong EUR, which appreciates by 6% in 2015 under this scenario.

Figure 4.12 presents the mean term structure of interest under each economic scenario. We understand that these term structures of interest do not necessarily result in the mean IR portfolio values. However, a direct look at these term structures still helps in understanding the portfolio performance under each scenario. Both USD and EUR term structures are in shapes similar to that in present year, which guarantees the reasonableness of the generated curves. As expected, the curves generally present a pattern of upward shifting as the economic conditions become more favorable. Only a minor variation is observed in the EUR rates among different economic scenarios due to the construction of the scenarios, except for the

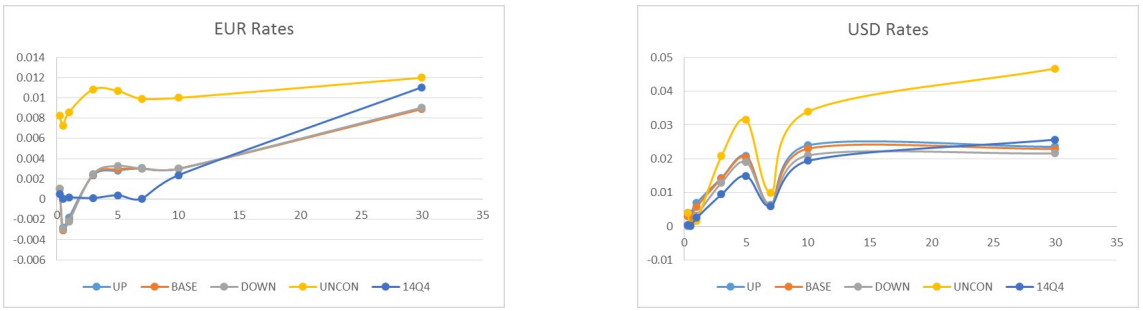


Figure 4.12: The mean term structure of interest under each scenario.

Unconstrained scenario, in which most economic variables deviates significantly from those in the other three scenarios. Due to the model specification for the short end of the EUR term structure, we observe negative interest rates. As mentioned previously, this may be justifiable under the current low interest rate environment.

Even more complicated is the variation of returns under different economic scenarios for the credit portfolio. A general idea is that the spreads on CDS/CDX reflect the likelihood of default for the underlying entity/basket of entities and thus under favorable economic conditions we would expect a narrower spread and vice versa. The simulation results confirm our conjecture: The mean returns under the UP, BASE and DOWN scenarios are respectively 4.2%, 10.7% and 12.4%. The standard deviations of spreads on each credit derivative under each scenario are all greater than the means, which may lead to high volatility in the portfolio return. In addition, we do not attempt to hedge any of these positions. The variations in the interest rate term structure also contribute to the volatility of the CR portfolio. As presented in Figure 4.13, the return can be as high as 400% and as low as -150%. Note that a return lower than -100% is feasible for CDS/CDX as long as the deal spread is greater than 0.

In Figure 4.14 we present the mean spreads of each credit derivative under each economic scenario. In general the spreads vary in the direction as expected, with minor exceptions of a single name IG CDS and a HY CDS in EUR, for which other factors may dominate in terms of the variations.

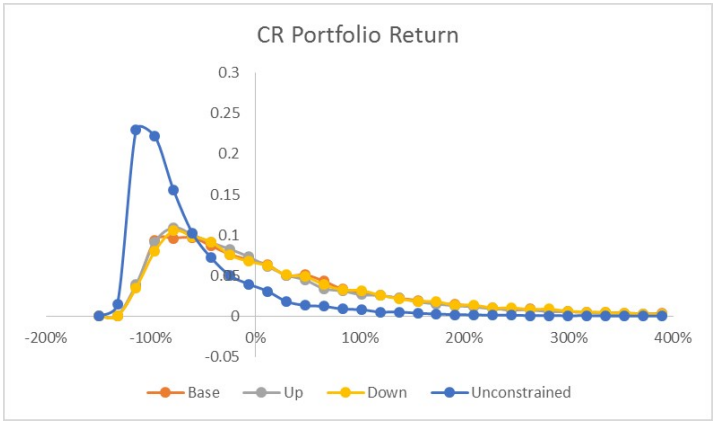


Figure 4.13: Full Conditional Distribution of Return on the Credit Portfolio.

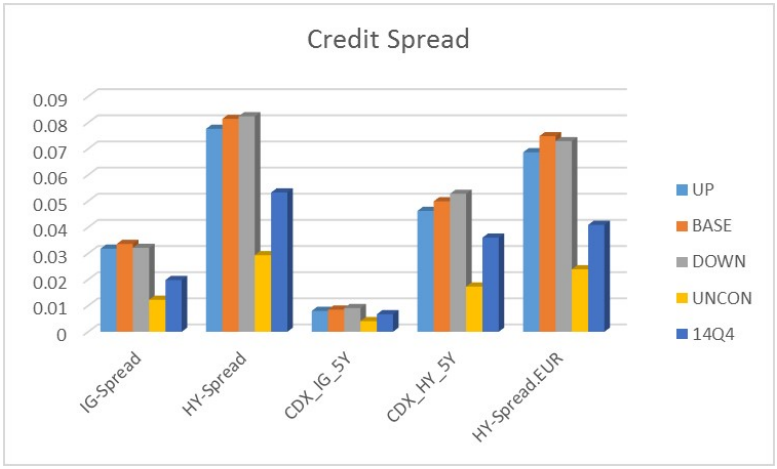


Figure 4.14: Mean Spreads Under Each Economic Scenario.

