FINANCIAL INTERLINKAGES IN EAST AFRICAN COUNTRIES: CONNECTING THE BAYESIAN MODEL AVERAGING (BMA) APPROACH TO GLOBAL VECTOR AUTOREGRESSIVE (GVAR)

DANIEL NGUGI NJOORA

MASTER OF SCIENCE

(Mathematics-Financial Option)

PAN AFRICAN UNIVERSITY

INSTITUTE FOR BASIC SCIENCES TECHNOLOGY AND INNOVATION

2014

Financial Interlinkages in East African Countries: Connecting the Bayesian Model Averaging (BMA) Approach to Global Vector Autoregressive (GVAR)

Daniel Ngugi Njoora

MF 300-0003/12

A thesis submitted to The Pan African University Institute for Basic Sciences, Technology and Innovation in partial fulfilment of the requirements for the degree of Master of Science in Mathematics (Financial option).

2014

Declaration

I declare that this thesis is my own work and has not been submitted in any university or other institution for the award of a degree or any other award.

Signature:.................................. Date:.........................

Daniel Ngugi Njoora

This thesis report has been submitted for examination with our approval as University supervisors.

Signature: Date: 14/11/2014

Dr. O. E. Olubusoye(University of Ibadan, Nigeria)

Signature:.. Date:.................................

Prof. Patrick Weke (University of Nairobi, Kenya)

Dedication

To my parents, wife and daughter.

Acknowledgement

I thank the Almighty God for helping and sustaining me this far. I also acknowledge the support I received from both staff and students of the Pan African University, Institute for Basic Sciences Technology and Innovation (PAUISTI) during my study.

I wish to extend my sincere gratitude to my supervisor, Dr. O.E. Olubusoye for sparing time to guide, direct and assist me whenever I encountered challenges. Despite the long distance separating us apart, his support made this work a reality. My second supervisor Prof. Patrick Weke for his patience, advice and guidance during and before this work.

I also thank Stanley, Jonathan, Anne and Rose-all of Mt. Kenya University- for their friendship during and before this work.

Lastly, I thank my parents, my wife, daughter, brothers and sister, the entire family and friends for their moral support and encouragement during my studies.

God bless you all.

Table of Contents

List of Tables

List of Figures

Abstract

The global Vector autoregressive (GVAR) approach directly models the interlinkages using trade- weighted observable macroeconomic aggregates and financial variables and allows for interdependencies at a variety of levels in a transparent manner that can be empirically evaluated. However, with a modeling task of this size, it would be surprising if a single model were universally preferred over any other. Recognizing that a broader set of models might be needed to tackle the problem, we apply Bayesian Model Averaging (BMA) with a Minnesota prior to appropriate GVAR models-for five East African countries on a quarterly basis from the year 2000 to the year 2013-in order to arrive at better overall forecasts. The results obtained indicate that the GVAR models are dynamically stable. The results also indicate that the GVAR methodology is valid since the weak exogeneity is rejected for most of the foreign variables and that the trade weights are 'granular'. The study also uses the AIC, SBC and log-likelihood selection criterion in order to select the lag orders of the domestic and foreign variables respectively. Our results also show that, taking international linkages into account improves forecasts in inflation and exchange rates while for interest rates forecasts of the univariate benchmark models remain difficult to beat. Moreover, there is a notable outcome whereby the standard GVAR model employed in our literature does not perform well than the simple models.

CHAPTER 1

INTRODUCTION

This chapter outlines a brief history of BMA and GVAR. It also outlines the application of BMA to macroeconomic models. In addition it states the statement problem, the study objectives and research questions. It also outlines the significance, scope and assumptions of the study and the definition of terms used. Organization of the study is also outlined in this chapter.

1.2 Background Study

Financial systems in advanced and emerging economies have undergone remarkable changes over time. Cross border ownership of assets and investments has increased, revealing important benefits and new risks associated with financial integration. Cross border ownership of assets and investment exposes financial institutions such as banks to macroeconomic, financial and price fluctuations in the countries where they hold positions. Increasingly complex linkages across market segments and borders make transmission of shocks in the international economy and the pattern of risk dispersion more opaque, creating uncertainty for agents and policy makers about where the ultimate risks lie. Adjusting well to shocks means having a model that is not only resilient but also forecasts more efficiently across sectors and across firms.

In order to shed light on linkages across markets, firms or countries, it is crucial to look into the cross-country or cross-market transmission of financial shocks. This will also account for regional interdependencies. In this perspective, country-specific models are estimated, where the domestic macro-economic variables are related to the corresponding foreign variables. In this way each country is potentially affected by the developments in other countries. Various models are used to model these regional interdependencies. These models are usually referred to as multiple equation models. This study will only review the Vector Autoregressive (VAR) and Vector Error Correction (VEC) models in this class of multiple equation models.

1.3 Vector Autoregressive (VAR) models

Considering a column vector of k different variables

$$\mathbf{y_t} = [y_{1t}, y_{2t}, \dots y_{kt}]' \tag{1.1}$$

Modeling this in terms of past values of the vector; the result is a vector auto regression (VAR) model. The VAR (p) process is represented as

$$\mathbf{y_t} = \mathbf{v} + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \mathbf{u_t} \tag{1.2}$$

Where A_1 through $A_{\mathbf{p}}$ are $k \times k$ matrices of coefficients, $\mathbf{v}$ is a $k \times 1$ vector of constants,

$\mathbf{u_t}$ is a vector of white noise or innovation process

Assumptions

i) $E(u_t) = 0$ for all t

ii) $E(u_t u_s) = \begin{cases} \sum_u \text{ for } s = t \\ 0 \text{ for } s \neq t \end{cases}$

The covariance matrix $\sum_u$ is assumed to be $\text{non} - \text{singular}$.

$$\mathbf{y_t} = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{1.3}$$

A VAR process thus describes a system in which each variable is a function of its own lag and the lag of other variables in the system.

1.4 Global Vector Autoregressive (GVAR) Model

The recent financial crisis has shown yet again how world economies are globally interlinked, via a complex net of transmission channels. When it comes, however, to build econometric frameworks aimed at analyzing such interlinkages, modelers are faced with what is called the "curse of dimensionality": there are far too many parameters to be estimated with respect to the available observations, Mauro and Pesaran (2013).

The GVAR model is a VAR based model of the global economy which offers a solution to this problem. The basic model is composed of large number of country specific VAR models, comprising domestic, foreign and purely global variables.

In this work the interdependencies among countries or markets is investigated. The GVAR modeling will be used since it is a multicountry framework which allows the investigation of interdependencies among countries Dees et al (2007).

Different GVAR models are obtained but given the size of the modeling task, the problem is that of selecting a single model-as it is a common practice in empirical research-from which conclusions can be drawn on this model acting as if the model chosen is the true one. This procedure tends to understate the real uncertainty and thus conclusions might not be sufficiently conservative. To overcome this problem all the candidate models are averaged and the weighted average of all the estimates computed. This is known as the model averaging approach. Frequentist model averaging (FMA) and Bayesian Model Averaging (BMA) are the main approaches to model averaging in literature.

1.5 Vector Error Correction (VEC) Models

When the variables in VAR are integrated of order one or more, unrestricted estimation is subject to hazards of regressions involving non-stationary variables. The presence of non-stationary variables raises the possibility of cointegrating relations. The relevant procedure then consists of three steps. First, determine the cointegrating rank, that is, the number of cointegrating relations. Secondly, estimate the matrix of cointegrating vectors, β, and the associated weighting matrix, α. This step amounts to determining the factorization π=αβ. The third step involves estimating the VAR while incorporating the cointegrating relations from the previous step.

Several methods of tackling these steps exist as discussed in Johnston and Dinardo (1996).

There is considerable evidence that these models when used in economic and financial forecasting are often unstable and subject to structural breaks, despite their success relative to other alternatives Stock and Watson (1996). Other studies that document the instability of these models include Alogoskoufis and Smith (1991) and Garcia and Perron (1996).

1.6Frequentist Model Averaging (FMA)

Let us take the linear model in matrix form as shown below

$$\mathbf{y} = \mathrm{X_A}\beta + \mathrm{X_B}\Upsilon + \mathcal{E} \tag{1.4}$$

Where $\mathbf{y},$ and $\mathcal{E}$are $\mathrm{T} \times 1$ vectors of the dependent variable, the treatment variable of interest and the random shocks respectively,$\mathrm{X_A}$ is a $\mathrm{T} \times \mathrm{q}$, $X_{\boldsymbol{B}}$is a $\mathrm{T} \times \mathrm{q}$ matrix of doubtful control variables that may or may not be included in the model and β andΥ $(\mathrm{q} \times 1)$ contain the parameters to be estimated. T represents the number of observations in the sample. If the components of $\boldsymbol{\Upsilon} = (\Upsilon_1 , \Upsilon_2 \dots , \Upsilon_q)'$ to be zeros, there are a total of 2^q candidate models to be estimated. The coefficient of interest is now β. Let $\beta_M$be the estimator of β under the candidate model M with$M \in (M_1, M_2, \dots, M_{2^q})$. The most common approach in research is to take the selected model as given and base the inference on this single estimate β_M while the actual estimator is

$$\beta = \begin{cases} \beta_{M_1} \text{ if first model is selected} \\ \beta_{M_2} \text{if second model is selected} \\ \vdots \\ \beta_{M_{2^q}} \text{if } 2^q - \text{th model is selected} \end{cases} \tag{1.5}$$

The above estimator can also be re-written as

$$\beta = \sum_{j=1}^{2^q} \omega_{M_j} \beta_{M_j} \dots \dots \dots \dots \text{(FMA i)} \tag{1.6}$$

Where

$$\omega_{M_j} = \begin{cases} 1 \text{ if the candidate model } M_j \text{is selected} \\ 0 \qquad\qquad\qquad\qquad \text{otherwise} \end{cases}$$

If we restrict $0 \leq \omega_{M_j} \leq 1$ and $\sum_{j=1}^{2^q} \omega_{M_j} = 1$ in FMA i thenβ is the FMA estimator ofβ.

Another method used in model selection is the stepwise methods for model selection in logistic regressions commonly applied in medical research to identify important risk factors of mobidity and mortality outcomes. The stepwise method may be implemented automatically or manually Miller (1990). The automated technique consists of sequentially adding and deleting variables guided by approximate asymptotic ratio test. It leads to the construction of a single “optimal” model. In this model variables without statistically significant associations to the outcomes are excluded.

The model combination or selection techniques discussed above leave policy makers with a wide range of choices, such as what the set of admissible models should be selected and what averaging scheme should be used to combine the forecasts from each model Pesaran et al. (2008). To fix this problem and that of uncertainty, this work will use Bayes rule to arrive at BMA.

1.7 A Bayesian Approach to Estimating a Linear Regression Model

Consider the task of estimating the following regression model

$$\mathbf{Y_t} = \mathrm{AX_t} + \mathbf{v_t} \tag{1.7}$$

Where $\mathbf{v_t} \sim \mathrm{N}(0, \sigma^2)$ and $\mathbf{Y}_\mathrm{t}$ is a $\mathrm{T} \times 1$ matrix of the dependent variable, $\mathrm{X_t}$ is a $\mathrm{T} \times \mathrm{K}$ matrix of the independent variables and deterministic terms. In this case we are concerned with estimating the $\mathrm{K} \times 1$ vector of coefficients B and the variance of the error term σ^2.

A classical econometrician solves this problem by estimating and maximizing a likelihood function resulting in the end the estimator for matrix A and the estimated variance errors. So the classical econometrician is based on utilization of all data on Y_t and X_t and writes down the likelihood function of the model

$$\mathrm{F}(\mathrm{Y_t}|\mathrm{B}, \sigma^2) = (2\pi\sigma^2)^{-\mathrm{T}/2} \exp\left(-\frac{(\mathrm{Y_t} - \mathrm{BX_t})'(\mathrm{Y_t} - \mathrm{BX_t})}{2\sigma^2}\right) \tag{1.8}$$

and obtains estimates $\widehat{\mathrm{A}}$ and $\widehat{\sigma}^2$ by maximizing the likelihood.

By following the step above the classical econometrician obtains the familiar OLS estimator for the coefficients $\widehat{\mathrm{A}}_{\mathrm{OLS}} = (\mathrm{X_t}'\mathrm{X_t})^{-1}(\mathrm{X_t}'\mathrm{Y_t})$ and the biased maximum likelihood estimator for the error variance $\widehat{\sigma}^2 = \frac{\mathrm{v_t}'\mathrm{v_t}}{\mathrm{T}}$.

Bayesian econometrics on the other hand offers the following solution that involves three steps as explained in in Blake and Mumtaz (2012)

*Step 1.*The researcher forms a prior belief about the parameters to be estimated. The prior belief usually represents information the researcher has about A and σ^2 which is not derived using the data Y_t and X_t. These prior beliefs may have been formed through past experience or by examining studies (estimating similar models) using other datasets. For instance, the prior on coefficients A is expressed as

$$P(A) \sim N(A_0, \Sigma_0) \tag{1.9}$$

Where the mean A_0 represents the actual beliefs about the elements of A.

*Step 2.*The researcher collects data on Y_t and X_t and write down the likelihood function of the model

$$F(Y_t|A, \sigma^2) = (2\pi\sigma^2)^{-T/2} \exp\left(-\frac{(Y_t - AX_t)'(Y_t - AX_t)}{2\sigma^2}\right) \tag{1.10}$$

This step is identical to the approach of the classical econometrician and represents the information about the model parameters contained in the data.

*Step 3.*The researcher updates her prior beliefs on model parameters in step 1 by incorporating the steps contained in the data. The researcher combines the prior distribution $P(A, \sigma^2)$ and the likelihood function $F(Y_t|A, \sigma^2)$ to obtain the posterior distribution $H(A, \sigma^2|Y_t)$. This steps applies Bayes law to obtain the posterior

$$H(A, \sigma^2|Y_t) = \frac{F(Y_t|A, \sigma^2) \times P(A, \sigma^2)}{F(Y)} \tag{1.11}$$

The marginal likelihood$F(Y)$ is a scalar and will not have any operational significance as far as estimation is concerned though it is crucial for model comparison. Therefore the Bayes law can be written as

$$H(A, \sigma^2 | Y_t) \propto F(Y_t | A, \sigma^2) \times P(A, \sigma^2) \tag{1.12}$$

Equation (1.12) above states that the posterior distribution is proportional to the likelihood times prior. Blake and Mumtaz (2012), argue that the Bayes law in equation (1.11) can be easily derived by considering the joint density of the data and parameters $A, \sigma^2, G(Y_t, A, \sigma^2)$ and observing that it can be factored in two ways

$$G(Y_t, A, \sigma^2) = F(Y_t) \times H(A, \sigma^2 | Y_t) = F(Y_t | A, \sigma^2) \times P(A, \sigma^2) \tag{1.13}$$

To understand in details the basics applied in Bayesian econometrics, we will consider the derivation of the posterior distribution in three circumstances. First we consider estimating the posterior distribution of A under the assumption that σ^2 is known. Next we consider estimating the posterior distribution of A under the assumption that A is known and finally we consider the general case when both sets of the parameters are unknown.

Case 1: The posterior distribution of A assuming σ^2 is known.

We consider a scenario where the econometrician wants to estimate $\boldsymbol{\beta}$ in equation (1.7) but knows the value of σ^2 already. As discussed above, the posterior distribution is derived using three steps.

Step 1: Setting the prior

A prior normal is set for **β**. A normally distributed prior $\mathrm{P(A)\sim N(A_0, \sum_0)}$ for the coefficients is a conjugate prior. That is, when this prior is combined with the likelihood function this result in a posterior with the same distribution as the prior. The prior distribution is given by

$$(2\pi)^{-K/2} | \textstyle\sum_0 |^{-\frac{1}{2}} \exp[-0.5(A - A_0)' \textstyle\sum_0^{-1}(A - A_0)] \tag{1.14}$$

Step 2: Setting up the likelihood

In the second step, the researcher collects the data and defines the likelihood as

$$F(Y_t|A, \sigma^2) = (2\pi\sigma^2)^{-T/2} \exp\left(-\frac{(Y_t - AX_t)'(Y_t - AX_t)}{2\sigma^2}\right) \tag{1.15}$$

The first term in the right hand side of (1.15) is also a constant since σ^2 is assumed to be known.