Chapter 5

Conclusion

In this thesis, we have proposed a GVAR-based framework which allows risk managers to simulate a large set of risk factors that are relevant to a complex multi-currency portfolio of various classes of assets, conditioning on views of the future evolution of the global economy. Thanks to the application of the GVAR model, a large number of risk factors can be modeled endogenously, which effectively captures the interdependence between risk factors within one market or from different markets, with a relatively small sample size. Moreover, the GVAR model is capable of modeling correlations between portfolio risk factors and economic variables. Whether to model economic variables endogenously or exogenously is a flexible choice that can be made by practitioners. In this study, our implementation uses endogenous risk factors with exogenous economic variables. As illustrated in Chapter 4, such a specification allows the users to draw meaningful simulations for risk analytics, which combines the views on future economic conditions and historical paths of portfolio risk factors. When the conditioned set of economic variables is carefully selected and predicted, the simulated evolution of risk factors can be fairly precise. However, the success of an exogenous specification for economic variables does not preclude the possibility of improvement from modeling them endogenously. While a certain level of programming skill is required as the practitioners are likely to be working with high-dimensional data, the implementation is manageable in general. Also, due to the linear framework under which the GVAR model is devised, the estimation and simulation are both relatively tractable. Further, the risk analytics are quite tractable and easily explainable under this linear framework.

Although not demonstrated in this thesis, the GVAR model can be easily extended to simulation of risk factors conditioning on a range of outcomes of economic variables. This can be implemented by using a nested Monte-Carlo approach, where simulations are first drawn from the range of economic variables and then risk factors are simulated on top of the point economic simulation. However, the computational cost can be substantial.

In this study we have focused on exploring the feasibility of applying the GVAR model to the conditional scenario generation problem. The more advanced models for residuals are not implemented in this thesis. Incorporating the features of autocorrelated, non-normal innovation processes, heteroskedasticity and regime-switching can be considered in future work.

In addition, we should mention a few caveats to this thesis. The GVAR methodology has several attractive features, not all of which are exploited in this thesis. Firstly it allows for a systematic treatment of long-run properties through cointegration analysis, and a flexible dynamic specification through vector error correction modeling. In the thesis, we modeled the variables in first difference form, precluding information at the low frequency being utilized in the estimation and, hence, forecast. In addition, since weak exogeneity plays an important role in the GVAR model by allowing conditional subsystem analysis on a sector-by-sector basis, weak exogeneity allows us to focus on the conditional error correction model, which is a GVAR model, and allows us to ignore the marginal error correction model. Secondly, the GVAR model implicitly assumes parameter constancy. Given the sample period used in this thesis, this assumption is likely to be violated, potentially affecting our forecast performances. Third, in a GVAR model, an important marginalization involves aggregation across the variables of the individual foreign countries. This marginalization is testable in principle for a GVAR because the variables of all individual sub-sectors are available: the variables for a given sub-sector are the "domestic" variables for the VARX* for specification for which that particular country is treated as domestic rather than foreign. However testing this marginalization involves a practical difficulty involving degrees of freedom.

Lastly, in this thesis, we have matched frequencies of variables by interpolating low frequency variables to the higher frequency. This is commonly done by adopting a two-step procedure, where first missing data are interpolated, then model parameters are estimated using newly

augmented series. Ideally this procedure should take into account measurement errors induced by disaggregation involved in the interpolation exercise. Both steps can be conveniently and jointly run in a Kalman filter set-up, starting with a state-space representation of the model (see e.g. Harvey, 1990). This interpolation option is generally unsatisfactory since commonly used interpolation methods, including the method used in this thesis, do not fully exploit the available sample information.

Bibliography

Abdalla, Issam S.A. and Victor Murinde (1997). "Exchange rate and stock price interactions in emerging financial markets: evidence on India, Korea, Pakistan and the Philippines". *Applied financial economics* 7.1, pp. 25–35.

Ahn, Sung K. (1988). "Distribution for residual autocovariances in multivariate autoregressive models with structured parameterization". *Biometrika* 75.3, pp. 590–593.

Alexander, Mood, Franklin A. Graybill, and C. Duane (1974). *Introduction to the Theory of Statistics*. McGraw-Hill.

Bańbura, Marta, Domenico Giannone, and Michele Lenza (2015). "Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections". *International Journal of Forecasting* 31.3, pp. 739–756.

Bańbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2010). "Large Bayesian vector auto regressions". *Journal of Applied Econometrics* 25.1, pp. 71–92.

Banerjee, Anindya, Juan J. Dolado, John W. Galbraith, and David Hendry (1993). "Co-integration, error correction, and the econometric analysis of non-stationary data". *OUP Catalogue*.

Beinstein, Eric and Andrew Scott (2006). "Credit derivatives handbook". *JP Morgan Corporate Quantitative Research*, pp. 1–180.

Bernanke, Ben S., Jean Boivin, and Piotr Eliasz (2004). *Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach*. Tech. rep. National Bureau of Economic Research.

Bussière, Matthieu, Alexander Chudik, and Giulia Sestieri (2009). "Modelling global trade flows: results from a GVAR model".

Cairns, Andrew J.G. (2004). "A Family of Term Structure Models for Long Term Risk Management and Derivative Pricing". *Mathematical Finance* 14.3, pp. 415–444.

Chudik, Alexander, Valerie Grossman, and M. Hashem Pesaran (2016). "A multi-country approach to forecasting output growth using PMIs". *Journal of Econometrics* 192.2, pp. 349–365.

Chudik, Alexander and M. Hashem Pesaran (2011). "Infinite-dimensional VARs and factor models". *Journal of Econometrics* 163.1, pp. 4–22.

– (2014). "Theory and practice of GVAR modelling". *Journal of Economic Surveys*.

Cont, Rama (2001). "Empirical properties of asset returns: stylized facts and statistical issues".

Davies, Nelville, C.M. Triggs, and Paul Newbold (1977). "Significance levels of the Box-Pierce portmanteau statistic in finite samples". *Biometrika* 64.3, pp. 517–522.

Dees, Stephane, Filippo di Mauro, M. Hashem Pesaran, and L. Vanessa Smith (2007). "Exploring the international linkages of the euro area: a global VAR analysis". *Journal of Applied Econometrics* 22.1, pp. 1–38.

Diebold, Francis X. and Canlin Li (2006). "Forecasting the term structure of government bond yields". *Journal of Econometrics* 130.2, pp. 337–364.

Engle, Robert F., David F. Hendry, and Jean-Francois Richard (1983). "Exogeneity". *Econometrica: Journal of the Econometric Society*, pp. 277–304.

Fuller, Wayne A. (2009). *Introduction to statistical time series*. Vol. 428. John Wiley & Sons.

Granger, Clive W.J. (1969). "Investigating causal relations by econometric models and cross-spectral methods". *Econometrica: Journal of the Econometric Society*, pp. 424–438.

Hanousek, Jan and Randall K. Filer (2000). "The relationship between economic factors and equity markets in Central Europe". *Economics of transition* 8.3, pp. 623–638.

Harvey, Andrew C (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.

Hosking, John R.M. (1980). "The multivariate portmanteau statistic". *Journal of the American Statistical Association* 75.371, pp. 602–608.

Jarociński, Marek (2010). "Conditional forecasts and uncertainty about forecast revisions in vector autoregressions". *Economics Letters* 108.3, pp. 257–259.

Jarque, Carlos M. and Anil K. Bera (1987). "A test for normality of observations and regression residuals". *International Statistical Review/Revue Internationale de Statistique*, pp. 163–172.

Kadiyala, K. Rao and Sune Karlsson (1997). "Numerical methods for estimation and inference in Bayesian VAR-models". *Journal of Applied Econometrics* 12.2, pp. 99–132.

Kim, Chang-Jin and Charles R. Nelson (1999). *State-space models with regime switching: classical and Gibbs-sampling approaches with applications*. Vol. 2. MIT press Cambridge, MA.

Kim, Tong Suk, Yuen Jung Park, and Jaewon Park (2013). "Macroeconomic Conditions and Credit Default Swap (CDS) Spread Changes". *Available at SSRN 2530836*.

Kirkulak, Berna and Hassan Ezzat (2014). "Market Randomness and Weak-Form Efficiency Revisited: A Global Investigation". *Available at SSRN 2511014*.

Litterman, Robert B. (1986). "Forecasting with Bayesian vector autoregressions—five years of experience". *Journal of Business & Economic Statistics* 4.1, pp. 25–38.

Ljung, Greta M. and George E.P. Box (1978). "On a measure of lack of fit in time series models". *Biometrika* 65.2, pp. 297–303.

Lomnicki, Zbigniew A. (1961). "Tests for departure from normality in the case of linear stochastic processes". *Metrika* 4.1, pp. 37–62.

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

Meucci, Attilio (2008). "Fully flexible views: Theory and practice". *Risk* 21.10, pp. 97–102.

– (2013). "Estimation and stress-testing via time-and market-conditional flexible probabilities". *Available at SSRN 2312126*.

Mukherjee, Tarun K. and Atsuyuki Naka (1995). "Dynamic relations between macroeconomic variables and the Japanese stock market: an application of a vector error correction model". *Journal of Financial Research* 18.2, pp. 223–237.

Pesaran, M. Hashem, Til Schuermann, and Scott M. Weiner (2004). "Modeling regional interdependencies using a global error-correcting macroeconometric model". *Journal of Business & Economic Statistics* 22.2, pp. 129–162.

Rebonato, Riccardo, Sukhdeep Mahal, Mark Joshi, Lars-Dierk Buchholz, and Ken Nyholm (2005). "Evolving yield curves in the real-world measures: a semi-parametric approach". *The Journal of Risk* 7.3, p. 1.

Rosen, Dan and David Saunders (2015). "Regress under stress: A simple least-squares method for integrating economic scenarios with risk simulations". *Available at SSRN 2699474*.

Schmidt, Peter (1976). *Econometrics*. Marcel Dekker, New York.

Serfling, Robert J. (1980). *Approximation theorems of mathematical statistics*. Vol. 162. John Wiley & Sons.

Sims, Christopher A. (1992). "Interpreting the macroeconomic time series facts: The effects of monetary policy". *European Economic Review* 36.5, pp. 975–1000.

Sims, Christopher A. and Tao Zha (1998). "Bayesian methods for dynamic multivariate models". *International Economic Review*, pp. 949–968.

Stock, James H. and Mark W. Watson (2002). "Macroeconomic forecasting using diffusion indexes". *Journal of Business & Economic Statistics* 20.2, pp. 147–162.

Tanaka-Yamawaki, Mieko (2011). "Cross correlation of intra-day stock prices in comparison to Random Matrix Theory". *Intelligent Information Management* 3.03, p. 65.

Tripathi, Vanita and Ritika Seth (2014). "Stock market performance and macroeconomic factors: The study of the Indian equity market". *Global Business Review* 15.2, pp. 291–316.

Waggoner, Daniel F. and Tao Zha (1999). "Conditional forecasts in dynamic multivariate models". *Review of Economics and Statistics* 81.4, pp. 639–651.

White, Halbert and Glenn M. MacDonald (1980). "Some large-sample tests for nonnormality in the linear regression model". *Journal of the American Statistical Association* 75.369, pp. 16–28.

Wilkie, A. David (1984). "A Stochastic Investment Model for Actuarial Use." *Transactions of the Faculty of Actuaries* 39, pp. 341–403.

Xiang, Ju and Xiaoneng Zhu (2013). "A Regime-Switching Nelson–Siegel Term Structure Model and Interest Rate Forecasts". *Journal of Financial Econometrics*, nbs021.

Zellner, Arnold (1962). "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias". *Journal of the American Statistical Association* 57.298, pp. 348–368.

Appendix A

Matrices

We reproduce part of the Appendix A in Lütkepohl (2005) for reference. The following sections elucidate some operators and rules used in the chapters of this thesis.

A.1 Kronecker Product

Let $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ be $(m \times n)$ and $(p \times q)$ matrices respectively. The $(mp \times nq)$ matrix

$$\mathbf{A} \otimes \mathbf{B} := \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix} \tag{A.1}$$

is a *Kronecker product* or *direct product* of matrices $\mathbf{A}$ and $\mathbf{B}$. Assuming that all matrices below are of conformable dimensions, here we list a number of rules for the *Kronecker product*:

(1) $\mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A}$ in general.

(2) $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$.

(3) $\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C}$.

(4) $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$.

(5) If $\mathbf{A}$ and $\mathbf{B}$ are invertible, then $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$.