Step 3: Calculating the posterior

The posterior distribution is proportional to the likelihood times prior. Therefore, to find the posterior distribution for *A* (conditional on knowing σ^2) the researcher multiplies equation (1.14) and (1.15) to obtain

$$H(A|\sigma^2, Y_t) \propto \exp[-0.5(A - A_0)' \textstyle\sum_0^{-1}(A - A_0)] \times \exp\left(-\frac{(Y_t - AX_t)'(Y_t - AX_t)}{2\sigma^2}\right) \tag{1.16}$$

Equation (1.16) is simply a product of two normal distributions and the result is also a normal distribution. Hence the posterior distribution *A* conditional on σ^2 is given by:

$$H(A|\sigma^2, Y_t) \sim N(M^*, V^*) \tag{1.17}$$

Koop (2003) uses the following formulae for the mean and variance of this normal distribution:

$$M^* = \left(\Sigma_0^{-1} + \frac{1}{\sigma^2} X_t{}' X_t\right)^{-1} \left(\Sigma_0^{-1} A_0 + \frac{1}{\sigma^2} X_t{}' Y_t\right)$$

$$V^* = \left(\Sigma_0^{-1} + \frac{1}{\sigma^2} X_t{}' X_t\right)^{-1} \tag{1.18}$$

The final term $X_t{}'Y_t$ in M^* can be re-written as $X_t{}'X_tA_{OLS}$where $A_{OLS} = (X_t{}'X_t)^{-1}X_t{}'Y_t$. That is

$$M^* = \left(\Sigma_0^{-1} + \frac{1}{\sigma^2} X_t{}' X_t\right)^{-1} \left(\Sigma_0^{-1} A_0 + \frac{1}{\sigma^2} X_t{}' X_t A_{OLS}\right) \tag{1.19}$$

Case 2: The posterior distribution of σ^2 assuming *A* is known

We consider the estimation of σ^2 in equation (1.7) assuming that the value of *A* is known. The derivation of the posterior distribution of σ^2 proceeds in exactly the same three steps.

Step 1: setting the prior

The normal distribution allows for negative numbers and is therefore not appropriate as a prior distribution for σ^2. A conjugate prior for σ^2 is the inverse Gamma distribution or equivalently a conjugate prior for $\frac{1}{\sigma^2}$ is the Gamma distribution.

Definition 1.1 (Gamma Distribution):suppose we have T iid numbers from the normal distribution $v_t \sim N(0, \frac{1}{\theta})$. If we calculate the sum of squares of $W = \sum_{t=1}^{T} v_t^2$, then W is distributed as Gamma distribution with T degrees of freedom and a scale parameter θ

$$W \sim \Gamma\left(\frac{T}{2}, \frac{\theta}{2}\right) \tag{1.20}$$

The probability density function for the Gamma distribution has a simple form and is given by

$$g(W) \propto W^{\frac{T}{2}-1} \exp\left(\frac{-W\theta}{2}\right) \tag{1.21}$$

Where the mean of the distribution is defined as $\mathrm{E(W)} = \frac{\mathrm{T}}{\theta}$.

We set a Gamma prior for $\frac{1}{\sigma^2}$. That is $\mathrm{p}(\frac{1}{\sigma^2}) \sim \Gamma(\frac{T_0}{2}, \frac{\theta_0}{2})$, where T_0 denotes the prior degrees of freedom and θ_0 denotes the prior scale parameter. From equation (1.21) the prior density has the following form

$$\frac{1}{\sigma^2}^{\frac{T_0}{2}-1} \exp\left(\frac{-\theta_0}{2\sigma^2}\right) \qquad (1.22)$$

Step 2: Setting up the likelihood function

In the second step, the researcher collects the data and forms the likelihood function

$$\mathrm{F}(\mathrm{Y_t}|\mathrm{A}, \sigma^2) = (2\pi\sigma^2)^{-\mathrm{T}/2} \exp\left(-\frac{(\mathrm{Y_t}-\mathrm{AX_t})\prime(\mathrm{Y_t}-\mathrm{AX_t})}{2\sigma^2}\right) \qquad (1.23)$$

Step 3: calculating the posterior distribution

To calculate the posterior distribution of $\frac{1}{\sigma^2}$ (conditional on *A*) we multiply the prior distribution in equation (1.22) and the likelihood function in (1.23) to obtain

$$\mathrm{H}\left(\frac{1}{\sigma^2}\middle|\mathrm{A}, \mathrm{Y_t}\right) \propto \frac{1}{\sigma^2}^{\frac{T_0}{2}-1} \exp\left(\frac{-\theta_0}{2\sigma^2}\right) \times (\sigma^2)^{-\mathrm{T}/2} \exp\left(-\frac{(\mathrm{Y_t}-\mathrm{AX_t})\prime(\mathrm{Y_t}-\mathrm{AX_t})}{2\sigma^2}\right) \qquad (1.24)$$

The expression in (1.24) can be re-arranged to form (1.25) below

$$\mathrm{H}\left(\frac{1}{\sigma^2}\middle|\mathrm{A}, \mathrm{Y_t}\right) \propto \frac{1}{\sigma^2}^{\frac{T_0}{2}-1+\frac{\mathrm{T}}{2}} \exp\left(\frac{-1}{2\sigma^2}\left[\theta_0 + (\mathrm{Y_t} - \mathrm{AX_t})'(\mathrm{Y_t} - \mathrm{AX_t})\right]\right) \qquad (1.25)$$

Or

$$\mathrm{H}\left(\frac{1}{\sigma^2}\middle|\mathrm{A}, \mathrm{Y_t}\right) \propto \frac{1}{\sigma^2}^{\frac{T_1}{2}-1} \exp\left(\frac{-\theta_1}{\sigma^2}\right) \qquad (1.26)$$

The resulting conditional posterior for $\frac{1}{\sigma^2}$ in (1.24) is a Gamma distribution with $T_1 = \frac{T_0+T}{2}$ and scale parameter $\theta_1 = \frac{\theta_0+(Y_t-AX_t)'(Y_t-AX_t)}{2}$.

Case 3: the posterior distribution of $\boldsymbol{\sigma^2}$ and *A*

We now turn to the empirically relevant case when both the coefficient vector *A* and the variance $\frac{1}{\sigma^2}$ in equation (1.7) are unknown. We proceed in exactly the same three steps

Step 1: Setting the prior

We set the joint prior density for

$$p\left(A, \frac{1}{\sigma^2}\right) = p\left(\frac{1}{\sigma^2}\right) \times p\left(A \middle| \frac{1}{\sigma^2}\right) \tag{1.27}$$

where $p\left(A \middle| \frac{1}{\sigma^2}\right) \sim N(A_0, \sigma^2 \sum_0)$ and $p\left(\frac{1}{\sigma^2}\right) \sim \Gamma(\frac{T_0}{2}, \frac{\theta_0}{2})$. That is: $\frac{1}{\sigma^2}^{\frac{T_0}{2}-1} \exp\left(\frac{-\theta_0}{2\sigma^2}\right)$and $p\left(A \middle| \frac{1}{\sigma^2}\right) = (2\pi)^{-K/2} |\sigma^2 \sum_0|^{-1/2} \exp[-0.5(A - A_0)'(\sigma^2 \sum_0)^{-1} (A - A_0)]$.

Step 2: setting up the likelihood function

The likelihood function is given by

$$F(Y_t|A, \sigma^2) = (2\pi\sigma^2)^{-T/2} \exp\left(-\frac{(Y_t-AX_t)'(Y_t-AX_t)}{2\sigma^2}\right) \tag{1.28}$$

Step 3: calculating the posterior

The joint posterior distribution of *A* and the variance $\frac{1}{\sigma^2}$ is obtained by combining (1.27) and (1.28) to get

$$H\left(\frac{1}{\sigma^2} \middle| A, Y_t\right) \propto p\left(A, \frac{1}{\sigma^2}\right) \times F(Y_t|A, \sigma^2) \tag{1.29}$$

Note that equation (1.29) is a joint posterior distribution involving $\frac{1}{\sigma^2}$and A. Its form is more complicated than the conditional distributions for A and $\frac{1}{\sigma^2}$ shown in case 1 and 2.

To proceed further in terms of inference, the researcher has to 'isolate' the component of the posterior relevant to A or $\frac{1}{\sigma^2}$. For instance, to conduct inference about A, the researcher should derive the marginal posterior distribution for A. Similarly, inference on $\frac{1}{\sigma^2}$ is based on the marginal posterior distribution for $\frac{1}{\sigma^2}$. The marginal posterior for A is defined as

$$\mathrm{H}(\mathrm{A}|\mathrm{Y_t}) = \int_0^\infty \mathrm{H}\left(\frac{1}{\sigma^2}\middle|\mathrm{A}, \mathrm{Y_t}\right) \mathrm{d}\,\frac{1}{\sigma^2} \tag{1.30}$$

The marginal posterior for $\frac{1}{\sigma^2}$ is given by

$$\mathrm{H}\left(\frac{1}{\sigma^2}\middle|\mathrm{Y_t}\right) = \int_0^\infty \mathrm{H}\left(\frac{1}{\sigma^2}\middle|\mathrm{B}, \mathrm{Y_t}\right) \mathrm{dA} \tag{1.31}$$

In the case of a simple linear regression model under the natural conjugate prior, analytical results for these integrals are available. An in depth description of these analytical results can be found in Koop (2003) chapter 2.

However, for the linear regression model with other prior distributions (for example where the prior for the coefficients are set independently from the prior for the variance) analytical derivation of the joint posterior and then the marginal posterior distribution is not possible. Similarly, in more complex models with a larger set of unknown parameters (i.e. models that may be more useful for inference and forecasting) these analytical results may be difficult to obtain. This may happen if the form of the joint posterior is unknown or is too complex for analytical integration.

This need for analytical integration to calculate the marginal posterior distribution is the main stumbling block of Bayesian analysis making it difficult for applied researchers. In order to deal with this difficulty/challenge, Gibbs sampling simulation method which greatly simplifies the integration step discussed above and make it possible to easily extend Bayesian analysis to our model, can be applied.

1.8 Statement of the Problem

The recent financial crisis raises important issues about the transmission of financial shocks across borders. Therefore the need to link specific country models to others by a set of country specific foreign variables Galesi and Sgherri (2009).

Popular autoregressive models used in economic and financial forecasting can capture these linkages and interdependencies but are often unstable and subject to structural breaks Pesaran et al, (2008). Structural instability is a key factor in poor forecast performance, Clements and Hendry (1998). It is important to note that the conditional models, for instance, each country specific VAR models in the GVAR are structurally stable but the unconditional models used to generate forecasts could be subject to structural breaks.

Data analysts typically select a model from some class of models and then proceed as if the selected model had generated the data. This approach ignores uncertainty in model selection, leading to overconfident inferences and decisions that are more risky than one thinks they could be. In reality, some or all of the models under consideration could be subject to structural breaks and different choices of estimation samples might be warranted. When there is little certainty about which GVAR model is the right one and in addition if models are subject to structural breaks, model averaging techniques come into play.

1.9 Objectives of the Study

1.9.1 General objective

To model financial interlinkages among five East African countries by BMA to GVAR.

1.9.2 Specific Objectives

1. To specify and estimate the VAR models for Kenya, Uganda, Tanzania, Rwanda and Burundi
2. To connect the parameter estimates of the country specific VAR models
3. To determine the dynamic properties of the GVAR models obtained.
4. To estimate forecasts in Inflation rate, exchange rates and interest rates
5. To evaluate the performance of the averaged models over the random walk and autoregressive models

1.11 Significance of the Study

This study will be important to various users of the research information that will be generated. The various users will include financial institutions, investors, policy makers and academia.

1.11.1 Financial Institutions

This research will be useful to financial institutions as it will address the issue of choosing appropriate modeling technique that will address the complex linkages across market segments or borders. In this way, financial institutions will be in a position to forecast macroeconomic risks that could affect the financial institutions' portfolios. The financial institutions can apply the GVAR models for applications in credit-risk stress testing and forecast market uncertainties.

1.11.2 Investors

The sizable cross-border financial linkages across countries require models that will highlight the vulnerabilities on foreign investments/funding. An appropriate model as is the basis of this study must give the sense of the magnitude and distribution of exposures to foreign spillovers/shocks to investors.

1.11.3 Policy Makers

Since the GVAR model is a multi-country framework, it will enable a country's policy makers to assess and analyze the interdependencies among countries. It will allow the policy makers to use this global macro-econometric modeling approach in analyzing the regional propagation of shocks.

1.11.4 Academia

This research work will add useful information to the existing knowledge. It will also arouse deep thoughts and interest on the subject matter for further research.

1.12 Scope of the Study

This study will consider the following East Africa countries: Kenya, Tanzania, Uganda, Rwanda and Burundi. The period of study will be from 2000 to 2013 on a quarterly basis.

1.13 Assumption of the study

The VAR model approach assumes that the domestic variables are related to country specific variables in a constant manner.

1.14 Organization of the Study

This study is organized as follows; chapter two offers literature review, while chapter three is the methodology of the GVAR model and estimation procedure of the parameters in a GVAR model. Chapter four offers the Bayesian model averaging implementation in a GVAR model while data, empirical results and discussion of these results is offered in chapter five. Finally, chapter 6 includes summary of findings, recommendations and conclusion.

1.15 Conclusion

This chapter has introduced the basics behind a Global Vector Autoregressive model and model averaging techniques. Moreover, specific objectives, research questions and significance of the study have been discussed in this section too.

CHAPTER 2

Literature Review

2.1 Introduction

A GVAR model is a model that is composed of individual country vector Autoregressive models in which the core domestic variables are related to country-specific foreign variables. It is a large model which allows for a high degree of interdependence and dynamics. On the other hand Bayesian Model Averaging provides a coherent and effective approach for incorporating model uncertainty predictions and inferences based on a set of models, rather than a single model. Bayesian Model Averaging ensures that each model in the mixture distribution contributes proportionally to the support it receives from the observed data.

This chapter therefore reviews related literature based on the concept of interlinkages across economies. The literature related to previous work by other researchers concerning the GVAR model and Bayesian GVAR implementation has also been reviewed in this section.

In 1818, Laplacederived and compared the properties of two estimators, one being least squares and the other a kind of a weighted median. He also analyzed the joint distribution of the two and proposed a combining formula that resulted in a better estimator than either. A brief description of Laplace's work can be found in the work of Stigler (1973).

Apart from Laplace, other early work on combining multiple estimates came from the statistical literature of Edgerton and Kolbe (1936) who proposed to combine different estimates in such a way that the combining weights result from minimizing the sum of squares of the differences of the scores.

Horst (1938) derived a formula for combining multiple measures in which the criterion was obtaining maximum separation among the individual population members. Halperin (1961) provided a minimum squared-error combination of estimates.

By the 1970s the ideas of combining estimates was present in several studies in the field of statistics. This is evidenced by the work of de Finetti (1972), Davis (1979) and Geisser and Eddy (1979).

In the forecasting literature, many papers about combining different forecasts were generated in the 1960s and the 1970s. Barnard (1963), Bates and Granger (1969) provide a lot of literature on forecast combination. This means that up to date the idea of combining forecasts is well established in the available literature. Timmermann (2006) provides a good overview of recent advances in this literature.

The papers of Hjort and Claeskens (2003) and Hansen (2007) use the Frequentist Model Averaging (FMA) approach in forecasting combination. They use the FMA approach in model averaging to address the issue of model uncertainty. Their work is an extension of the traditional approach to model uncertainty which focused on model selection rather than model averaging.

In BMA approaches, Geisser (1965), Roberts (1965) and Geisel (1973) have the earliest Bayesian approaches to combining estimates. However, the work of Leamer (1978) presents the first comprehensive description of the basis paradigm for BMA. Leamer pointed out the fundamental idea that BMA accounts for the uncertainty involved in selecting the model and published a book on this work. With a few exceptions such as Moulton (1991), the BMA approach was ignored in economic applications until the late 1990s and 2000s, Benito and de Espania (2013). The reason why BMA approach is gaining popularity is because of the presence of more powerful computers and dramatic increases in numerical methods such as Monte Carlo Markov-Chain model composition (MC^3) which enabled researchers to overcome the troubles encountered while implementing BMA. MC^3 would explore large model spaces in more sensible ways.

The work of Hoeting et al (1999) gives a tutorial on BMA. In their work Hoeting et al present several difficulties encountered while implementing BMA and discuss several solutions to these implementation difficulties. However, their work concentrated on the theoretical implementation of Bayesian model averaging on linear models which in most cases are not applicable in real cases.

The application of BMA in economics and financial models received attention in the late 1990s when theoretical developments and computational power enabled researchers to overcome the difficulties to implementing BMA.

Draper (1995), Chatfield (1995) and Kass and Raftery (1995) all review BMA and the cost of ignoring model uncertainty. George (1999) reviews Bayesian model selection and discusses BMA in the context of decision theory. Raftery et al (1997) offer two alternative approaches to BMA for linear regression models. Raftery et al describe the “Occam’s window” which indicates a small set of models over which a model average can be computed. Secondly, Raftery et al describe a Markov chain Monte-Carlo approach that directly approximates the exact solution. The results of these two approaches in the presence of model uncertainty is that they both provide better predictive performance than any single model that might reasonably have been selected. Their work only concentrates on two procedures that account for model uncertainty in variable selection for linear regression models but do not address the uncertainty involved in the identification of outliers and in the choice of transformations in regression.

To broaden the flexibility of the procedures addressed in Raftery et al, Hoeting et al (1999) have extended BMA to include transformation selection and outlier identification.

In their article, Bauwens and Rombouts (2007) consider the estimation of a large number of GARCH models of the order of several hundreds. They identify common structures in the volatility dynamics of the univariate time series. They classify the series in an unknown number of clusters, whereby within each cluster, the series share the same model and the same parameters. However, they use normal densities for ease of illustration. Flexible densities such as the student's t-distribution and skew t distributions could probably be used.

Other research work in econometrics considers univariate models which do not take into account the increasing interdependencies that exist between countries which mean dealing with very high dimensional systems. This is because most univariate models can only deal with a relatively small number of variables, Pesaran and Smith (2006).

The GVAR approach advanced in Pesaran et al. (2004) provides a simple solution where country specific VAR models are estimated relating a vector of domestic variables to their foreign counterparts and are then consistently combined to form a Global VAR.

However most work in global macro models as is seen in the work of Pesaran et al. (2004), Pesaran and Smith (2006), Dees et al.(2007), Galesi and Sgherri (2009), Bache et al (2009) and others consider the selection of a single GVAR model and drawing conclusions as if the model selected were the true model.

The work of Pesaran et al (2008) states that the conditional country specific models in the GVAR models are structurally stable but the unconditional model used to generate forecasts could be subject to structural breaks. For this reason, Pesaran et al consider a situation whereby some or all of the models under consideration could be subject to structural breaks and different choices of estimation samples might be warranted. With that in mind Pesaran et al averaged each model over different sampling windows allowing for both model and estimation window uncertainty. They apply BMA to GVAR by specifying model weights, namely the prior probability of a model and the prior probability of a model's coefficients. They examine and evaluate some of the choices in the context of a global VAR covering 33 countries grouped into 26 regions. They then address the issue of model and window uncertainties by averaging GVAR forecasts over different model specifications and estimation windows. Their results tend to outperform forecasts based on individual models. However Pesaran et al did not use GVAR to examine the context to which forecasts for an individual economy are likely to be enhanced by allowing for global interactions. Their work does not address the foreign inter-linkages of fast growing economies such as China and India.

However, there has been relatively few little research work in the field of macroeconomic forecasting using large macroeconomic models, Pesaran et al (2008) is a notable example. Hence, the main purpose of this research is to draw on ideas from the statistical literature on Bayesian model averaging and apply them to the problem of macroeconomic forecasting.

2.2 Research Gap

The standard approach in global macro models in the relevant literature has been to choose a single model and present empirical results based on this model. There are two potential problems with this approach. First, when the researcher selects a single model, statistical evidence from other plausible models is ignored. The second problem with this approach is that evaluating an information criterion for every model can be computationally prohibitive since, if K is the number of factors and models are defined by the inclusion or exclusion of each factor, then 2^K possible models exist. For a large value of K, say $K > 20$ or so, direct computation of an information criterion for every model is cumbersome. This calls for factor analysis which has some drawbacks Koop and Potter (2003). Suffice it to note that there are theoretical and practical reasons for basing inference not on a single model, but on averages across models. This has motivated an explosion of research work which uses BMA in many fields of applied statistics and econometrics. However, there has been little work on BMA to global macro models, specifically BMA to global VAR. In this contribution we combine the virtues of the standard GVAR model and Bayesian literature and propose a Bayesian GVAR (B-GVAR) model. Akin to the GVAR framework we assume that links among economies are determined exogenously, while we borrow strength from the Bayesian literature in estimating the individual country models. This allows us to keep the virtues of the GVAR framework offering a coherent way for policy and counterfactual analysis. Our model includes standard variables that are typically employed in small-country VARs such as inflation, exchange rates, interest rates and the oil price as a global control variable.

2.3 Conclusion

This chapter is a brief review of the previous work by other researchers on GVAR modeling and model averaging techniques. On the research gap, the researcher has criticized the selection of a single model using model selection criterion and presenting empirical results as if the selected model is the true model. Moreover, previous work has indicated clearly that there is little work on BMA to global macro models, specifically BMA to global VAR hence the main purpose of this research.

CHAPTER 3

The GVAR Model

3.1 Introduction

In order to assess the importance of financial interlinkages among East African countries, we build a GVAR model, following Pesaran et al (2004) and Dees et al (2007).

The GVAR model is a multicountry framework which allows the investigation of interlinkages among countries. It is generally composed by several country economies modeled by corresponding Vector Autoregressive (VAR) models.

This chapter outlines the estimation procedure for the global vector autoregressive model which is the standard model employed in our literature.

3.2 Vector Autoregressive (VAR) Models

In this section, we will briefly discuss Vector Autoregression commonly referred to as VAR. VAR constitutes one of the easiest types of multivariate time series models to estimate and they have been widely used in economics in recent years. The VAR model is a general framework to describe the dynamic interrelationship between stationary variables.

3.2.1 The VAR Model Specification

Consider a column vector of k-different variables $\boldsymbol{y_t} = [y_{1t}, y_{2t}, \dots, y_{kt}]'$ and model this in terms of past values of the vector. The result is a Vector Autoregressive order p or a VAR (p) process written as

$$\boldsymbol{y_t} = \boldsymbol{v} + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \boldsymbol{u_t} \tag{3.1}$$

Where $\mathrm{A_1}$ through $\mathrm{A_p}$ are $\mathrm{k} \times \mathrm{k}$ matrices of coefficients, $\boldsymbol{v}$ is a $\mathrm{k} \times 1$ vector of constants, $\mathbf{u_t}$ is a vector of white noise or innovation process

Assumptions

iii) $\mathrm{E(u_t) = 0}$ for all t

iv) $\mathrm{E(u_t u_s)} = \begin{cases} \sum_\mathrm{u} \text{ for } \mathrm{s = t} \\ 0 \text{ for } \mathrm{s \neq t} \end{cases}$

The covariance matrix $\sum_\mathrm{u}$ is assumed to be non − singular.

The VAR (p) process in (3.1) can also be written in lag operator notation, the lag L operator is defined such that $Ly_t = y_{t-1}$, that is, it lags (shifts back) the index by one period. Using this operator equation (3.1) can be written as

$$\boldsymbol{y_t} = \boldsymbol{v} + \left(A_1 L + A_2 L^2 + \cdots + A_p L^p\right) y_t + \boldsymbol{u_t}$$

Or

$$A(L)\boldsymbol{y_t} = \boldsymbol{v} + \boldsymbol{u_t}$$

Where

$$A(L) = I_K - A_1 L - A_2 L^2 - \cdots - A_p L^p$$

3.2.2 Properties of a VAR (p) Model

3.2.2.1 Stability Condition for a VAR (p) Process

The VAR (p) process in equation (3.1) is stable if its reverse characteristic polynomial has no roots in and on the complex unit circle, that is, y_t is stable if

$$\det(I_K - A_1 z - A_2 z^2 - \cdots - A_p z^p) \neq 0 \; for \; |z| \leq 1 \tag{3.2}$$

Proposition 3.1: Stationarity condition

A stable VAR (p) process y_t is stationary for $t = 0, \pm 1, \pm 2, \ldots$

Because stability implies stationarity, the stability condition stated in (3.2) is often referred to as stationarity in time series. The converse of proposition 3.1 is not true. In other words, an unstable process is not necessarily non-stationary.

3.3 Estimation of a VAR (p) Model

In the VAR (p) in equation (3.1) where $\mathrm{u_t} \sim$ IID $(0, \Omega)$, if $\mathrm{y_{it}}$ denotes the $\mathrm{i^{th}}$ element of $\mathrm{y_t}$ and $\mathrm{A_{j,ki}}$ denotes the $\mathrm{ki^{th}}$ element of A_j, the i^{th} row of equation (3.1) can be written as

$$\mathrm{y_{it}} = \mathrm{m_i} + \textstyle\sum_{\mathrm{k=1}}^{\mathrm{m}} \sum_{\mathrm{j=1}}^{\mathrm{p}} \mathrm{A_{ki,j}} \mathrm{y_{k,t-j}} + \mathrm{u_{it}} \tag{3.3}$$

This is just a linear regression in which $\mathrm{y_{it}}$depend on a constant term and lags 1 through p of all of the m variables in the system. Because exactly the same variables appear on the right hand side of equation (3.3) for all i, the OLS estimates for each equation are identical to GLS estimates for equation (3.1) as a whole. This is a consequence of Kruskal's theorem and Zellner (1962).

Different possibilities for estimating a VAR (p) process exist and each method has its own consequences during forecasting. This section will discuss the Multivariate Least Squares (MLS) estimation.

3.3.1 Multivariate Least Squares Estimation

The estimator obtained for the standard form (3.1) of a VAR (p) process is considered. Some properties of the estimator are derived in the subsequent section.

3.3.1.1 The Estimator

It is assumed that a time series $y_1, \dots, y_T$ of the y variables is available, that is, we have a sample of size T for each of the k variables for the same sample period. In addition p pre-sample values for each variable, $y_{-p+1}, \dots, y_0$ are assumed to be available.

The VAR (p) model in (3.1) can be written compactly as

$$Y = \boldsymbol{B}Z + \boldsymbol{U}$$

where

$$Y = (y_1, \dots, y_T)\,, \quad (K \times T)$$

$$\boldsymbol{B} = \left(v \quad A_1 \dots A_p\right),\ K \times (Kp + 1)$$

$$\boldsymbol{Z_t} = \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}, ((Kp + 1) \times 1\ )$$

$$Z = (Z_0, \dots, Z_{T-1}), \left((Kp + 1) \times T\ \right)$$

$$\boldsymbol{U} = (u_1, \dots, u_T)\,, (K \times T) \tag{3.4}$$

$\boldsymbol{y} = vec\ (Y)\ , KT \times 1$

$\boldsymbol{\beta} = vec\ (B)\ , \left((K^2p + K) \times 1\right)$

$\boldsymbol{b} = vec\ (B')\ , \left((K^2p + K) \times 1\right)$

$\boldsymbol{u} = vec\ (U)\ , (KT \times 1)$

Here vec is the column stacking operator as defined in appendix A2.

$$vec\ (Y) = vec\ (BZ) + vec\ (U)$$

$$= (Z' \otimes I_K) vec\ (B) + vec\ (\boldsymbol{U})$$

Or

$$\boldsymbol{y} = (Z' \otimes I_{\boldsymbol{K}})\boldsymbol{\beta} + \boldsymbol{u}$$

Note that the covariance matrix of u is

$$\sum\nolimits_u = \boldsymbol{I}_T \otimes \Sigma_u$$

Where $\otimes$ is the Kronecker product as defined in appendix A1.

The multivariate LS estimation (or GLS estimation) of $\boldsymbol{\beta}$ means to choose the estimator that minimizes

$$S(\beta) = u'(I_T \otimes \sum\nolimits_u)^{-1} u = u'(I_T \otimes \Sigma_u^{-1})u$$

$$= [\boldsymbol{y} - (Z' \otimes I_K)\boldsymbol{\beta}]'(I_T \otimes \Sigma_u^{-1})[\boldsymbol{y} - (Z' \otimes I_K)\boldsymbol{\beta}]$$

$$= vec\ (Y - BZ)'(I_T \otimes \textstyle\sum_u^{-1})vec\ (Y - BZ)$$

$$= tr[(Y - BZ)'\textstyle\sum_u^{-1}(Y - BZ)]$$

In order to find the minimum of this function we note that

$$S(\beta) = \boldsymbol{y}'(I_{\boldsymbol{T}} \otimes \textstyle\sum_u^{-1})\boldsymbol{y} + \boldsymbol{\beta}'(Z \otimes I_K)(I_T \otimes \textstyle\sum_u^{-1})(Z' \otimes I_K)\boldsymbol{\beta} - 2\boldsymbol{\beta}'(Z \otimes I_K)(I_T \otimes \textstyle\sum_u^{-1})\boldsymbol{y}$$

$$= \boldsymbol{y}'(I_T \otimes \textstyle\sum_u^{-1})\boldsymbol{y} + \boldsymbol{\beta}'(ZZ' \otimes \textstyle\sum_u^{-1})\boldsymbol{\beta} - 2\boldsymbol{\beta}'(Z \otimes \textstyle\sum_u^{-1})\boldsymbol{y}$$

Hence

$$\frac{\partial S(\beta)}{\partial \beta} = 2(ZZ' \otimes \sum_u^{-1})\boldsymbol{\beta} - 2(Z \otimes \sum_u^{-1})\,\boldsymbol{y}$$

Equating to zero gives the normal equations

$$(ZZ' \otimes \sum_u^{-1})\boldsymbol{\beta} = (Z \otimes \sum_u^{-1})\,\boldsymbol{y}$$

This implies that

$$\widehat{\boldsymbol{\beta}} = ((ZZ')^{-1} \otimes \sum_u)\,(Z \otimes \sum_u^{-1})\,\boldsymbol{y}$$

$$= ((ZZ')^{-1}Z \otimes I_K)\boldsymbol{y}$$

The Hessian of $S(\beta)$

$$\frac{\partial^2 S}{\partial \beta \partial \beta'} = 2(ZZ' \otimes \textstyle\sum_u^{-1})$$

is positive definite which confirms that $\widehat{\boldsymbol{\beta}}$ is indeed a minimizing vector. It has to be assumed that ZZ' is non-singular for these results to hold.

The multivariate LS estimator $\widehat{\boldsymbol{\beta}}$ obtained above is identical to the OLS estimator obtained by minimizing

$$\bar{S}(\beta) = \boldsymbol{u}'\boldsymbol{u} = [\boldsymbol{y} - (Z'\otimes I_K)\boldsymbol{\beta}]'[\boldsymbol{y} - (Z'\otimes I_K)\boldsymbol{\beta}]$$

This result is due to Zellner (1962) who showed that GLS and LS estimation in a multiple equation model are identical if the regressors in all equations are the same.

The LS estimator can also be written as

$$\widehat{\boldsymbol{\beta}} = ((ZZ')^{-1}Z\otimes I_K)[(Z'\otimes I_K)\boldsymbol{\beta} + \boldsymbol{u}]$$

$$= \boldsymbol{\beta} + ((ZZ')^{-1}Z\otimes I_K)\boldsymbol{u} \tag{3.5}$$

Or

$$vec\left(\hat{B}\right) = \widehat{\boldsymbol{\beta}} = ((ZZ')^{-1}Z\otimes I_K)vec\,(\boldsymbol{Y})$$

$$= vec(YZ'(ZZ')^{-1}$$

Thus

$$\widehat{\boldsymbol{B}} = YZ'(ZZ')^{-1}$$

$$= (\boldsymbol{B}Z + \boldsymbol{U})Z'(ZZ')^{-1}$$

$$= \boldsymbol{B} + \boldsymbol{U}Z'(ZZ')^{-1} \tag{3.6}$$

Another possibility to write the LS estimator is

$$\widehat{\boldsymbol{b}} = vec\left(\hat{B}'\right) = (I_K\otimes(ZZ')^{-1}Z)vec(\boldsymbol{Y}') \tag{3.7}$$

If we let b'_k be the k^{th} row of $\boldsymbol{B}$, that is, b_k contains all the parameters of the k^{th} equation. Obviously $\boldsymbol{b}' = (b'_1, \dots, b'_k)$.

If we let $y_{(k)} = (y_{k1}, \dots, y_{kT})'$ be the time series available for the k^{th} variable, so that

$$vec(Y') = \begin{bmatrix} y_{(1)} \\ \vdots \\ y_{(K)} \end{bmatrix}$$

then$\hat{b}_k = (ZZ')^{-1}Zy_{(K)}$ is the OLS estimator of the model

$$y_{(K)} = Z'^{b_k} + u_{(k)} \tag{3.8}$$

Where $u_k = (u_{k1}, \dots, u_{kT})$ and $\hat{b}' = (\hat{b}'_1, \dots, \hat{b}'_k)$.

Comparing equation (3.7) and equation (3.8), it is easy to see that the GLS estimator is equivalent to OLS estimator of each of the K equations in model (3.1) separately.

3.4 Asymptotic Properties of GLS Estimators

Consistency and asymptotic normality are established if the following results hold:

$$\Gamma \coloneqq plim \frac{ZZ\prime}{T} \tag{3.9}$$

exist and is non-singular and

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} vec\ (\boldsymbol{u}_t Z'_{t-1}) = \frac{1}{\sqrt{T}} vec\ (\boldsymbol{U}Z') = \frac{1}{\sqrt{T}} (Z \otimes I_K)\boldsymbol{u} \xrightarrow{d} N(0, \Gamma \otimes \textstyle\sum_u)\ \ as\ T \to \infty \tag{3.10}$$

where $\xrightarrow{d}$ denotes convergence in distribution.

The theorem due to Mann and Wald (1943) shows that these results are true under suitable conditions for $\boldsymbol{u_t}$, if $\boldsymbol{y_t}$ is a stationary, stable VAR (p). For instance, the conditions stated in the following definition are sufficient.

Definition 3.1: Standard White Noise

A white noise process $\boldsymbol{u_t} = (u_{1t}, \dots, u_{kt})'$ is called standard white noise if the u_t are continuous random vectors satisfying $E(u_t) = 0$, $\sum_u = E(u_t u'_t)$ is non-singular, u_t and u_s are independent for $s \neq t,$ and, for some finite constant c

$$E\left|u_{it}u_{jt}u_{kt}u_{mt}\right| \leq c\ for\ i, j, k, m = 1, \dots, K\ and\ all\ t \tag{3.11}$$

Condition (3.11) means that all fourth moments exist and are bounded. It is obvious that if the $\boldsymbol{u_t}$ are normally distributed (Gaussian), then they satisfy the moment requirements.