(6) If $\mathbf{A}$ and $\mathbf{B}$ are square matrices, then $\operatorname{tr}(\mathbf{A} \otimes \mathbf{B}) = \operatorname{tr}(\mathbf{A}) \operatorname{tr}(\mathbf{B})$.

A.2 The vec and vech Operators

Let $\mathbf{A} = (a_1, \ldots, a_n)$ be an $(m \times n)$ matrix with $(m \times 1)$ columns a_i. The *vec operator* transforms $\mathbf{A}$ into an $(mn \times 1)$ vector by stacking the columns, namely,

$$\text{vec}(\mathbf{A}) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}. \tag{A.2}$$

Assume matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are of conformable dimensions, we have following rules:

(1) $\text{vec}(\mathbf{A} + \mathbf{B}) = \text{vec}(\mathbf{A}) + \text{vec}(\mathbf{B})$.

(2) $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\,\text{vec}(\mathbf{B})$.

(3) $\text{vec}(\mathbf{AB}) = (\mathbf{I} \otimes \mathbf{A})\,\text{vec}(\mathbf{B}) = (\mathbf{B}' \otimes \mathbf{I})\,\text{vec}(\mathbf{A})$.

(4) $\text{vec}(\mathbf{ABC}) = (\mathbf{IAB})\,\text{vec}(\mathbf{C}) = (\mathbf{C}'\mathbf{B} \otimes \mathbf{I})\,\text{vec}(\mathbf{A})$.

(5) $\text{vec}(\mathbf{B}')'\,\text{vec}(\mathbf{A}) = \text{tr}(\mathbf{BA}) = \text{tr}(\mathbf{AB}) = \text{vec}(\mathbf{A}')'\,\text{vec}(\mathbf{B})$.

(6)

$$\begin{aligned}
\text{tr}(\mathbf{ABC}) &= \text{vec}(\mathbf{A}')'(\mathbf{C}' \otimes \mathbf{I})\,\text{vec}(\mathbf{B}) \\
&= \text{vec}(\mathbf{A}')'(\mathbf{I} \otimes \mathbf{B})\,\text{vec}(\mathbf{C}) \\
&= \text{vec}(\mathbf{B}')'(\mathbf{A}' \otimes \mathbf{I})\,\text{vec}(\mathbf{C}) \\
&= \text{vec}(\mathbf{B}')'(\mathbf{I} \otimes \mathbf{C})\,\text{vec}(\mathbf{A}) \\
&= \text{vec}(\mathbf{C}')'(\mathbf{B}' \otimes \mathbf{I})\,\text{vec}(\mathbf{A}) \\
&= \text{vec}(\mathbf{C}')'(\mathbf{I} \otimes \mathbf{A})\,\text{vec}(\mathbf{B})
\end{aligned}$$

The *vech operator* is similar to vec, which stacks only the elements on and below the main diagonal of a square matrix. In general, for an $(m \times m)$ matrix $\mathbf{A}$, $\text{vech}(\mathbf{A})$ is an $m(m+1)/2$-dimensional vector. The vech operator is usually applied to a symmetric matrix to collect distinct elements only.

A.3 Vector and Matrix Differentiation

In optimization with respect to a multivariate function, using vector/matrix differentiation helps tidy up the derivation. In essence, differentiation of vector/matrix form is equivalent to summarizing all derivatives of a certain function in a vector/matrix. Assuming all derivatives exist and are continuous, let $f(\beta)$ be a scalar function that depends on the $(n \times 1)$ vector $\beta = (\beta_1, \ldots, \beta_n)'$. We define the following:

$$\frac{\partial f}{\partial \beta} := \begin{bmatrix} \frac{\partial f}{\partial \beta_1} \\ \vdots \\ \frac{\partial f}{\partial \beta_n} \end{bmatrix}, \qquad \frac{\partial f}{\partial \beta'} := \begin{bmatrix} \frac{\partial f}{\partial \beta_1}, \ldots, \frac{\partial f}{\partial \beta_n} \end{bmatrix},$$

which are $(n \times 1)$ and $(1 \times n)$ vectors of first order partial derivatives, respectively. Moreover,

$$\frac{\partial^2 f}{\partial \beta \partial \beta'} := \begin{bmatrix} \frac{\partial^2 f}{\partial \beta_i \partial \beta_j} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 f}{\partial \beta_1 \partial \beta_1} & \cdots & \frac{\partial^2 f}{\partial \beta_1 \partial \beta_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial \beta_n \partial \beta_1} & \cdots & \frac{\partial^2 f}{\partial \beta_n \partial \beta_n} \end{bmatrix}$$

is the $(n \times n)$ *Hessian matrix* of second order partial derivatives. If $f(\mathbf{A})$ is a scalar function of an $(m \times n)$ matrix $\mathbf{A} = (a_{ij})$, then

$$\frac{\partial f}{\mathbf{A}} := \begin{bmatrix} \frac{\partial f}{\partial a_{ij}} \end{bmatrix} \tag{A.3}$$

is an $(m \times n)$ matrix of partial derivatives. Based on the above definitions, we give some rules of vector/matrix differentiation without proof:

(1) Let $\mathbf{A}$ be an $(m \times n)$ matrix and β be an $(n \times 1)$ vector. Then

$$\frac{\partial \mathbf{A}\beta}{\partial \beta'} = \mathbf{A} \text{ and } \frac{\partial \beta' \mathbf{A}'}{\partial \beta} = \mathbf{A}'. \tag{A.4}$$

(2) Let $\mathbf{A}$ be $(m \times m)$ and β be $(m \times 1)$, then

$$\frac{\partial \beta' \mathbf{A} \beta}{\partial \beta} = (\mathbf{A} + \mathbf{A}')\beta \text{ and } \frac{\partial \beta' \mathbf{A} \beta}{\partial \beta'} = \beta'(\mathbf{A}' + \mathbf{A}). \tag{A.5}$$

(3) Let $\mathbf{A}$ be $(m \times m)$ and β be $(m \times 1)$, then

$$\frac{\partial^2 \beta' \mathbf{A} \beta}{\partial \beta \partial \beta'} = \mathbf{A} + \mathbf{A}'. \tag{A.6}$$

(4) If $\mathbf{A}$ is a symmetric $(m \times m)$ matrix and β an $(m \times 1)$ vector then

$$\frac{\partial^2 \beta' \mathbf{A} \beta}{\partial \beta \partial \beta'} = 2\mathbf{A}. \tag{A.7}$$

(5) Let Ω be a symmetric $(n \times n)$ matrix and $c(\beta)$ an $(n \times 1)$ vector that depends on the $(m \times 1)$ vector β. Then

$$\frac{\partial c(\beta)' \Omega c(\beta)}{\partial \beta} = 2c(\beta)' \frac{\partial c(\beta)}{\partial \beta'} \tag{A.8}$$

and

$$\frac{\partial^2 c(\beta)' \Omega c(\beta)}{\partial \beta \partial \beta'} = 2\left[\frac{\partial c(\beta)'}{\partial \beta} \Omega \frac{\partial c(\beta)}{\partial \beta'} + [c(\beta)' \Omega \otimes \mathbf{I}_m] \frac{\partial \operatorname{vec}(\partial c(\beta)'/\partial \beta)}{\partial \beta'}\right]. \tag{A.9}$$

In particular, if y is an $(n \times 1)$ vector and $\mathbf{X}$ is an $(n \times m)$ matrix, then

$$\frac{\partial (y - \mathbf{X}\beta)' \Omega (y - \mathbf{X}\beta)}{\partial \beta'} = -2(y - \mathbf{X}\beta)' \Omega \mathbf{X} \tag{A.10}$$

and

$$\frac{\partial^2 (y - \mathbf{X}\beta)' \Omega (y - \mathbf{X}\beta)}{\partial \beta \partial \beta'} = 2\mathbf{X}' \Omega \mathbf{X}. \tag{A.11}$$

Appendix B

Stochastic Convergence

In the derivation of the limiting distribution for the estimator, we applied some results in the theory of stochastic convergence. In this appendix we reproduce part of Appendix C in Lütkepohl (2005) for reference. This appendix is organized as follows: In Section B.1, convergence of a sequence of random variables is defined in a number of ways. These concepts are then extended to random vectors. Section B.1 is then concluded with some properties of stochastic convergence. Some results regarding determination of the asymptotic distribution of estimators and test statistics are given in Section B.2.

B.1 Basic Concepts of Stochastic Convergence

We give definitions on four types of convergence of a sequence of random variables.

Definition B.1 (Convergence in Probability)**.** *Let x_1, x_2, ..., or $\{x_T\}$, $T = 1, 2, \ldots$, be a sequence of scalar random variables which are all defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The sequence $\{x_T\}$ converges in probability to the random variable x (which is also defined on $(\Omega, \mathcal{F}, \mathbb{P})$) if for every $\epsilon > 0$,*

$$\lim_{T\to\infty} \mathbb{P}\{|x_T - x| > \epsilon\} = 0 \tag{B.1}$$

or equivalently,

$$\lim_{T\to\infty} \mathbb{P}\{|x_T - x| < \epsilon\} = 1. \tag{B.2}$$

This type of convergence is abbreviated as

$$\operatorname{plim} x_T = x \quad \text{or} \quad x_T \xrightarrow{p} x,$$

where x can be either a random variable or a fixed, non-stochastic real number which is a so-called a degenerate random variable that takes only one value a with probability one.

Definition B.2 (Almost Sure Convergence)**.** *The sequence* $\{x_T\}$ *converges almost surely or with probability one to the random variable* x *if for every* $\epsilon > 0$*,*

$$\mathbb{P}\{\lim_{T\to\infty} |x_T - x| < \epsilon\} = 1. \tag{B.3}$$

This type of convergence is often written as

$$x_T \xrightarrow{a.s.} x$$

and is also called strong convergence.

Definition B.3 (Convergence in Quadratic Mean)**.** *The sequence* $\{x_T\}$ *converges in quadratic mean or mean square error to* x *if*

$$\lim_{T\to\infty} \mathbb{E}(x_T - x)^2 = 0. \tag{B.4}$$

This type of convergence is written as

$$x_T \xrightarrow{q.m.} x$$

and requires that both mean and variance exist for x_T, $T = 1, 2, \ldots$ and x.
Denoting the distribution functions of x_T and x by F_T and F respectively, we define the *convergence in distribution* or *weak convergence* or *convergence in law* in following definition.

Definition B.4 (Convergence in Distribution)**.** *The sequence* $\{x_T\}$ *converges in distribution to* x *if for all real numbers* c *for which* F *is continuous,*

$$\lim_{T\to\infty} F_T(c) = F(c). \tag{B.5}$$

This type of convergence is abbreviated as

$$x_T \xrightarrow{d} x.$$

One important point to note that convergence in distribution does not imply convergence in probability density function (p.d.f.) in general. Moreover, it does not even require p.d.f to exist.

All of the aforementioned definitions can be easily extended to a sequence of random vectors. Denote a sequence of K-dimensional random vectors as $\{X_T = (x_{1T}, \ldots, x'_{KT})\}$, $T = 1, 2, \ldots$ and a K-dimensional random vector as $X = (x_1, \ldots, x'_K)$. Further, denote the joint distribution of X_T and X as F_T and F respectively. Then the following are corresponding to the four definitions above.

$$\begin{aligned}
&\operatorname{plim} X_T = X \text{ or } X_T \xrightarrow{p} X \text{ if } \operatorname{plim} x_{kT} = x_k \text{ for } k = 1, \ldots, K.\\
&X_T \xrightarrow{a.s.} X \text{ if } x_{kT} \xrightarrow{a.s.} x_k \text{ for } k = 1, \ldots, K.\\
&X_T \xrightarrow{q.m.} \text{ if } \lim_{T\to\infty} \mathbb{E}[(X_T - X)'(X_T - X)] = 0.\\
&X_T \xrightarrow{d} X \text{ if } \lim_{T\to\infty} F_T(c) = F(c) \text{ for all continuity points of } F.
\end{aligned}$$

The properties of stochastic convergence defined above are given in the following two propositions in concluding the appendix.