With this definition it is easy to state conditions for consistency and asymptotic normality of the LS estimator. The following lemma is important for proving the large sample results.

Lemma 3.1

If y_t is a stable, K-dimensional VAR (p) process in equation (3.1) with standard white noise residuals $\boldsymbol{u_t}$ then results (3.9) and (3.10) hold.

Proof

The proof of this lemma can be found in theorem 8.2.3 of Fuller (1976) page 340.

The Lemma holds also for other definitions of standard white noise. For example, the convergence result in equation (3.10) follows from a CLT for martingale differences or martingale differences arrays (See proposition 1) by noting that $\omega_t = vec(u_t Z'_{t-1})$ is a martingale difference sequence under quite general conditions. The convergence result in (3.9) may then be obtained from the weak Law of Large Numbers (see proposition B1 in appendix B).

We now formally state the asymptotic properties of the LS estimator.

Proposition 3.2 (Asymptotic Properties of the LS estimator)

Let $\boldsymbol{y_t}$ be a stable, K-dimensional VAR (p) process as in (3.1) with standard white noise residuals, $\widehat{\boldsymbol{B}} = \boldsymbol{Y}Z'(ZZ')^{-1}$ is the LS estimator of the VAR coefficients B. Then

$$plim\ \widehat{\boldsymbol{B}} = \boldsymbol{B}$$

And

$$\sqrt{T}\big(\hat{\beta} - \beta\big) = \sqrt{T}\ vec(\hat{B} - B) \xrightarrow{d} N(0, \Gamma^{-1} \otimes \sum\nolimits_{u})$$

Or equivalently

$$\sqrt{T}\big(\widehat{\boldsymbol{b}} - \boldsymbol{b}\big) = \sqrt{T}vec\big(\widehat{\boldsymbol{B}}' - \boldsymbol{B}'\big) \xrightarrow{d} N\left(0, \sum\nolimits_{u} \otimes \Gamma^{-1}\right)$$

Where $\Gamma \coloneqq plim \frac{ZZ\prime}{T}$

Proof

Using equation (3.6)

$$plim\ \big(\widehat{\boldsymbol{B}} - \boldsymbol{B}\big) = plim\ (\frac{UZ'}{T})plim(\frac{ZZ'}{T})^{-1} = 0$$

By Lemma 1, because 3.10implies $plim\ \frac{UZ'}{T} = 0$. Thus the consistency of $\widehat{\boldsymbol{B}}$ is established.

Using equation (3.5)

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)=\sqrt{T}((ZZ')^{-1}Z\otimes I_K)\boldsymbol{u}$$

$$=(\left(\frac{1}{T}ZZ')^{-1}\otimes I_K\right)\frac{1}{\sqrt{T}}(Z\otimes I_K)\boldsymbol{u}$$

By proposition D2 (4) of appendix D, $\sqrt{T}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)$ has the same asymptotic distribution as

$$\left[plim(\frac{1}{\sqrt{T}}ZZ')^{-1}\otimes I_K\right]\frac{1}{\sqrt{T}}(Z\otimes I_K)u=(\Gamma^{-1}\otimes I_K)\frac{1}{\sqrt{T}}(Z\otimes I_K)u$$

Hence the asymptotic distribution of $\sqrt{T}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)$ is normal by Lemma 1 and the covariance matrix is

$$(\Gamma^{-1}\otimes I_K)\left(\Gamma\otimes\sum\nolimits_u)\,(\Gamma^{-1}\otimes I_K\right)=\Gamma^{-1}\otimes\textstyle\sum_U$$

3.5 Asymptotic Properties of the White Noise Covariance Estimators

In order to assess the asymptotic dispersion of the LS estimator, we need to know the matrices Γ and $\sum_u$.From (3.9) a consistent estimator for Γ is

$$\widehat{\Gamma}=\frac{ZZ'}{T}$$

Because $\sum_u=E(u_t u'_t)$, a plausible estimator for the covariance matrix is

$$\widetilde{\textstyle\sum}_u=\frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{u}}_t\widehat{\boldsymbol{u}}'_t=\frac{1}{T}\widehat{\boldsymbol{U}}\widehat{\boldsymbol{U}}'=\frac{1}{T}\left(\boldsymbol{Y}-\widehat{\boldsymbol{B}}Z\right)\left(\boldsymbol{Y}-\widehat{\boldsymbol{B}}Z\right)'$$

$$=\frac{1}{T}[\boldsymbol{Y}-\boldsymbol{Y}Z'(ZZ')^{-1}Z][\boldsymbol{Y}-\boldsymbol{Y}Z'(ZZ')^{-1}Z]'$$

$$= \frac{1}{T}\boldsymbol{Y}[I_T - Z'(ZZ')^{-1}Z][I_T - Z'(ZZ')^{-1}Z]'$$

$$= \frac{1}{T}\boldsymbol{Y}(I_T - Z'(ZZ')^{-1}Z)\boldsymbol{Y'}$$

An adjustment for degrees of freedom is desired because in a regression with fixed, non-stochastic regressors this leads to an unbiased estimator of the covariance matrix. Thus

$$\widehat{\Sigma}_u = \frac{T}{T - Kp - 1}\widetilde{\Sigma}_u$$

may be considered.

Proposition 3.3: Asymptotic properties of the white noise covariance estimators

Let $\boldsymbol{y_t}$ be a stable, K-dimensional VAR (p) process as in (3.1) with standard white noise innovations and let $\overline{\boldsymbol{B}}$ be an estimator of the VAR coefficients $\boldsymbol{B}$ so that $\sqrt{T}vec(\overline{\boldsymbol{B}} - \boldsymbol{B})$ converges in distribution. Using the symbols from (3.4) suppose that

$$\frac{\overline{\Sigma_u =} (\boldsymbol{Y} - \overline{\boldsymbol{B}}Z)(\boldsymbol{Y} - \overline{\boldsymbol{B}}Z)'}{T - c}$$

where c is a fixed constant. Then

$$plim\sqrt{T}\left(\frac{\overline{\Sigma_u -}\boldsymbol{UU'}}{T}\right) = 0$$

Proof

$$\frac{1}{T}(\boldsymbol{Y} - \overline{\boldsymbol{B}}Z)(\boldsymbol{Y} - \overline{\boldsymbol{B}}Z)' = (\boldsymbol{B} - \overline{\boldsymbol{B}})\left(\frac{ZZ'}{T}\right)(\boldsymbol{B} - \overline{\boldsymbol{B}})' + (\boldsymbol{B} - \overline{\boldsymbol{B}})\frac{Z\boldsymbol{U}'}{T} + \frac{\boldsymbol{U}Z'}{T}(\boldsymbol{B} - \overline{\boldsymbol{B}})' + \frac{\boldsymbol{UU}'}{T}$$

Under the conditions of the proposition, $plim(\boldsymbol{B}-\overline{\boldsymbol{B}})=0$. Hence by Lemma 1

$$plim(\boldsymbol{B}-\overline{\boldsymbol{B}})\frac{Z\boldsymbol{U}'}{\sqrt{T}}=0$$

And

$plim\left[(\boldsymbol{B}-\overline{\boldsymbol{B}})\frac{\boldsymbol{ZZ'}}{T}\sqrt{T}(\boldsymbol{B}-\overline{\boldsymbol{B}})'\right]=0$ (See appendix C1)

Thus

$$plim\sqrt{T}\left[\frac{(\boldsymbol{Y}-\overline{\boldsymbol{B}}Z)(\boldsymbol{Y}-\overline{\boldsymbol{B}}Z)'}{T}-\frac{\boldsymbol{UU'}}{T}\right]=0$$

The proposition follows by noting that as $T\to\infty$, then $\frac{T}{T-c}\to 1$

The proposition covers both $\widehat{\sum}_u$ and $\widetilde{\sum}_u$ This implies that both $\widehat{\sum}_u$ $and\widetilde{\sum}_u$have the same asymptotic properties as the estimator

$$\frac{\boldsymbol{UU'}}{T}=\frac{1}{T}\sum_{t=1}^{T}u_t u'_t$$

which is based on the unknown true residuals and therefore not feasible in practice.

In particular, if $\sqrt{T}vec(\frac{\boldsymbol{UU'}}{T}-\sum_u)$ converges in distribution, $\sqrt{T}vec(\widehat{\sum}_u-\sum_u)$ and $\sqrt{T}vec(\widetilde{\sum}_u-\sum_u)$ will have the same limiting distribution (see proposition D2 of appendix D1)

Corollary 3.1

Under the conditions of proposition 3.3

$$plim\widehat{\Sigma}_u = plim\widetilde{\Sigma}_u = plim\frac{\boldsymbol{UU'}}{T} = \Sigma_u$$

Proof

By proposition 3.2 it suffices to show that $plim\frac{\boldsymbol{UU'}}{T} = \Sigma_u$, which follows from proposition D3 (4) of appendix D because

$$E\left(\frac{1}{T}\boldsymbol{UU'}\right) = \frac{1}{T}\sum_{t=1}^{T} E(u_t u'_t) = \Sigma_u$$

And

$$Var\left(\frac{1}{T}vec(\boldsymbol{UU'})\right) = \frac{1}{T^2}\sum_{t=1}^{T} Var[vec(u_t u'_t)] \leq \frac{T}{T^2}g \rightarrow 0 \; as \; T \rightarrow \infty$$

where$g$ is the constant upper bound for $Var[vec(u_t u'_t)]$. This bound exists because the fourth moments of u_t are bounded by definition 3.1.

3.6 Formulation of Each Country VARY*Model

In this model each country will be linked with the others by including foreign-specific variables. Since all countries are modeled individually, then in each country VARY* model, country specific variables are related to deterministic variables such as time trend (t) and a set of country specific foreign variables. Each country will be modeled as a VARY* model as shown below

$$\boldsymbol{y_{it}} = \boldsymbol{a_{i0}} + a_{i1}t + \phi_{i1}\boldsymbol{y_{i,t-1}} + \cdots + \phi_{ip}\boldsymbol{y_{i,t-p}} + \Lambda_{i0}\boldsymbol{y^*}_{\boldsymbol{it}} + \Lambda_{i1}\boldsymbol{y^*}_{\boldsymbol{i,t-1}} + \cdots + \Lambda_{iq}\boldsymbol{y^*}_{\boldsymbol{i,t-q}} + \boldsymbol{u}_{it} \quad (3.12)$$

Where

$$t = 1, \ldots, T$$

$$i = 1, \ldots, N$$

$\boldsymbol{y_{it}}$ is a $k_i \times 1$ vector of country specific domestic variables

$\boldsymbol{y^*_{it}}$ is the $k^*{}_i \times 1$ vector of foreign variables specific to country i

ϕ_{ip} is a $k_i \times k_i$ matrix of coefficients associated to lagged domestic variables

Λ_{i0} is a $k_i \times k^*{}_i$ matrices of coefficients related to contemporaneous foreign variables

Λ_{ij} is a $k_i \times k^*{}_i$ matrices of coefficients related to the lagged foreign variables $(j = 1, \ldots, q)$

$\boldsymbol{a_{i0}}$ is a $k_i \times 1$ vector of fixed intercepts

$\boldsymbol{a_{i1}}$ is a $k_i \times 1$ vector of coefficients of the deterministic time trend

$\boldsymbol{u_{it}}$is a $k_i \times 1$ vector of country specific shocks assumed to be serially uncorrelated with a zero mean and a non-singular covariance matrix. Specifically $u_{it} \sim (0, \sum_u)$

The domestic variables and foreign variables are grouped as

$$Z_{it} = \left(y_{it}, y^*{}_{it}\right)$$

Each country model in (3.12) is then written as

$$\boldsymbol{A_{i0}} Z_{it} = \boldsymbol{a_{i0}} + a_{i1} t + A_{i1} Z_{i,t-1} + \cdots + A_{i1} Z_{i,t-p} + \boldsymbol{u_{it}} \qquad (3.13)$$

where it is assumed that $p = q$ for ease of computation

In equation (3.13)

$$A_{i0} = (I_{ki}, -\Lambda_{i0})$$

$$A_{i1} = (\phi_{i1}, \Lambda_{i1})$$
$$\vdots$$
$$A_{ip} = (\phi_{ip}, \Lambda_{ip})$$

And the A_{ip} coefficient matrices are all of size $k_i \times (k_i + k^*{}_i)$. Equation (3.13) can be treated like a VAR (p) model by multiplying throughout by ${A_{i0}}^{-1}$.

3.7 GVAR Model Specification

To examine the endogeneity of the foreign variable y_{it}^*, we need to solve the entire (global) model. Stacking over the countries model can be written as

$$\boldsymbol{y_t} = \boldsymbol{a_0} + \boldsymbol{a_1}t + \Phi_1\boldsymbol{y_{t-1}} + \cdots + \Phi_p\boldsymbol{y_{t-p}} + \Lambda_0 w\boldsymbol{y}_t + \Lambda_1 w\boldsymbol{y_{t-1}} + \cdots + \Lambda_p w\boldsymbol{y_{t-p}} + \boldsymbol{u_t}$$

where

$$\mathbf{y_t} \text{ is } \mathrm{Nk} \times 1$$

$$\mathrm{a}_0 \text{ is } \mathrm{Nk} \times 1$$

$$\mathrm{a}_1 \text{ is } \mathrm{Nk} \times 1$$

$$\Phi_1 \ldots \Phi_\mathrm{p} \text{ is } \mathrm{Nk} \times \mathrm{Nk}$$

$$\mathrm{y}_{\mathrm{t}-1} \ldots \mathrm{y}_{\mathrm{t}-\mathrm{p}} \text{ is } \mathrm{Nk} \times 1$$

$$\Lambda_0, \Lambda_1, \ldots, \Lambda_\mathrm{p} \text{ is } \mathrm{Nk} \times \mathrm{Nk}$$

The solution of the stacked model is obtained as

$$\boldsymbol{y_t} = (I_{kN} - \Lambda_0 w)^{-1}\left(\boldsymbol{a_0} + a_1 t + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \Lambda_1 w y_{t-1} + \cdots + \Lambda_p w y_{t-p} + \boldsymbol{u_t}\right) \tag{3.14}$$

Provided the innovations u_t are independent in the time dimension, the endogeneity of the regressors wy_t follows from

$$E(wy_t u_t) = w(I_{kN} - \Lambda_0 w)^{-1} E(u_t u_t')$$

Pesaran et al. (2002) assumed that the weight matrices w_{ij} are diagonal with

$w_{ij} = diag\left(w_{ij}^1, \dots, w_{ij}^k\right)$and that $\sum_{j=0}^{N}\left(w_{ij}^m\right)^2 \to 0$, as $N \to \infty, for\ all\ i\ and\ m$

However, this implies that asymptotically the foreign variables have no explanatory power in the model. Asymptotic properties of such model should not be used as small sample guidance for our estimators if we actually expect some degree of cross sectional dependence in our model.

The assumption $\sum_{j=0}^{N}\left|w_{ij}^m\right| \le c < \infty, for\ all\ i\ and\ m$, where the constant c does not depend on the sample size N. This is clearly a weaker assumption but it turns out to be powerful enough to allow us derive asymptotic properties of our model.

3.7.1 Assumptions

The general assumptions that are maintained/applied throughout this section are listed below. The importance of such assumptions is also explained.

Assumption 1

The disturbances u_{it} are generated from

$$u_t = R_{t,N}\ \eta_t$$

Where $\eta_t = (\eta_{1t}, \dots, \eta_{Nt})$ $is\ Nk \times 1\ with\ \eta_{it}\ being\ a\ k \times 1\ vector\ of\ innovations$ and

a) The innovations η_{mit} are totally independent (w.r.t $i, t\ and\ m\ indexes$) and have uniformly bounded absolute $4 + \delta$ moments for some $\delta > 0$

b) The sequence of $Nk \times Nk$ matrices $R_{t,N}$ has uniformly bounded absolute row sums i.e. denoting $r_{ij,t,N}$ the ij^{th} element of $R_{t,N}$, it holds that

$$\sum_{j=1}^{Nk} |r_{ij,t,N}| \leq k_r < \infty$$

Where the constant k_r does not depend on T or N.

This assumption allows for a general heterogeneity structure within a given time period. However, it imposes the restriction that the disturbances at different time periods are independent. Part (a) is a standard restriction required for deriving asymptotic results, while part (b) guarantees that the amount of heterogeneity in the disturbances is asymptotically limited as the number of countries in the sample increases.