Proposition B.1 (Convergence Properties of Sequences of Random Variables)**.** *Suppose* $\{x_T\}$ *is a sequence of random variables. Then the following relations hold:*

(1) $x_T \xrightarrow{a.s.} x \Rightarrow x_T \xrightarrow{p} x \Rightarrow x_T \xrightarrow{d} x.$

(2) $x_T \xrightarrow{q.m.} x \Rightarrow x_T \xrightarrow{p} x \Rightarrow x_T \xrightarrow{d} x.$

(3) If x *is a fixed, non-stochastic vector, then*

$$x_T \xrightarrow{q.m.} x \Leftrightarrow [\lim_{T\to\infty} \mathbb{E}(x_T) = x \text{ and } \lim_{T\to\infty} \mathbb{E}[(x_T - \mathbb{E}(x_T))'(x_T - \mathbb{E}(x_T))] = 0].$$

(4) If x *is a fixed, non-stochastic random vector, then*

$$x_T \xrightarrow{p} x \Leftrightarrow x_T \xrightarrow{d} x.$$

(5) (Slutsky's Theorem) If $g : \mathbb{R}^K \to \mathbb{R}^m$ is a continuous function, then

$$x_T \xrightarrow{p} x \Rightarrow g(x_T) \xrightarrow{p} g(x)[\operatorname{plim} g(x_T) = g(\operatorname{plim} x_T)],$$
$$x_T \xrightarrow{d} x \Rightarrow g(x_T) \xrightarrow{d} g(x),$$

and

$$x_T \xrightarrow{a.s.} x \Rightarrow g(x_T) \xrightarrow{a.s.} g(x).$$

Proposition B.2 (Properties of Convergence in Probability and in Distribution)**.** *Suppose $\{x_t\}$ and $\{y_T\}$ are sequences of $(K \times 1)$ random vectors, $\{\mathbf{A}_T\}$ is a sequence of $(K \times K)$ random matrices, x is a $(K \times 1)$ random vector, c is a fixed $(K \times 1)$ vector, and $\mathbf{A}$ is a fixed $(K \times K)$ matrix.*

(1) If $\operatorname{plim} x_T$, $\operatorname{plim} y_T$, and $\operatorname{plim} \mathbf{A}_T$ exist, then

- *(a)* $\operatorname{plim}(x_T \pm y_T) = \operatorname{plim} x_T \pm \operatorname{plim} y_T$;
- *(b)* $\operatorname{plim}(c' x_T) = c'(\operatorname{plim} x_T)$;
- *(c)* $\operatorname{plim} x_T' y_T = (\operatorname{plim} x_T)'(\operatorname{plim} y_T)$;
- *(d)* $\operatorname{plim} \mathbf{A}_T x_T = \operatorname{plim}(\mathbf{A}_T) \operatorname{plim}(x_T)$.

(2) If $x_T \xrightarrow{d} x$ and $\operatorname{plim}(x_T - y_T) = 0$, then $y_T \xrightarrow{d} x$.

(3) If $x_T \xrightarrow{d} x$ and $\operatorname{plim} y_T = c$, then

- *(a)* $x_T \pm y_T \xrightarrow{d} x \pm c$;
- *(b)* $y_T' x_T \xrightarrow{d} c' x$.

(4) If $x_T \xrightarrow{d} x$ and $\operatorname{plim} \mathbf{A}_T = \mathbf{A}$, then $\mathbf{A}_T x_T \xrightarrow{d} \mathbf{A} x$.

(5) If $x_T \xrightarrow{d} x$ and $\operatorname{plim} \mathbf{A}_T = 0$, then $\operatorname{plim} \mathbf{A}_T x_T = 0$.

B.2 Standard Asymptotic Properties of Estimators and Test Statistics

Let $\hat{\beta}_T$ be an estimator of a $(K \times 1)$ parameter vector, where the subscript T denotes the size of the sample with which the estimator is computed. Then $\{\hat{\beta}_T\}$ for $T = 1, 2, \ldots$ is a sequence of random variables. Applying the concepts of stochastic convergence from the previous section, the estimator is said to have an *asymptotic normal distribution* if

$$\sqrt{T}(\hat{\beta}_T - \beta) \xrightarrow{d} N(0, \Sigma). \tag{B.6}$$

Once we obtain the estimator and its asymptotic distribution, we may be interested in testing the significance of our estimates. To cover as many cases as possible, we usually derive test statistics as a general function of estimators. The following results are given to assist such derivations.

Proposition B.3 (Asymptotic Properties of Estimators)**.** *Suppose $\hat{\beta}_T$ is an estimator of the $(K \times 1)$ vector β with $\sqrt{T}(\hat{\beta}_T - \beta) \xrightarrow{d} N(0, \Sigma)$. Then the following rules hold:*

(1) (A special case of (2)) If $\mathbf{R} \neq \mathbf{0}$ is a $(M \times K)$ matrix, then $\sqrt{T}(\mathbf{R}\hat{\beta}_T - \mathbf{R}\beta) \xrightarrow{d} N(0, \mathbf{R}\Sigma\mathbf{R}')$.

(2) For a sequence of $(M \times K)$ matrices, $\{\hat{\mathbf{A}}_T\}$, and a fixed matrix $\mathbf{A}$, if $\operatorname{plim} \hat{\mathbf{A}}_T = \mathbf{A}$, then $\sqrt{T}\hat{\mathbf{A}}(\hat{\beta}_T - \beta) \xrightarrow{d} N(0, \mathbf{A}\Sigma\mathbf{A}')$ (see Schmidt, 1976, p. 251).