Assumption 2

a) The sequence of the weight matrices w has uniformly bounded absolute row and column sums i.e. denoting $w_{ij,qm}$, the (q, m)-th element of w_{ij}, it holds that $\sum_{j=1}^{N} \sum_{m=1}^{k} |w_{ij,qm}| \leq k_w < \infty$where the constant k_w does not depend on $T\ or\ N$ and the choice of indexes $i\ and\ q$. But can partially depend on other parameters of the model.

b) The sequences of matrices $(I_{kN} - \Lambda_0 w)^{-1}$ and $\left[I_{kN} - (I_{kN} - \Lambda_0 w)^{-1}(\Phi_1 + \Lambda_1 w)^{-1} \dots (\Phi_p + \Lambda_p w)^{-1}\right]^{-1}$ are well defined (the inverses exist) and have uniformly bounded absolute row and column sums.

c) The parameter space is uniformly bounded i.e. the matrices $\Phi_1, \dots, \Phi_p, \Lambda_0, \dots, \Lambda_p$ have uniformly bounded absolute row sums and the vectors $a_0 \; and \; a_1$ have elements uniformly bounded in absolute value.

Assumption 2 guarantees that the degree of international interactions in the data does not explode as the sample size (number of countries) increases.

The existence of the inverses in the above assumption will be guaranteed by the following assumptions that imposes stability of the process in both $N \; and \; T$ dimensions (i.e. assumptions 3 and 4)

Notation

Let A be any square $n \times n$ matrix with real entries. We denote its spectral radius as

$$\rho(A) \coloneqq max\{|\lambda| : \lambda \; is \; an \; eigen \; value \; of \; A\}$$

Assumption 3

The spectral radius of $\Lambda_0 w$ is uniformly less than one i.e. $\rho(\Lambda_0 w) \leq k < 1$ where the constant k does not depend on $N \; or \; T$

Assumption 4

The spectral radius of $(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w)$ and of $(I_{kN} - \Lambda_0 w)^{-1}\left(\Phi_1 + \Phi_p + \Lambda_1 w + \Lambda_p w\right)$ are uniformly less than one.

Assumption 5

The initial observations y_o are drawn from $y_0 = R_0 u$

Where

a) The innovations collected in the $Nk \times 1$ vector u are totally independent of each other as well as of innovations $\eta_t \; for \; t > 0$ and the elements of u have uniformly bounded absolute $4 + \delta$ moments for some $\delta > 0$

b) The sequence of $Nk \times Nk$ matrices R_0 has uniformly bounded absolute row sums i.e. $\sum_{j=1}^{Nk} |r_{ij,0}| \leq k_0 < \infty$ where the constant k_0 does not depend on $N \; and \; T$.

This assumption is about the initial starting values of the process and it will enable us demonstrate that the observable process is a well-defined transformation of the underlying innovations.

The following lemma will be useful in testing the stability of a GVAR model.

Lemma 3.2

Let A, B and C be square matrices with same dimensions and let $\|A\|$ and $\|B\|$ be less than one for some matrix norm. Then the matrix $S = \sum_{n=0}^{\infty} A^n \, CB^n$ is well defined and

$$vec(S) = [I - (B' \otimes A)]^{-1} vec(C)$$

Furthermore, the finite sum $S_t = \sum_{n=0}^{t} A^n \, CB^n$ can be expressed as

$$S_t = S - A^{t+1} S B^{t+1}$$

3.8 Stability Conditions of the GVAR Process

Inspecting the solution of the global model given in equation 3.53, it follows that to determine whether the GVAR model is stable, it is not sufficient to examine the stability of the country-by-country models separately, ignoring the endogeneity of y_{it}^{*} i.e. to examine the Eigen values of Φ_i *and* Λ_i. Instead the stability of the global model is determined by the spectral radius of

$$(I_{kN} - \Lambda_0 w)^{-1}\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right)$$

Hence it does not suffice to impose stability of each country model (i.e. require that $\rho(\Phi) < 1$). Accounting for the autocorrelation in the foreign variable i.e. imposing that $\rho\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right) < 1$) is also not sufficient.

Instead, the stability of the process also depends on the strength of the contemporaneous global links in the models (i.e. on the parameters collected in Λ_0) and it must be determined by the spectral radius of the entire matrix

$$(I_{kN} - \Lambda_0 w)^{-1}\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 + \cdots + \Lambda_p\right)$$

In general when both N *and* T are allowed to tend to infinity, the claim that this is sufficient is not straight forward and is demonstrated in the following proposition.

Proposition 3.4

Under the assumptions 1-5, y_t has well defined uniformly bounded $4+\delta$ moments for some $\delta > 0$. Furthermore, if $a_1 = 0$, then in the limit as $T \to \infty$, y_T converges in quadratic means to a random variable y_∞ which has well defined finite absolute $4+\delta$ moments for some $\delta > 0$ with

$$E(y_\infty) = \left[I_{kN} - (I_{kN} - \Lambda_0 w)^{-1}\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 + \cdots + \Lambda_p\right)\right]^{-1}(I_{kN} - \Lambda_0 w)^{-1} a_0$$

If additionally, $\lim_{T\to\infty} E(u_t u_t') = \Omega_u$, we have

$$vech[VC(y_\infty)] = \{I_{N^2k^2} - [A(I_{kN} - \Lambda_0 w)^{-1} \otimes A(I_{kN} - \Lambda_0 w)^{-1}]\}^{-1}.Dvech(\Omega_u)$$

Where

$$A = (I_{kN} - \Lambda_0 w)^{-1}\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right)$$

and$D$ is a duplication matrix such that

$$vec(\Omega_u) = Dvech(\Omega_u)$$

Proof

Given assumption 3, the matrix $(I_{kN} - \Lambda_0 w)$ is invertible (cf lemma 5.6.10 and corollary 5.6.16 in Horn and Johnson, 1985) and the endogenous variable y_t can be expressed as in equation (3.14), that is,

$$\boldsymbol{y_t} = (I_{kN} - \Lambda_0 w)^{-1}\left(\boldsymbol{a_0} + \boldsymbol{a_1}t + \Phi_1\boldsymbol{y_{t-1}} + \cdots + \Phi_p\boldsymbol{y_{t-p}} + \Lambda_1 w\boldsymbol{y_{t-1}} + \cdots + \Lambda_p w\boldsymbol{y_{t-p}} + \boldsymbol{u_t}\right)$$

By backward substitution, we then obtain

$$\boldsymbol{y_t} = b_{1t} + b_{2t} + b_{3t} + b_{4t}$$

Where

$$b_{1t} = \sum_{s=0}^{t-1}[A]^s\,(I_{kN} - \Lambda_0 w)^{-1}a_0$$

$$b_{2t} = \sum_{s=0}^{t-1}[A]^s\,(I_{kN} - \Lambda_0 w)^{-1}a_1 s$$

$$b_{3t} = \sum_{s=0}^{t-1}[A]^s\,(I_{kN} - \Lambda_0 w)^{-1}u_s$$

$$b_{4t} = [A]^t y_0$$

By assumption (2b), $b_{1t}\, and\, b_{2t}$ have elements uniformly bounded in absolute value. We demonstrate that the sequences of the stochastic vectors $b_{3t}\, and\, b_{4t}$ have elements with uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$. The claim in the proposition will then follow from Minkowski inequality (appendix C).

Consider the stochastic term $b_{3t} = \sum_{s=0}^{t-1} [A]^s \,(I_{kN} - \Lambda_0 w)^{-1} u_s$

Note that given assumption 1, the random vector η_t and the sequence of matrices $R_{t,N}$ satisfy the conditions of lemma B2 in Mutl (2006). Therefore the elements of the random vector u_t have uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$. From assumption 2, we have that the absolute row sums of $A^s(I_{kN} - \Lambda_0 w)^{-1}$ are uniformly bounded in absolute value. Hence by repeated application of the lemma B2 in Mutl (2006) we have that $A^s(I_{kN} - \Lambda_0 w)^{-1} u_s$ has elements with uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$. By Minikowski inequality, we then have that b_{3t} has elements with uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$.

Next, consider the stochastic term $b_{4t} = [A]^t y_0$.

Again by assumption 2, the matrix A^t has uniformly bounded absolute row sums and hence given assumption 5, we have by the same lemma B2 that the elements of b_{4t} have uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$.

We now turn to the asymptotic moments of $y_t$as $t \to \infty$, assuming that $a_1 = 0$. Using lemma A1 and theorem 5.6.12 in Horn and Johnson (1985), it follows that b_{1t} converges to

$$b_1 = \lim_{t\to\infty} b_{1t} = \lim_{t\to\infty} (I_{kN} - A)^{-1}\,(I_{kN} - A^t)^{-1}(I_{kN} - \Lambda_0 w)^{-1} a_0$$

$$= (I_{kN} - \Lambda)^{-1}(I_{kN} - \Lambda_0 w)^{-1} a_0$$

Given assumption (2b), it follows that b_1 has elements uniformly bounded in absolute value and it suffices to show that the elements of b_{3t} *and* b_{4t} converge in quadratic means to random variables b_3 *and* b_4 with finite $4 + \delta$ moments for some $\delta > 0$. By assumption 1, the elements of b_{3t} are independent of the elements of b_{4t}.

Denote the matrix $B_{3s} = A^s(I_{kN} - \Lambda_0 w)^{-1} R_s$ and note that from assumption 1 and 2b it follows that

$$\sum_{s=0}^{\infty} \|B_{3s}\|_1 \leq \|A^s(I_{kN} - \Lambda_0 w)^{-1}\|_1 . k_r$$

$$\leq \|A^s\|_1 . k_1 k_r = \|(I_{Nk} - A)^{-1}\|_1 . k_1 k_r$$

$$\leq k_2 k_1 k_r < \infty$$

Where k_r is the uniform bound for absolute row sums of matrices R_t, k_1 *and* k_2 are uniform bounds for absolute row sums of matrices $(I_{kN} - \Lambda_0 w)^{-1}$and $(I_{kN} - A)^{-1}$.

Given assumption 1, the elements of b_{3t} satisfy conditions of lemma B1 in Mutl (2006) and hence converge in quadratic means to a random variable with uniformly bounded absolute $4 + \delta$ moments for some $\delta > 0$.

Finally, note that from assumption 4 and theorem 5.6.12, it follows that

$$\lim_{t \to \infty} A^t = 0$$

And hence given assumption 5, we have that the elements of b_{4t} converge in quadratic means to zero.

Therefore, the random variable y_∞ is well defined and we have

$$y_\infty = \lim_{t\to\infty} B_{0t}a_0 + \sum_{s=0}^{\infty} A^s \,(I_{kN} - \Lambda_0 w)^{-1} u_s$$

$$= (I_{kN} - A)^{-1}(I_{kN} - \Lambda_0 w)^{-1} a_0 + \sum_{s=0}^{\infty} A^s \,(I_{kN} - \Lambda_0 w)^{-1} u_s$$

Hence

$$E(y_\infty) = (I_{kN} - A)^{-1}(I_{kN} - \Lambda_0 w)^{-1} a_0$$

And using the independence of $u_t \; and \; u_s \; for \; t \neq s$

$$VC(y_\infty) = \sum_{s=0}^{\infty} A^s \,(I_{kN} - \Lambda_0 w)^{-1}\Omega_u (I_{kN} - w'\Lambda_0{}')^{-1} A^{s'}$$

Using lemma 2, we find that

$$vech[VC(y_\infty)] = \{I_{N^2k^2} - [A(I_{kN} - \Lambda_0 w)^{-1} \otimes A(I_{kN} - \Lambda_0 w)^{-1}]\}^{-1} D.\, vech(\Omega_u)$$

Where D is a duplication matrix.

We now examine the sufficient conditions for stability in more details. Note that, for any matrix norm, the spectral radius $\rho(A)$ is smaller than the norm $\|A\|$, Horn and Johnson (1985), hence using the sub-multiplicative property of the matrix norm, we have that

$$\rho[(I_{kN} - \Lambda_0 w)^{-1}\Phi] \leq \left\|(I_{kN} - \Lambda_0 w)^{-1}\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right)\right\|$$

$$\leq \|(I_{kN} - \Lambda_0 w)^{-1}\|. \left\|\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right)\right\|$$

Note that from assumption 3 and lemma 5.6.10 in Horn and Johnson (1985), we have by corollary 5.6.16 in Horn and Johnson (1985) that the inverse $(I_{kN} - \Lambda_0 w)^{-1}$ can be expanded as an infinite sum. Therefore, any norm of $(I_{kN} - \Lambda_0 w)^{-1}$ can be bounded above by

$$\|(I_{kN} - \Lambda_0 w)^{-1}\| \leq \sum_{s=0}^{\infty}(\|w\|. \|\Lambda_0\|)^s \tag{3.15}$$

Often, the weight matrices are row normalized. In this case we have that $\|w\|_1 = 1$ and hence

$$\|(I_{kN} - \Lambda_0 w)^{-1}\|_1 \leq \sum_{s=0}^{\infty} \|\Lambda_0\|_1^s$$

$$= \frac{1}{1 - \|\Lambda_0\|_1}$$

$$= \frac{1}{1 - max_{1\leq i\leq N}\{\|\Lambda_{i0}\|_1\}}$$

To satisfy assumption 3 (in the case of $\|w\|_1 = 1$) we can, for example, require that $0 \leq max_{1\leq i\leq N}\{\|\Lambda_{i0}\|_1\} < 1$. However, if there are global feedbacks in the model we have $max_{1\leq i\leq N}\{\|\Lambda_{i0}\|_1\} > 0$ and hence

$$\frac{1}{1 - max_{1\leq i\leq N}\{\|\Lambda_{i0}\|_1\}} > 1$$

In this case the requirement that $\left\|\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right)\right\|_1 < 1$ which is a stronger requirement than $\rho\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right) < 1$ does not necessarily guarantee that the process is stable. This is due to the fact that the requirement $\left\|\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right)\right\|_1 < 1$ is a sufficient condition for $\rho\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right) < 1$.

The following proposition provides a sufficient condition under which the process is stable.

Proposition 3.5

Assume that the maximum absolute row sums of W are less or equal to k_w, i.e. $\|w\|_1 \leq k_w$. Suppose that

$$\|\Phi\|_1 + k_w \left(\|\Lambda_0\|_1 + \cdots + \left\|\Lambda_p\right\|_1 \right) < 1$$

Then the spectral radius of $(I_{kN} - \Lambda_0 w)^{-1}\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right)$ is less than one.

Proof

Observe by equation (3.15) and the assumption in the proposition we have that

$$\rho\left[(I_{kN} - \Lambda_0 w)^{-1}(\Phi_1 + \cdots + \Phi_p)\right] \leq \|(I_{kN} - \Lambda_0 w)^{-1}\|_1 \left\|\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right)\right\|_1$$

$$\leq \left[\sum_{s=0}^{\infty} (k_w \|\Lambda_0\|_1)^s\right] \left[\left\|\Phi_1 + \cdots + \Phi_p\right\|_1 k_w \left\|\Lambda_1 + \cdots + \Lambda_p\right\|_1\right]$$

$$= \frac{\left\|\Phi_1 + \cdots + \Phi_p\right\|_1 + k_w \left\|\Lambda_1 + \cdots + \Lambda_p\right\|_1}{1 - k_w \|\Lambda_0\|_1}$$

Next, note that from the condition in the proposition

($\left\|\Phi_1 + \cdots + \Phi_p\right\|_1 + k_w \left\|\Lambda_1 + \cdots + \Lambda_p\right\|_1 + k_w \|\Lambda_0\|_1 < 1$)it follows that

$$\left\|\Phi_1 + \cdots + \Phi_p\right\|_1 + k_w \left\|\Lambda_1 + \cdots + \Lambda_p\right\|_1 < 1 - k_w \|\Lambda_0\|_1$$

And thus, observe that the condition also implies that

$$k_w \|\Lambda_0\|_1 < 1, thus\ also\ 1 - k_w \|\Lambda_0\|_1 > 0$$

Hence,

$\frac{\|\Phi_1+\cdots+\Phi_p\|_1+k_w\|\Lambda_1+\cdots+\Lambda_p\|_1}{1-k_w\|\Lambda_0\|_1} < 1$which proves the claim.

The above proposition provides a simpler alternative to checking the Eigen values of the entire matrix$(I_{kN} - \Lambda_0 w)^{-1}(\Phi_1 + \cdots + \Phi_p + \Lambda_1 + \cdots + \Lambda_p)$. Note that, when the weights are normalized to add up to one, we have $k_w = 1$ and it suffices to check whether for all country models it holds that the row sums of

$|\Phi_1| + \cdots + |\Phi_p| + |\Lambda_{io}| + |\Lambda_{i1}| + \cdots + |\Lambda_{ip}|$are less than one.

However, the above proposition provides only a sufficient condition for stability. Necessary condition is that the spectral radius of

$(I_{kN} - \Lambda_0 w)^{-1}(\Phi_1 + \cdots + \Phi_p + \Lambda_1 + \cdots + \Lambda_p)$ is less than one.

3.9 The GVAR Framework

We can build a simple version of our GVAR model from each country models represented by equation (3.12) as follows.

We collect all the domestic variables of all the countries to create the global vector

$$\boldsymbol{y_t} = \begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Nt} \end{pmatrix} \tag{3.16}$$

which is a $k \times 1$ vector containing all endogeneous variables, where $k = \sum_{i=1}^{N} k_i$. Following the step that gives rise to equation (3.12) and the one above, we obtain the identity

$$Z_{it} = w_i y_t \tag{3.17}$$

For $i = 1, \dots, N, where\, w_i$ is a country-specific link matrix of dimensions $(k_i + k_i^*) \times k$ constructed on the basis of trade weights. This identity allows writing each country model in terms of the global vector in (3.16). By substituting (3.17) in (3.12), we obtain

$$A_i w_i y_{it} = a_{i0} + a_{i1} t + B_{i1} w_i y_{i,t-1} + \cdots + B_{ip} w_i y_{i,t-p}$$

The individual country models are then stacked, yielding the model for all the variables in the global model y_t to obtain

$$G y_t = a_0 + a_1 t + \sum_{j=1}^{p} H_j\, y_{t-j} + u_t \tag{3.18}$$

Where

$$G = \begin{pmatrix} A_{1,0} w_1 \\ \vdots \\ A_{N,0} w_N \end{pmatrix}, H_j = \begin{pmatrix} A_{1,j} w_1 \\ \vdots \\ A_{N,j} w_N \end{pmatrix}, a_0 = \begin{pmatrix} a_{1,0} \\ \vdots \\ a_{N,0} \end{pmatrix}, a_1 = \begin{pmatrix} a_{1,1} \\ \vdots \\ a_{N,1} \end{pmatrix}, u_t = \begin{pmatrix} u_{1,t} \\ \vdots \\ a_{N,t} \end{pmatrix}$$

Pre-multiplying equation (3.18) by G^{-1} yields an autoregressive representation of the GVAR (p) model shown below

$$\boldsymbol{y_t} = b_0 + b_1 t + \sum_{j=1}^{p} F_j\, y_{t-j} + \varepsilon_{\boldsymbol{t}} \tag{3.19}$$

Where

$$F_j = G^{-1} H_j, b_0 = G^{-1} a_0, b_1 = G^{-1} a_1 \; and \; \varepsilon_t = G^{-1} u_t$$

Equation (3.19) can be treated like any other VAR equation of order p.

3.10 Parameter Estimates for the GVAR model

If we let $B = \left(v, F_1, \dots, F_p\right)$ and $U = (\varepsilon_1, \dots, \varepsilon_T)$ in (3.4) for the GVAR model, and follow the same procedure as explained in the LS estimation of the VAR (p) model, we obtain the GLS estimator for the GVAR coefficients as

$$\hat{B} = YZ'(ZZ')^{-1} \tag{3.20}$$

3.11 Conclusion

This chapter has explained the estimation procedure of the GVAR model parameters. Also the asymptotic properties of the estimators have been outlined in this section. Stability conditions have also been given in this section and it has been shown that it is not sufficient to check the stability of individual country models ignoring the endogeneity of the foreign variables. We determine the Eigen values of the entire matrix in F in order to examine stability conditions.

CHAPTER 4

Bayesian Model Averaging Implementation

4.1 Introduction

This chapter explains Bayesian model averaging implementation for a GVAR model. By incorporating prior information into the estimation process, the estimates obtained using Bayesian methods are generally more precise than those obtained using the standard classical approach, Blake and Mumtaz (2012). In addition, Gibbs sampling method has been used not only to obtain point estimates but also to characterize the uncertainty around those point estimates. Therefore we focus on estimation of GVAR models via Gibbs sampling in this chapter.

4.2 Bayesian theory

Consider Random variables A and B. The rules of probability imply

$$\mathrm{p}(\mathrm{A},\mathrm{B}) = \mathrm{p}(\mathrm{A}|\mathrm{B})\mathrm{p}(\mathrm{B}) \qquad (4.1)$$

Where

$\mathrm{p}(\mathrm{A},\mathrm{B})$ is the joint probability of A and B occurring.

$\mathrm{p}(\mathrm{A}|\mathrm{B})$ is the probability of A occurring conditional on B having occurred.

$\mathrm{p}(\mathrm{B})$ is the marginal probability of B.

Alternatively, we can reverse the roles of A and B and find the joint probability of A and B

$$\mathrm{p}(\mathrm{A},\mathrm{B}) = \mathrm{p}(B|A)\mathrm{p}(\mathrm{A}) \qquad (4.2)$$

Equating equations (4.1) and (4.2) for $\mathrm{p(A,B)}$ and re-arranging provides us with Bayes rule, which lies at the heart of Bayesian econometrics

$$\mathrm{p(B|A)p(A) = p(A|B)p(B) \Rightarrow p(B|A) = \frac{p(A|B)p(B)}{p(A)}} \tag{4.3}$$

Let y be a vector or matrix of data and θ be a vector or matrix which contains the parameters for a model which seeks to explain y. We are interested in learning about θ based on the data y. Bayesian econometrics uses Bayes rule to do so. In other words, the Bayesian would replace B by θ and A by y in (4.3) to obtain

$$\mathrm{p(\theta|y) = \frac{p(y/\theta)p(\theta)}{p(y)}}$$

Bayesians treat $\mathrm{p(\theta|y)}$ as being of fundamental interest. That is, it directly addresses the question, "given the data, what do we know aboutθ?"

Since we are only interested in learning about θ, we can ignore the term $p(y)$ since it does not involve θ. We can then write

$$\mathrm{p(\theta|y)}\alpha p(y|\theta)p(\theta) \tag{4.4}$$

The term $\mathrm{p(\theta|y)}$ is referred to as the posterior density. The p.d.f for the data given the parameters of the model, $p(y|\theta)$ is the likelihood function and $p(\theta)$ as the prior density. Relationship (4.4) is often referred to as "posterior is proportional to likelihood times prior"

The prior $p(\theta)$ does not depend upon the data. Accordingly, it contains any non-data information available about θ. In other words, it summarizes what you know about θ prior to seeing the data.

The likelihood function $p(y|\theta)$ is the density of the data conditional on the parameters of the model. It is often referred to as the data generating process.

The posterior p(θ|y) is the density which is of fundamental interest. It summarizes all we know about θ after i.e. posterior to, seeing the data.

Equation (4.4) can be thought of as an updating rule, where the data allows us to update our prior views about θ. The result is the posterior which combines both data and non-data information.

In addition to learning about parameters of a model, an econometrician might be interested in comparing different models. A model is formally defined by a likelihood function and a prior. Suppose we have m different models, $M_i \, for \, i = 1, \dots, m$, which all seek to explain y. $M_i$depends upon parameters θ^i. In cases where many models are being entertained, it is important to be explicit about which model is under consideration. Hence, the posterior for the parameters calculated using M_i is written as

$$p\left(\theta^i \middle| y, M_i\right) = \frac{p(y/\theta^i, M_i)p(\theta^i|M_i)}{p(y|M_i)} \tag{4.5}$$

And the notation makes clear that we now have a posterior, likelihood and a prior for each model. The logic of Bayesian econometrics suggests that we use Bayes' rule to derive a probability statement about what we do not know (i.e. whether a model is a correct one or not) conditional on what we do know (i.e. the data). This means the posterior model probability can be used to assess the degree of support for M_i.

Using 4.3 with $B = M_i \; and \; A = y$, we obtain

$$p(M_i|y) = \frac{p(y|M_i)p(M_i)}{p(y)} \tag{4.6}$$

$p(M_i)$ is referred to as the prior model probability since it does not involve the data, it measures how likely we believe in M_i to be the correct one before seeing the data.

$p(y|M_i)$ is called the marginal likelihood and is calculated using (4.5) and a few simple manipulations. In particular, if we integrate both sides of (4.5) with respect to θ^i, use the fact that $\int p(\theta^i|y, M_i)d\theta^i = 1$ (probability density functions integrate to one) and rearrange we obtain

$$p(y|M_i) = \int p(y|\theta^i, M_i)p(\theta^i|M_i)d\theta^i$$

Note that the marginal likelihood depends only upon the prior and the likelihood.

Since the denominator in (4.6) is often hard to calculate directly, it is common to compare two models $i \ and \ j$, using the posterior odds ratio, which is simply the ratio of their posterior model probabilities

$$pO_{ij} = \frac{p(M_i|y)}{p(M_j|y)} = \frac{p(y|M_i)p(M_i)}{p(y|M_j)p(M_j)}$$

Posterior odds ratio can be used to calculate the posterior model probabilities given in (4.6).

Econometricians are often interested in prediction. That is, given the observed data,y, the econometrician may be interested in predicting some future unobserved data y^*. Our Bayesian reasoning says that we should summarize our uncertainty about what we don't know (i.e. y^*) through a conditional probability statement. That is, prediction should be based on the predictive density $p(y^*|y)$, or if we have many models, we would want to make explicit the dependence of a prediction on a particular model and write $p(y^*|y, M_i)$

Using a few simple rules of probability, we can write $p(y^*|y)$in a convenient form. In particular, since a marginal density can be obtained from a joint density through integration, we can write

$$p(y^*|y) = \int p(y^*|\theta, y)\, d\theta$$

However, the term inside the integral can be written using another simple rule of probability as

$$p(y^*|y) = \int p(y^*|\theta, y)\, p(\theta|y) d\theta \tag{4.7}$$

The form for the predictive in (4.7) is quite convenient, since it involves the posterior.

4.3 Bayesian Model Implementation for the GVAR Model

Bayesian analysis of the GVAR model requires the elicitation of the prior distributions for all parameters of the model. In this work, we use Minnesota prior structures that have been developed for VAR specifications for the individual country-specific models together with standard prior settings for the parameters corresponding to (weakly) exogenous variables and combine the posterior results to obtain the Bayesian Global Vector Autoregressive (B-GVAR) model, Cuaresma et al. (2014)

We now turn to the key prior distribution for our VARY* models proposed in literature, that is, the Minnesota prior. Though other key prior distributions have been proposed in literature, the choice of the Minnesota prior in this work is due to its flexibility and hence easy to implement in empirical studies.

4.3.1The Minnesota Prior

The Minnesota prior incorporates the prior belief that the endogenous variables included in the VARY* follow a random walk process or an AR (1) process. The prior mean corresponding to the lag of the endogenous variables is thus a priori set to one and the rest of the parameters to zero. In other words, the mean of the Minnesota prior for the VARY* coefficients in equation (3.12) implies the following form

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{kt} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} b_{11}^0 & 0 & \dots & 0 \\ 0 & b_{22}^0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & b_{kk}^0 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \vdots \\ y_{k,t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} y_{1,t-2} \\ y_{2,t-2} \\ \vdots \\ y_{k,t-s} \end{pmatrix} +$$

$$\begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} y_{1,t-j}^* \\ y_{2,t-j}^* \\ \vdots \\ y_{k,t-j}^* \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{kt} \end{pmatrix} \tag{4.8}$$

where $j = 0, \dots, q$.

Equation (4.8) states that the Minnesota prior incorporates the belief that the variables on the left hand side follow an AR (1) process or a random walk process if $b_{11}^0 = b_{22}^0 = b_{kk}^0 = 1$. If the variables on the left hand side of (4.8) are stationary it may be more realistic to incorporate the prior that they follow an AR (1) process. In this case the mean of the Minnesota prior distribution for the VAR coefficients (i.e. $\tilde{b}_0$from $p(b) \sim N(\tilde{b}_0, H)$) is given by the vector

$$\tilde{b}_0 = \begin{pmatrix} 0 \\ b_{11}^0 \\ \vdots \\ 0_{1,p+1} \\ 0 \\ 0 \\ b_{22}^0 \\ \vdots \\ 0_{2,p+1} \\ \vdots \\ 0 \\ 0 \\ \vdots \\ b_{kk}^0 \\ \vdots \\ 0_{k,p+1} \end{pmatrix} \tag{4.9}$$

Where the first $p + 1\, rows$ correspond to the coefficients of the first equation and the second $p + 1$ rows correspond to the coefficients for the second equation and the $k^{th}\, p + 1$ rows correspond to the coefficients for the k^{th} equation.

In our empirical case, we restrict ourselves to a VAR (2) model and k=4 is the number of variables in the model. Hence, $\tilde{b}_0$ in (4.9) above will be a 24×1 column matrix. While implementing Gibbs sampling we incorporate the belief that the variables follow an AR (1) process rather than random walk model priors. This is because; the variables have been transformed and are stationary. We also incorporate an inverse Wishart prior for the covariance matrix as explained in Cuaresma et al (2014) and hence depart from the strict form in other literatures where the covariance matrix is fixed and diagonal. The Bayesian GVAR model is then estimated using Gibbs sampling algorithm as explained below..

4.4 Gibbs Sampling for a VAR Model

As discussed in chapter 1, researchers are interested in marginal posterior distributions which may be difficult to derive analytically. In contrast, the conditional posterior distribution of each set of parameters is readily available. According to the definition below, one can approximate the marginal posterior distribution by sampling from the conditional distributions.

Definition 4.1 (Gibbs sampling)

Gibbs sampling is a numerical method that uses draws from conditional distributions to approximate joint and marginal distributions, Blake and Mumtaz (2012).

4.4.1 A General Description of Gibbs Sampling

Suppose we have a joint distribution of k variables

$$f(x_1, x_2, \dots, x_k) \tag{4.10}$$

and we are interested in obtaining the marginal distributions

$$f(x_i), i = 1, \dots, k$$

The standard of doing this is to integrate the joint distribution in (4.10). However, this integration may be difficult or infeasible in some cases. It may be that the exact form of (4.10) is unknown or too complicated for direct analytical integration.

If we assume that the form of the conditional distributions $f\left(x_i|x_j\right), i \neq j$ is known. A Gibbs sampling algorithm with the following steps can be used to approximate marginal distributions.

1) Set the starting values for $x_1, x_2, \dots, x_k$, that is, $x_1^0, x_2^0, \dots, x_k^0$, where the superscript 0 denotes the starting values.
2) Sample x_1^1 from the distribution of x_1 conditional on current values of $x_2, \dots, x_k$, that is, $f(x_1^1|x_2^0, \dots, x_k^0)$.

3) Sample x_2^1 from the distribution of x_2 conditional on current values of $x_1, x_2, \dots, x_k$, that is, $f(x_2^1 | x_1^1, x_3^0, \dots, x_k^0)$.

$$\vdots$$

k) Sample x_k^1 from the distribution of x_k conditional on current values of $x_1, x_2, \dots, x_{k-1}$, that is $f(x_k^1 | x_1^1, x_2^1, \dots, x_{k-1}^1)$ to complete one iteration of the Gibbs sampling algorithm.

As the number of Gibbs iterations increases to infinity, the samples or draws from the conditional distributions converge to the joint and marginal distributions of x_i at an exponential rate, Casella and George (1992). Therefore, after a large enough number of iterations, the marginal distributions can be approximated by the empirical distribution of x_i.

In other words, one repeats the Gibbs iterations M times (that is, a number of iterations large enough for convergence) and saves the last H draws of x_i (for e.g. H=1000). This implies that the researcher is left with H values for $x_1, x_2, \dots, x_k$. Thus an estimate of the mean of the marginal posterior distribution for x_i is simply the mean of the H retained draws.

$$\frac{1}{H}\sum_{b=1}^{h} x_i^b$$

where the superscript b indexes the retained Gibbs iterations.

4.4.2 Conditional posterior distribution of the VAR parameters

The posterior distribution of the VAR coefficients conditional on $\sum$ is normal as shown in Kadiyala and Karlsson (1997). That is, the conditional posterior for the coefficients is given by $H(b \backslash \sum, Y_t) \sim N(M^*, V^*)$ where

$$M^* = (H^{-1} + \sum^{-1} \otimes X'_t X_t)^{-1} \left(H^{-1} \tilde{b}_0 + \sum^{-1} \otimes X'_t X_t \tilde{b} \right)$$

$$V^* = (H^{-1} + \sum^{-1} \otimes X'_t X_t)^{-1} \tag{4.11}$$

Where $\tilde{b}$ is a $\left(N \times (N \times P + 1)\right) \times 1$ vector which denotes the OLS estimates of the VAR coefficients in vectorised format $\tilde{b} = vec\left((X'_t X_t)^{-1} (X'_t Y_t)\right)$. The format of the conditional posterior mean in equation (4.11) is very similar to the posterior mean for the linear regression model discussed in chapter 1. That is, the mean of the conditional posterior distribution is a weighted average of the OLS estimator $\tilde{b}$ and the prior $\tilde{b}_0$ with the weights given by the inverse of the variance of each ($\sum^{-1} \otimes X'_t X_t$ is the inverse of $\tilde{b}$ while H^{-1} is the inverse of the variance prior).

The conjugate prior for the VAR covariance matrix is an inverse Wishart distribution with prior scale matrix $\bar{S}$ and prior degrees of freedom α.

$$p(\sum) \sim IW(\bar{S}, \alpha) \tag{4.12}$$

Definition 4.2

If$\sum$ is a $n \times n$ positive definite matrix, it is distributed as an Inverse Wishart with $P(\sum) = k \frac{|H|^{v/2}}{|\sum|^{(v+n+1)/2}} exp(0.5 tr \sum^{-1} H)$ where $k^{-1} = 2^{vn/2} \pi^{n(n-1)/4} \prod_{i=1}^{n} \Gamma\left[(v + 1 - i)/2\right]$, H is the scale matrix and v denotes the degrees of freedom, Zellner (1971).