(3) (Delta Method) If $g(\beta) = (g_1(\beta), \ldots, g_m(\beta))$ is a vector-valued continuously differentiable function with $\frac{\partial g}{\partial \beta'} \neq 0$ at β, then

$$\sqrt{T}[g(\hat{\beta}_T) - g(\beta)] \xrightarrow{d} N\left(0, \frac{\partial g(\beta)}{\partial \beta'} \Sigma \frac{\partial g(\beta)'}{\partial \beta}\right).$$

$\frac{\partial g}{\partial \beta'} = 0$ at β, $\sqrt{T}[g(\hat{\beta}) - g(\beta)] \xrightarrow{p} 0$ (see Serfling, 1980, pp. 122 - 124).

(4) (A special case of (5)) If Σ is nonsingular, then $T(\hat{\beta}_T - \beta)'\Sigma^{-1}(\hat{\beta}_T - \beta) \xrightarrow{d} \chi^2(K)$.

(5) If Σ is nonsingular and $\operatorname{plim} \hat{\Sigma} = \Sigma$, then $T(\hat{\beta}_T - \beta)'\hat{\Sigma}^{-1}(\hat{\beta}_T - \beta) \xrightarrow{d} \chi^2(K)$.

(6) If $\Sigma = \mathbf{QA}$, where $\mathbf{Q}$ is symmetric, idempotent of rank n and $\mathbf{A}$ is positive definite, then $T(\hat{\beta}_T - \beta)'\mathbf{A}^{-1}(\hat{\beta}_T - \beta) \xrightarrow{d} \chi^2(n)$.

Appendix C

Tables and Figures

C.1 Augmented Dickey-Fuller Test for Economic Variables

Without carrying out an explicit model selection, we conduct an augmented Dickey-Fuller (ADF) test with lags 2, 3 and 4 on time series of levels and differences of economic variables. The treatment for each variable at first difference is tabulated. We only present the p-values of the test statistics here due to space limitations.

Table C.1: Augmented Dickey-Fuller Test for Economic Variables

	Level				Difference		
	Lag				Lag		
Variable	2	3	4	Treatment	2	3	4
GB_REAL_GDP	0.536105	0.603313	0.475915	LogDiff	0.01824	0.09015	0.07241
US_REAL_GDP	0.543994	0.500602	0.43435	LogDiff	0.07052	0.25060	0.19375
XE_REAL_GDP	0.380132	0.352727	0.371051	LogDiff	0.05101	0.05608	0.10234
JP_REAL_GDP	0.249296	0.307847	0.489821	LogDiff	0.01000	0.01000	0.01000
GB_CPI_INF	0.567786	0.585629	0.498254	LogDiff	0.01000	0.37942	0.37767
US_CPI_INF	0.474645	0.795632	0.504362	LogDiff	0.01000	0.01000	0.02019
XE_CPI_INF	0.65477	0.579008	0.166637	LogDiff	0.03585	0.39556	0.16026
JP_CPI_INF	0.952046	0.945243	0.643819	LogDiff	0.01000	0.30105	0.13937
US_UNEMPLOYMENT	0.360572	0.407272	0.530428	Diff	0.35287	0.27675	0.26806
US_RATE_SHORT	0.544276	0.234056	0.221397	Diff	0.19882	0.24223	0.33962
XE_RATE_SHORT	0.11274	0.187283	0.059326	Diff	0.01000	0.05234	0.08835
US_RATE_LONG	0.118292	0.074081	0.214464	Diff	0.01000	0.01000	0.01838
XE_RATE_LONG	0.293841	0.324591	0.452307	Diff	0.01000	0.01000	0.01000
GB_FX	0.659565	0.605172	0.637913	LogDiff	0.01000	0.01000	0.07831
XE_FX	0.664592	0.729616	0.910095	LogDiff	0.01954	0.01000	0.13722
JP_FX	0.973797	0.74439	0.565644	LogDiff	0.06304	0.25941	0.41830
WRLD_OIL	0.383644	0.294265	0.487478	LogDiff	0.01000	0.01000	0.01000
US_MKT_IDX	0.843911	0.864877	0.737064	LogDiff	0.01000	0.01923	0.07116

The tabulated values are p-values of the ADF test under the null hypothesis of the time series being integrated to order one.
The p-values are interpolated from Banerjee et al. (1993). If the computed statistic is outside the table of critical values, we use the left/right end value instead.

C.2 Augmented Dickey-Fuller Test for Risk Factors

Table C.2: Augmented Dickey-Fuller Test for Portfolio Risk Variables - Part I

	Level			Difference			
	Lag				Lag		
Variable	2	3	4	Treatment	2	3	4
IG-Spread	0.36485	0.14856	0.334556	LogDiff	0.010151	0.022712	0.048518
HY-Spread	0.377516	0.226805	0.30149	LogDiff	0.01	0.047204	0.059702
CDX_IG_5Y	0.383326	0.280532	0.405414	LogDiff	0.011429	0.024954	0.080388
CDX_HY_5Y	0.365938	0.280946	0.346332	LogDiff	0.01	0.03429	0.075508
HY-Spread.EUR	0.281546	0.303436	0.220777	LogDiff	0.013319	0.054813	0.112688
FTSE350	0.451649	0.405954	0.279899	LogDiff	0.01	0.035513	0.088225
MSCI.EMU	0.279591	0.475498	0.412191	LogDiff	0.01	0.017681	0.038522
Topix.500	0.492604	0.407282	0.359512	LogDiff	0.038733	0.065949	0.10273
SP500.ConsDisc	0.99	0.99	0.975881	LogDiff	0.01	0.018203	0.058906
SP500.ConsStaple	0.957742	0.986331	0.99	LogDiff	0.01	0.01	0.01
SP500.Energy	0.256649	0.33145	0.3507	LogDiff	0.01	0.016845	0.143889
SP500.Fin	0.612738	0.52816	0.472081	LogDiff	0.01	0.039317	0.175386
SP500.Health	0.99	0.99	0.99	LogDiff	0.01	0.01	0.195628
SP500.Indust	0.659596	0.711254	0.649092	LogDiff	0.01	0.01	0.057087
SP500.InfoTech	0.464236	0.526442	0.680733	LogDiff	0.01	0.010032	0.011606

The tabulated values are p-values of the ADF test under the null hypothesis of the time series being integrated to order one.
The p-values are interpolated from Banerjee et al. (1993). If the computed statistic is outside the table of critical values, we use the left/right end value instead.