Informally, one can think of the Inverse Wishart distribution as a multivariate version of the inverse Gamma distribution introduced in the context of the linear regression model in chapter 1.

Given the prior in (4.12), the posterior for $\sum$ conditional on b is also an inverse Wishart $H(\sum \backslash b, Y_t) \sim IW\left(\bar{\sum}, T + \alpha\right)$ where T is the sample size and

$$\bar{\bar{\sum}} = \bar{S} + (Y_t - X_t B)'(Y_t - X_t B)$$

Note that B denotes the VAR coefficients reshaped into $(N \times P + 1) \times N$ matrix

4.4.3 Gibbs sampling for the VAR model

The Gibbs sampling algorithm for VARs is very similar to that employed for the linear regression model in chapter 1. The key difference turns out to be the fact that setting up the prior in the VAR model is a more structured process than the linear regression case.

The Gibbs sampling for the VAR model consists of the following steps:

Step 1: Set priors for the VAR coefficients and the covariance matrix. As discussed above, the prior for the VAR coefficients is normal and given by $p(b) \sim N\left(\tilde{b}_0, H\right)$). The prior for the covariance matrix of the residual $\sum$ is Inverse Wishart and given by $IW(\bar{S}, \alpha)$. Set a starting value for $\sum$ (e.g. the OLS estimate of $\sum$).

Step 2: Sample the VAR coefficients from its conditional posterior distribution $H(b \backslash \sum, Y_t) \sim N(M^*, V^*)$where $M^*\ and\ V^*$ are as shown in (4.11). Once $M^*\ and\ V^*$ are calculated, the VAR coefficients are drawn from the normal distribution.

Step 3: Draw $\sum$ from its conditional distribution $H(\sum \backslash b, Y_t) \sim IW\left(\bar{\bar{\sum}}, T + \alpha\right)$ where $\bar{\bar{\sum}} = \bar{S} + (Y_t - X_t B^1)'(Y_t - X_t B^1)$ where B^1 is the previous draw of the VAR coefficients reshaped into a matrix with dimensions $(N \times P + 1) \times N$ so it is conformable with X_t.

Repeat steps 2 to 3 M times to obtain $B^1, \dots, B^M$and $(\sum)^1, \dots, (\sum)^M$. The last H values of B and $\sum$ from these iterations is used to form the empirical distribution of these parameters.

4.5 Conclusion

This chapter discusses the BMA implementation for a GVAR model. The main result in our discussion is that, Bayesian analysis of the GVAR model requires prior structures that have been developed for VAR specifications for the individual country-specific models together with standard prior settings for the parameters corresponding to (weakly) exogenous variables and combines the posterior results to obtain the Bayesian Global Vector Autoregressive (B-GVAR) model, Cuaresma et al. (2014). Since researchers are interested in obtaining the marginal posterior distributions which may be difficult to obtain analytically, Gibbs sampling is employed in order to sample from conditional distributions which are readily available.

CHAPTER 5

Research Methodology

5.1 Introduction

This chapter shows the materials and methods used to attain the specific objectives outlined in chapter one. Since there are five specific objectives, a method for each, thus five methods, are discussed in this section.

5.2 Estimation of Individual VARY Models

The VARY models are estimated separately for each country conditional on x_{it}^{*}, using reduced rank regression, taking into account the possibility of Cointegration both within x_{it} and across x_{it} and x_{it}^{*}. This way, the number of Cointegration relations, r_i, the speed of adjustment coefficients, α_i, and the cointegrating vectors β_i for each country are obtained.

Once the variables to be included in the individual country $VAR(p_i, q_i)$ models have been selected, the lag orders of the domestic and foreign variables, $p_i \; and \; q_i$ respectively, are chosen according to either the AIC or the SBC or can be inputted by the user. It should be noted that for each model, $p_i \; and \; q_i$ need not be the same. The lag order of the GVAR, denoted by p, is computed as the maximum value from the lag orders $p_i \; and \; q_i$ across all countries i. For $p > 1$, the lag order selection is performed under the conditions that $q_i \leq p - 1$ and $q_i \leq p_i$ for each country model.

5.2.1 Lag order selection of the individual VARY models

The lag orders p_i *and* q_i –of the domestic and foreign variables-of the individual country models as expressed in the general form of (3.12) can be selected using the Akaike Information Criterion (AIC) or the Scharz Bayesian criterion (SBC). These are computed as follows:

$$AIC_{i,pq} = -\frac{Tk_i}{2}(1 + \log 2\pi) - \frac{T}{2} log\left|\hat{\Sigma}_i\right| - k_i s_i \tag{5.1}$$

And

$$SBC_{i,pq} = -\frac{Tk_i}{2}(1 + \log 2\pi) - \frac{T}{2} log\left|\widehat{\Sigma}_i\right| - \frac{k_i s_i}{2} In\, T \tag{5.2}$$

Where the first two terms in equations (5.1) and (5.2) correspond to the maximized value of the log-likelihood function, $\hat{\Sigma}_i = \sum_{t=1}^{T} \hat{u}_{it}\hat{u}'_{it}/T$ computed based on the estimated residuals $\hat{u}_{it}$ obtained from the estimation of the individual VARY models given by (3.12), T is the sample size and $k_i = k_i^*$ is the number of the domestic and foreign variables respectively in the individual country models and $s_i = k_i p_i + k_i^* q_i$+2. The model with the highest AIC or SBC value is chosen.

5.3 Constructing Weight Matrices

The weights ω_{ij} used for the computation of the foreign variables and solving the GVAR have been constructed from flows data. The weights used throughout this work are trade weights. In this case, ω_{ij} is the share of country j in the trade (exports plus imports) of country i. The trade weights are constructed by dividing the total trade of each country i by the amount of trade with country j, such that the i^{th} row sums to one, for all i, Smith and Galesi (2011).

For the validity of the GVAR methodology as discussed in Pesaran et al. (2004), the trade weights should be 'granular' for each country; in other words they should not be too close to one

5.4 The Global Solution

5.4.1 Constructing the Link Matrices

The link matrix $\boldsymbol{W}_i$ for each country is a $(k_i \times k_i^*) \times k$ matrix of fixed (known) constants defined in terms of the country specific weights, ω_{ij}, where $k_i \; and \; k_i^*$ are the number of domestic and foreign variables respectively and $k = \sum_{i=0}^{N} k_i$. In our case $k_i = k_i^* = 3$ for each of the N countries in the system. The link matrix $\boldsymbol{W} = (\omega_{ij})$ is used in construction of the foreign variables and it is important that the weights reflect as close as possible the underlying economic linkages among countries.

5.4.2 Solving the GVAR model

Solving for the global model follows the procedure whereby country-specific models are properly reformatted and stacked involving the weight matrices that are taken or estimated as outlined above. The description of how to solve the global model proceeds as discussed in sections 3.7 and 3.9 in chapter 3.

5.5 Dynamic properties of the GVAR model

The dynamic properties of the GVAR model as expressed in (3.19) are analyzed by using the Generalized Forecast Error Variance Decomposition (GFEVD) and Generalized Impulse Response Functions (GIRFs) developed by Pesaran and Shin (1998). Although these methods do not allow a structural interpretation of the shocks, they overcome the identification problem by providing a meaningful characterization of the dynamic responses to observable shocks.

The long run properties of the GVAR model are also analyzed by looking at their persistence profiles, Pesaran and Shin (1996), which makes it possible to assess how long the system takes to revert to its steady-state path, after being hit by a system wide shock.

5.5.1 Persistence Profiles

Persistence profiles (PPs) refers to the time profiles of the effects of a system or variable-specific shocks on the cointegrating relations in the GVAR model, Pesaran and Shin (1996). PPs have a value of unity on impact, while they should tend to zero as the horizon $n \to \infty$, if the vector under consideration is a valid cointegrating vector. They provide information on the speed with which the cointegrating relationships return to their equilibrium states.

Consider the GVAR (p) model in (3.19). The moving average presentation of (3.19) is given by

$$y_t = d_t + \sum_{s=0}^{\infty} A_s \, \varepsilon_{t-s} \qquad (5.3)$$

Where d_t represents the deterministic component of y_t, and A_s can be derived recursively as

$$A_s = F_1 A_{s-1} + F_2 A_{s-2} + \cdots + F_p A_{s-p}, s = 1,2, \ldots \qquad (5.4)$$

With $A_0 = I_m, A_s = 0, for\ s < 0$.

The PPs of $\beta'_{ji} z_{it}$, with respect to a system-wide shock to ε_t are obtained as

$$PP\left(\beta'_{ji} z_{it}; \varepsilon_t, n\right) = \frac{\beta'_{ji} W_i A_n \Sigma_\varepsilon A'_n W'_i \beta_{ji}}{\beta'_{ji} W_i A_0 \Sigma_\varepsilon A'_0 W'_i \beta_{ji}}, n = 0,1,2, \ldots \qquad (5.5)$$

Where β'_{ji} is the j^{th} cointegrating relation in the i^{th}country $(j = 1,2, \dots, r_i), n$ is the horizon, $\sum_{\varepsilon}$ is the covariance matrix of ε_t. The A_n matrices are calculated based on equation (5.4).

5.5.3 Generalized Impulse Response Function (GIRFs)

If we consider model (3.18), the GIRFs are based on the definition

$$GIRF(y_t; u_{ilt}, n) = E\left(y_{t+n} | u_{ilt} = \sqrt{\sigma_{ii,ll,}} \mathfrak{T}_{t-1}\right) - E(y_{t+n} | \mathfrak{T}_{t-1}) \quad (5.6)$$

Where $\mathfrak{T}_{t-1}$ is the information set at time $t-1$, $\sigma_{ii,ll}$ is the diagonal element of the variance-covariance matrix $\sum_u$ corresponding to the l^{th} equation in the i^{th} country and n is the horizon.

5.5.4 Generalized Forecast Error Variance Decompositions (GFEVDs)

The GFEVDs consider the proportion of the $n - step$ forecast errors of y_t which is explained by conditioning on the non-orthogonalised shocks $u_{jt}, u_{j,t+1}, \dots, u_{j,t+n}$, for $j = 1, .., k$, while explicitly allowing for the contemporaneous correlations between these shocks and the shocks to the other equations in the systems. The GFEVDs for equation (3.18) will be derived as

$$GFEVD\left(y_{(l)t}; u_{(j)t}, n\right) = \frac{\sigma_{jj}^{-1} \sum_{s=0}^{n} \left(e_l' A_s G_0^{-1} \sum_u e_j\right)^2}{\sum_{s=0}^{n} e_l' A_s G_0^{-1} \sum_u G_0^{-1\prime} A\prime_s e_j} \quad (5.7)$$

5.6 Statistical Tests

5.6.1 Structural Stability tests

For the structural stability tests, consider the compact version of the l^{th} equation of the estimated i^{th} error correction model given by

$$y_{lt} = \theta'_{lt} z_t + \mathrm{e}_{lt} \tag{5.8}$$

A number of structural stability tests, similar to those considered by Stock and Watson (1996) are computed to detect the possibility of the presence of breaks. Among them are;

Tests based on the cumulative sum of OLS residuals

The maximal OLS CUSUM statistic proposed by Ploberger and Kramer (1992) is similar to the Brown et al. (1975) CUSUM statistic, although it is compared using OLS rather than recursive residuals. The mean square version of this is also considered. Let $\varsigma_T(\delta) = \hat{\sigma}_l^{-1} T^{1/2} \sum_{s=1}^{[T\delta]} \mathrm{e}_{ls}$, where $[.]$ is the greatest integer function then

$$PK_{sup} = \sup_{\delta\epsilon[0,1]} |\varsigma_{lT}(\delta)|$$

$$PK_{msq} = \int_0^1 \varsigma_{lT}(\delta)^2 \, d\delta$$

Random walk alternatives

Nyblom (1989) specifies as the alternative that θ_{lt} follows a random walk and proposes the following statistic

$$\mathfrak{N}_l = T^{-2} \sum_{t=1}^{T} S'_{lt} \hat{V}_l^{-1} S_{lt} \tag{5.9}$$

Where $S_{lt} = \sum_{s=1}^{t} z_s \, \mathrm{e}_{ls}$and $\hat{V}_l = (T^{-1} \sum_{t=1}^{T} z_t z'_t)\hat{\sigma}_t^2$.

The heteroskedasticity-robust version of the $\mathfrak{N}_l$ statistic is obtained by replacing $\hat{V}_l$ in (5.9) with $\hat{V}_l = T^{-1} \sum_{t=1}^{T} \mathrm{e}_{lt}^2 z_t z'_t$.

Sequential Wald statistics

i. Quandt (1960) likelihood ratio (QLR) statistic in Wald form $QLR = \underset{\delta\epsilon(\delta_0,\delta_1)}{sup} F_{lt}(\delta)$

ii. The mean Wald statistic in Hansen (1992) is derived as $MW = \int_{\delta_0}^{\delta_1} F_{lt}(\delta)d\delta$

iii. The exponential average Wald statistic by Andrews and Ploberger (1994) $APW = In\left\{\int_{\delta_0}^{\delta_1} F_{lt}(\delta)/2\ d\delta\right\}$

5.6.2 Unit Root tests

Unit root tests are used to test for a non-stationary situation. The following test are used in testing for unit roots in this study

Augmented Dickey Fuller (ADF) test

Augmented Dickey Fuller assumes that the error term ε is unlikely to be white noise (autocorrelated). ADF augment the equations of a model by adding the lagged values of the dependent variable in order to eliminate autocorrelation.

Weighted Symmetric (WS) tests

The WS tests exploit the time reversibility of stationary autoregressive process in order to increase their power performance. Leybourne et al. (2005) provide evidence that the WS statistic has a superior performance compared to the standard ADF tests. The lag order employed in the ADF and WS unit-root tests is selected either by AIC or SBC criterion discussed in section 5.2.1.

5.6.4 Contemporaneous effects of foreign variables on their domestic counterparts

The contemporaneous effects of foreign variables on their domestic counterparts are provided together with t-ratios computed based on standard as well as White and Newey-West adjusted variance matrices. They are informative as regards the international linkages between the domestic and foreign variables. High elasticities between the domestic and foreign variables imply strong co-movements between the two. Considering the l^{th} equation of the i^{th} country error correction model given by

$$\Delta y_{it,l} = \hat{\mu}_{il} + \sum_{j=1}^{\hat{r}_i} \hat{\gamma}_{ij,l} E\hat{C}M_{ij,t-1} + \sum_{n=1}^{\hat{p}_i-1} \hat{\boldsymbol{\varphi}}'_{in,l}\, \Delta\mathbf{x}_{\mathrm{i,t-n}} + \sum_{s=0}^{\hat{q}_i-1} \boldsymbol{\vartheta}'_{is,l}\, \Delta\mathbf{x}^*_{\mathrm{i,t-s}} + e_{it,l}$$

This can be written compactly as

$$y_{\boldsymbol{it,l}} = \boldsymbol{\theta}'_{\boldsymbol{il}} \boldsymbol{z_{it}} + e_{it,l} \tag{5.10}$$

Where $y_{\boldsymbol{it,l}} = \Delta y_{it,l}, \boldsymbol{z_{it}} = \left(1, E\hat{C}M_{ij,t-1}, \Delta\mathbf{x}'_{\mathrm{i,t-n}}, \Delta\mathbf{x}'^{*}_{\mathrm{i,t-s}}\right)', E\hat{C}M_{ij,t-1}, j = 1,2, \dots, r_i$ are the estimated error correction terms corresponding to the r_i cointegrating relations found for the i^{th} country model and $\boldsymbol{\theta}'_{\boldsymbol{il}} = \left(\hat{\mu}_{il}, \hat{\gamma}_{ij,l}, \hat{\boldsymbol{\varphi}}'_{in,l}, \boldsymbol{\vartheta}'_{is,l}\right)'$. Let $e_{it,l}$ be the residuals from the estimated model (5.10) and $\hat{\sigma}^2_{il} = \mathrm{T}^{-1} \sum_{t=1}^{T} e^2_{it,l}$ the corresponding estimated error variance-the White and Newey-West adjusted variance estimates of the individual VECMY regressors collected in $\boldsymbol{\theta}'_{\boldsymbol{il}}$are computed as follows

White's Adjusted Standard Error

The program computes a degrees of freedom corrected version of White's (1980) heteroskedasticity-consistent variance estimator for $\boldsymbol{\theta}'_{\boldsymbol{il}}$using the following formula

$$H\hat{C}V(\boldsymbol{\theta}'_{il}) = \left(\frac{\mathrm{T}}{\mathrm{T} - \kappa_i}\right)(\boldsymbol{Z}'_i\boldsymbol{Z}_i)^{-1}\left(\sum_{t=1}^{T} e^2_{it,l}\, \boldsymbol{z}_{it}\boldsymbol{z}'_{it}\right)(\boldsymbol{Z}'_i\boldsymbol{Z}_i)^{-1}$$

Newey-West Adjusted Standard Error

The Newey-West adjusted variance estimates version of Newey and West (1987) allows for small sample correction and is computed as follows

$$\hat{V}(\boldsymbol{\theta}'_{il}) = \left(\frac{\mathrm{T}}{\mathrm{T} - \kappa_i}\right)(\boldsymbol{Z}'_i\boldsymbol{Z}_i)^{-1}\hat{\boldsymbol{S}}_{iT,l}(\boldsymbol{Z}'_i\boldsymbol{Z}_i)^{-1}$$

Where $\boldsymbol{Z}_i = (\boldsymbol{z}_{i1}, \boldsymbol{z}_{i2}, \dots, \boldsymbol{z}_{iT})'$

5.7 Weak exogeneity

The main assumption underlying the estimation of the individual country VARY models is the weak exogeneity of the foreign variables with respect to the long run parameters of the conditional model defined by the error correction form of a VARY model. This assumption is compatible with a certain degree of weak dependence across $\boldsymbol{u}_{it}$, as discussed in Pesaran et al. (2004). The weak exogeneity test assumption in the context of cointegrating models implies no long run feedback from the domestic variables to the foreign variables, without necessarily ruling out lagged short run feedback between the two sets of variables. A formal test of this assumption for the country specific-foreign variables and the observed global variables is conducted as explained in Johansen (1992).