Table C.3: Augmented Dickey-Fuller Test for Portfolio Risk Variables - Part II

	Level			Difference			
	Lag				Lag		
Variable	2	3	4	Treatment	2	3	4
SP500.Mat	0.166276	0.404451	0.247619	LogDiff	0.01	0.01	0.01
SP500.Telco	0.219307	0.304915	0.263025	LogDiff	0.01	0.01	0.089632
SP500.Utils	0.388003	0.41505	0.545121	LogDiff	0.01	0.01	0.147725
EUR_Govt_Zero.6m	0.265037	0.167249	0.092771	LogDiff	0.017688	0.048323	0.060962
EUR_Govt_Zero.1y	0.349329	0.234493	0.203469	LogDiff	0.01	0.03473	0.049862
EUR_Govt_Zero.3y	0.349778	0.307813	0.380749	LogDiff	0.01	0.01	0.026049
EUR_Govt_Zero.5y	0.313805	0.31825	0.426638	LogDiff	0.01	0.01	0.01
EUR_Govt_Zero.7y	0.289751	0.343889	0.447053	LogDiff	0.01	0.01	0.01
EUR_Govt_Zero.30y	0.182382	0.258407	0.463632	LogDiff	0.01	0.039318	0.07355
USD_Govt_US_Zero.6M	0.54159	0.206237	0.211683	LogDiff	0.243922	0.239513	0.34585
USD_Govt_US_Zero.1Y	0.572611	0.276462	0.183734	LogDiff	0.165392	0.257596	0.375877
USD_Govt_US_Zero.3Y	0.585677	0.461374	0.321545	LogDiff	0.01	0.069408	0.180836
USD_Govt_US_Zero.5Y	0.475685	0.393201	0.384522	LogDiff	0.01	0.018577	0.04824
USD_Govt_US_Zero.7Y	0.296442	0.211119	0.313282	LogDiff	0.092384	0.116612	0.253946
USD_Govt_US_Zero.30Y	0.013576	0.01	0.011999	LogDiff	0.01	0.01	0.01

The tabulated values are p-values of the ADF test under the null hypothesis of the time series being integrated to order one.
The p-values are interpolated from Banerjee et al. (1993). If the computed statistic is outside the table of critical values, we use the left/right end value instead.

C.3 Causality Tests

Table C.4: Causality Tests for VARX* Models

Variable	Granger Causality	Instantaneous Causality
IG-Spread	0.00968	0.106749
HY-Spread	7.73E-08	0.09784
CDX_IG_CDSI_GEN_5Y	0.005815	0.111969
CDX_HY_CDSI_GEN_5Y	7.52E-13	0.064042
HY-Spread.EUR	0.008859	0.099459
FTSE350	2.10E-06	0.034319
MSCI.EMU	0.000579	0.03262
Topix.500	7.19E-05	0.118333
SP500.ConsDisc	3.71E-06	0.026485
SP500.ConsStaple	5.09E-06	0.144094
SP500.Energy	0.001035	0.040948
SP500.Fin	3.76E-07	0.034076
SP500.Health	0.006188	0.084801
SP500.Indust	0.00063	0.031201
SP500.InfoTech	0.000722	0.033973
SP500.Mat	2.35E-13	0.071423
SP500.Telco	2.93E-14	0.073516
SP500.Utils	4.51E-08	0.147638
EUR_Govt.6m	1.35E-09	0.021548
EUR_Govt.1y	3.49E-10	0.028651
EUR_Govt.3y	7.14E-10	0.029599
EUR_Govt.5y	1.23E-09	0.023554
EUR_Govt.7y	2.77E-12	0.020563
EUR_Govt.30y	5.22E-11	0.038146
USD_Govt_US.6M	0	0.02166
USD_Govt_US.1Y	3.33E-14	0.028863
USD_Govt_US.3Y	5.04E-12	0.04522
USD_Govt_US.5Y	8.63E-07	0.033148
USD_Govt_US.7Y	5.24E-11	0.108061
USD_Govt_US.30Y	8.98E-06	0.030496

C.4 Model Selection

We fix the number of lags included for economic variables at zero to our view that all markets under consideration are at least semi-strong efficient. In practice, depending on data availability and development levels of the markets, more lags can be included.

Table C.5: Selected Lags for each Variable with AIC and BIC

	AIC			BIC		
	X_t	X_t^*	Y_t	X_t	X_t^*	Y_t
IG-Spread	4	2	0	1	0	0
HY-Spread	1	2	0	3	0	0
CDX_IG_5Y	1	2	0	2	0	0
CDX_HY_5Y	5	1	0	1	0	0
HY-Spread.EUR	2	2	0	1	1	0
FTSE350	2	2	0	1	0	0
MSCI.EMU	1	1	0	1	0	0
Topix.500	2	0	0	2	0	0
SP500.ConsDisc	1	2	0	1	0	0
SP500.ConsStaple	1	0	0	1	0	0
SP500.Energy	5	1	0	2	1	0
SP500.Fin	4	1	0	4	0	0
SP500.Health	1	2	0	1	0	0
SP500.Indust	5	2	0	1	1	0
SP500.InfoTech	1	1	0	1	0	0
SP500.Mat	5	2	0	5	2	0
SP500.Telco	2	2	0	1	0	0
SP500.Utils	4	2	0	1	0	0
EUR_Govt.6m	2	2	0	2	0	0
EUR_Govt.1y	2	2	0	2	0	0
EUR_Govt.3y	1	2	0	1	0	0
EUR_Govt.5y	5	2	0	1	0	0
EUR_Govt.7y	1	2	0	1	0	0
EUR_Govt.30y	4	2	0	1	0	0
USD_Govt_US.6M	3	2	0	1	0	0
USD_Govt_US.1Y	3	2	0	1	0	0
USD_Govt_US.3Y	5	2	0	1	0	0
USD_Govt_US.5Y	5	2	0	1	0	0
USD_Govt_US.7Y	5	2	0	5	1	0
USD_Govt_US.30Y	5	2	0	5	2	0