This involves a test of the joint significance of the estimated error correction terms in auxiliary equations for the country-specific foreign variables, y^*_{it}. In particular, for each l^{th} of y^*_{it} the following regression is carried out

$$\Delta y_{it,l}^{*} = a_{il} + \sum_{j=1}^{r_i} \zeta_{ij,l} E\hat{C}M_{ij,t-1} + \sum_{k=1}^{s_i} \phi_{ik,l}' \Delta y_{i,t-k} + \sum_{m=1}^{n_i} \psi_{im,l}' \Delta \tilde{y}_{i,t-m}^{*} + \eta_{it,l}$$

where $E\hat{C}M_{ij,t-1}, j = 1,2, \dots, r_i$, are the estimated error correction terms corresponding to the r_i cointegrating relations found for the i^{th} country model, and s_i and n_i are the lag orders of the lagged changes for the domestic and foreign variables respectively.

Thus testing for the weak exogeneity involves the marginal model for y_{it}^{*}, that is, the model for the foreign variables.

5.8 Benchmark models

One of our specific objectives is to compare the performance of the Bayesian Global Vector Autoregressive (AR) model to other models available in existing literature. This work compares forecasts from random walk and Autoregressive models, with and without a drift. The specification of these models is as follows

Random Walk: $y_t = y_{t-1} + \varepsilon_t$

Random Walk plus drift μ:$y_t = \mu + y_{t-1} + \varepsilon_t$

AR(1): $y_t = \alpha + \gamma y_{t-1} + \varepsilon_t$

AR (1) plus trend: $y_t = \alpha + \beta t + \gamma y_{t-1} + \varepsilon_t$

The choice of these models as benchmarks in our study stems from the fact that they are surprisingly hard to beat in empirical macro and financial studies. The standard GVAR model applied in our literature is also used as a benchmark model.

5.9 Conclusion

This chapter has outlined the methodology applied in achieving the specific objectives. The test statistics applied in our empirical implementation have also been explained in this section. Finally the random walk and autoregressive models with and without drifts have been explained as the benchmark models for comparing forecasts with the Bayesian GVAR model.

CHAPTER 6

Results

6.1 Introduction

This chapter presents the analysis and findings of the study on financial interlinkages in East African countries: A Bayesian Model Averaging approach to a Global Vector Autoregressive model. The results obtained have been presented in form of tables and graphs. A discussion of each result is given in each subsection that a result has been given.

6.2 Data

This work has used secondary data from the central banks, stock exchanges and the national bureau of statistics of the respective countries. The data span from the year 2000 to 2013. The variables used in our model are Inflation rate, interest rate, exchange rates and oil prices. We follow the bulk of the literature in including oil prices as a global control variable or the dominant variable in the model.

6.2.1 Notations and Data Sources

The list of notations used and data sources for each of the variables used in our study are shown in table 1 below.

Table 1: Variable notation and data sources

Variable	Variable notation Domestic Foreign	Data sources
Inflation	infinfs	National Bureaus of statistics for respective countries.
Exchange rates	excexcs	Central Banks/ National banks for respective countries
Interest rates	intints	Central Banks/ National banks for respective countries
Oil prices	poil	Brent crude oil prices

6.3 Empirical Results

This section describes and explains the output produced from running the East Africa data in matlab.

6.3.1 The Weight Matrix

The weights used throughout are trade weights. In this case, trade weights w_{ij} is the share of country j in the trade (exports plus imports) of country i. The country level trade shares are constructed by dividing the total trade of each country i by the amount of trade with country j, such that the i^{th} column sums to one for all i. These are given in table 2 below

Table 2: Weight Matrix based on fixed weights

Country	RWANDA	KENYA	UGANDA	TANZANIA	BURUNDI
RWANDA	0	0.0983761	0.1196532	0.0532244	0.1072535
KENYA	0.442186	0	0.7372607	0.7488049	0.2510997
UGANDA	0.3773395	0.517266	0	0.150186	0.4867709
TANZANIA	0.1112848	0.3483198	0.099574	0	0.1548759
BURUNDI	0.0691896	0.036038	0.0435121	0.0477847	0

The vast majority of the trade weights are 'granular' for each country; in other words they are not too close to one. The largest weights are observed for Kenya towards Uganda and Tanzania with 0.7372607 and 0.74880049 respectively. The trade weights in this case can be said to be relatively small such that $\sum_{j=1}^{N} w_{ij}^{2} \rightarrow 0 \; as \; N \rightarrow \infty, for \; i = 1,2, \ldots N$. This is one of the three further requirements as a sufficient condition for the validity of the GVAR methodology as discussed in Pesaran et al. (2004).

6.3.2 Stationarity Tests

In this section we discuss the ADF unit root t-statistics as well as those based on weighted symmetric estimation of ADF type regressions introduced by Park and Fuller (1995). The latter tests denoted by Ws, exploit the time reversibility of stationary autoregressive processes in order to increase their power performance. Leybourne et al. (2005) and Pantula et al. (1995) provide evidence of superior performance of the weighted symmetric test statistic compared to the standard ADF test or the GLS-ADF test proposed by Elliot et al. (1996). The lag length employed in the ADF and Ws unit root tests has been selected by the Akaike Information Criterion (AIC). Results of the ADF and WS statistics are provided for the level, first differences and second differences of all the country specific domestic and foreign variables as well as global variables. When testing the levels, two types of regressions have been computed: one including both an intercept and a trend, and another including an intercept only. When testing first and second differences, only the intercept is included. Asymptotic 5% critical values for both statistics have been employed.

Table 3: Unit root test results for domestic variables at the 5% significance level

Domestic Variables	Statistic	Critical Value	RWANDA	KENYA	UGANDA	TANZANIA	BURUNDI
inf (with trend)	ADF	-3.45	-5.4171	-4.1712	-3.9697	-6.3820	-4.8858
inf (with trend)	WS	-3.24	-5.7648	-4.4711	-5.0028	-6.6417	-5.1293
inf (no trend)	ADF	-2.89	-5.4186	-4.2241	-3.5906	-6.3971	-4.9299
inf (no trend)	WS	-2.55	-5.7941	-4.5176	-4.5980	-6.6598	-5.1691
Dinf	ADF	-2.89	-5.8135	-6.7319	-6.8172	-6.7142	-8.5082
Dinf	WS	-2.55	-6.1495	-7.0027	-3.4995	-7.0892	-8.7924
DDinf	ADF	-2.89	-6.2798	-6.2581	-4.0608	-8.0522	-6.6873
DDinf	WS	-2.55	-6.6166	-6.6947	-4.3282	-8.3351	-7.1333
exc (with trend)	ADF	-3.45	-3.6331	-6.0477	-5.1100	-5.9118	-4.4346
exc (with trend)	WS	-3.24	-2.3780	-6.1804	-4.7805	-6.0616	-4.4125
exc (no trend)	ADF	-2.89	-4.1532	-6.0651	-5.0802	-5.8028	-4.4635
exc (no trend)	WS	-2.55	-2.1383	-6.2273	-4.8339	-6.0020	-4.2994
Dexc	ADF	-2.89	-5.6601	-6.9817	-7.3188	-5.5866	-6.6787
Dexc	WS	-2.55	-5.5670	-7.3460	-7.6653	-5.4044	-5.7036
DDexc	ADF	-2.89	-7.1903	-7.2466	-7.5913	-7.2045	-8.0991
DDexc	WS	-2.55	-7.2784	-7.7113	-7.9295	-7.5641	-8.8900
int (with trend)	ADF	-3.45	-6.7020	-4.5298	-5.0533	-5.7577	-5.7907
int (with trend)	WS	-3.24	-6.3029	-4.5884	-4.9363	-5.4614	-5.9809

int (no trend)	ADF	-2.89	-6.4890	-4.4180	-4.8196	-5.1851	-5.8335
int (no trend)	WS	-2.55	-6.5372	-4.2741	-4.9275	-5.1624	-5.9970
Dint	ADF	-2.89	-9.7974	-7.4423	-6.0772	-6.5831	-7.5631
Dint	WS	-2.55	-6.4099	-7.5136	-5.8732	-6.4763	-8.0331
DDint	ADF	-2.89	-10.363	-5.9243	-7.0431	-9.9380	-10.8254
DDint	WS	-2.55	-7.7119	-6.0285	-6.9330	-9.8297	-11.3713

Table 4: Unit Root Tests for the Foreign Variables at the 5% Significance Level

Foreign variable	Statistic	Critical value	RWANDA	KENYA	UGANDA	TANZANIA	BURUNDI
infs (trend)	ADF	-3.45	-3.9697	-3.9697	-3.7402	-3.9697	-3.9697
infs (trend)	WS	-3.24	-5.0028	-5.0028	-3.9677	-5.0028	-5.0028
infs (no trend)	ADF	-2.89	-3.5906	-3.5906	-3.7612	-3.5906	-3.5906
infs (no trend)	WS	-2.55	-4.5980	-4.5980	-3.9956	-4.5980	-4.5980
Dinfs	ADF	-2.89	-6.8172	-6.8172	-6.7574	-6.8172	-6.8172
Dinfs	WS	-2.55	-3.4995	-3.4995	-7.0235	-3.4995	-3.4995
DDinf	ADF	-2.89	-4.0608	-4.0608	-6.2533	-4.0608	-4.0608
DDinf	WS	-2.55	-4.3282	-4.3282	-6.9302	-4.3282	-4.3282
excs (trend)	ADF	-3.45	-5.2984	-4.4104	-5.7549	-5.8798	-5.0718
excs (trend)	WS	-3.24	-5.3574	-4.5269	-5.8358	-5.9334	-5.1650
excs (no trend)	ADF	-2.89	-5.3364	-4.4483	-5.8081	-5.9114	-5.1213
excs (no trend)	WS	-2.55	-5.4098	-4.5199	-5.8917	-5.9900	-5.2099
Dexcs	ADF	-2.89	-7.4421	-6.2559	-6.7551	-7.3853	-6.9980
Dexcs	WS	-2.55	-7.8599	-6.6718	-7.1413	-7.7658	-7.4242
DDexc	ADF	-2.89	-7.4314	-6.7954	-7.0989	-7.3335	-7.1169
DDexc	WS	-2.55	-7.8816	-7.0697	-7.5909	-7.8271	-7.5565
ints (trend)	ADF	-3.45	-3.4725	-5.11	-4.3480	-4.0247	-3.5720
ints (trend)	WS	-3.24	-3.6872	-4.6388	-4.5056	-4.1784	-3.7095
ints (no trend)	ADF	-2.89	-3.2560	-4.8700	-4.1904	-3.9158	-3.2296
ints (no trend)	WS	-2.55	-3.4911	-4.6134	-4.1787	-3.9041	-3.4706
Dints	ADF	-2.89	-6.7967	-5.8527	-7.8125	-7.0893	-6.9757

Dints	WS	-2.55	-6.9505	-5.2322	-7.9063	-7.1888	-7.1028
DDint	ADF	-2.89	-8.6701	-7.4101	-7.9962	-6.5924	-6.7629
DDint	WS	-2.55	-8.6134	-6.5496	-8.3707	-7.0019	-6.8185

Table 5: unit root test for the global variable at the 5% significance level

Global Variables	Test	Critical Value	Statistic
poil (with trend)	ADF	-3.45	-5.2370335
poil (with trend)	WS	-3.24	-5.4398825
poil (no trend)	ADF	-2.89	-5.0521658
poil (no trend)	WS	-2.55	-5.2925515
Dpoil	ADF	-2.89	-6.8166702
Dpoil	WS	-2.55	-7.1716318
DDpoil	ADF	-2.89	-6.4942482
Dpoil	WS	-2.55	-6.9426192

The 95% critical values are indicated in the third column for regressions with and with no trend. The unit root hypothesis at the 5% level of significance is rejected for all domestic, foreign and the global variables.

6.3.3 Lag Orders and Statistics of Individual VARY* Models

The lag orders $\mathrm{p_i}$ and $\mathrm{q_i}$ of the domestic and foreign variables respectively of the individual VARY $(\mathrm{p_i}, \mathrm{q_i})$ models is selected using the Akaike Information Criterion (AIC), the Schwartz Bayesian criterion (SBC) and the log likelihood statistics.

Where the first two terms in equations (5.1) and (5.2) correspond to the maximized value of the log likelihood function, $\widehat{\sum}_i = \sum_{t=1}^{T} \hat{u}_{it} \hat{w'}_{it} / T$ computed based on the estimated residuals $\hat{u}_{it}$ obtained from estimation of the individual VARY models given by equation (3.12), T is the sample size, $|.|$ is the determinant of $\hat{\sum}_i$, k_i and k_i^* are the number of domestic and foreign variables respectively in the individual models and $s_i = k_i p_i + k_i^* q_i + 2$. The model with the highest AIC or SBC value is chosen.

In addition, the F-statistics for testing the residual serial correlation of the individual VARY model equations are computed with the corresponding 95% critical values and degrees of freedom.

Table 6: Choice Criteria for VARY order and serial correlation results

	p	q	AIC	SBC	Loglik	Df	Fcrit_0.05	Inf	Exc	Int	poil
RWANDA	1	1	326.83	289.16	365.83	F(4,34)	2.65	16.88	0.26	4.96	
	2	1	331.81	285.45	379.81	F(4,31)	2.68	12.46	0.72	3.80	
	2	2	334.07	276.12	394.07	F(4,27)	2.73	8.57	0.52	6.22	
	3	1	333.52	278.46	390.52	F(4,28)	2.71	14.30	1.74	3.04	
	3	2	337.38	270.73	406.38	F(4,24)	2.78	11.04	0.97	4.48	
	3	3	346.47	268.23	427.47	F(4,20)	2.87	2.12	1.44	6.38	
	4	1	347.63	283.88	413.63	F(4,25)	2.76	3.69	1.60	0.83	
	4	2	352.36	277.01	430.36	F(4,21)	2.84	3.04	0.46	2.43	
	4	3	362.53	275.59	452.53	F(4,17)	2.96	0.97	0.38	2.04	
KENYA	1	1	272.92	226.56	320.92	F(4,35)	2.64	3.24	1.36	0.60	30.80
	2	1	267.42	205.60	331.42	F(4,31)	2.68	4.18	0.17	0.15	35.59
	2	2	279.14	205.73	355.14	F(4,28)	2.71	2.58	0.77	0.07	17.13
	3	1	269.94	192.67	349.94	F(4,27)	2.73	3.12	0.18	0.65	25.20

	3	2	284.31	195.44	376.31	F(4,24)	2.78	2.43	0.72	0.18	17.32
	3	3	289.93	189.47	393.91	F(4,21)	2.84	1.22	0.41	0.72	32.76
	4	1	273.26	180.53	369.26	F(4,23)	2.80	1.70	0.66	1.64	23.32
	4	2	286.35	182.03	394.35	F(4,20)	2.87	1.37	2.29	2.93	10.82
	4	3	288.05	172.14	408.05	F(4,17)	2.96	1.09	1.87	1.65	20.59
UGANDA	1	1	-683.5	-721.2	-644.5	F(4,34)	2.65	2.95	2.59	0.73	
	2	1	-687.9	-734.2	-639.9	F(4,31)	2.68	2.55	2.68	1.25	
	2	2	-689.7	-747.6	-629.7	F(4,27)	2.73	10.86	1.30	0.97	
	3	1	-688.2	-743.2	-631.2	F(4,28)	2.71	4.27	1.50	1.00	
	3	2	-689.7	-756.3	-620.7	F(4,24)	2.78	5.79	1.75	1.07	
	3	3	-685.0	-763.2	-603.9	F(4,20)	2.87	21.08	3.26	0.22	
	4	1	-689.3	-753.1	-623.3	F(4,25)	2.76	3.65	1.37	1.39	
	4	2	-688.1	-763.4	-610.1	F(4,21)	2.84	4.86	2.07	1.41	
	4	3	-680.2	-767.1	-590.2	F(4,17	2.96	5.27	4.93	0.46	
TANZANIA	1	1	281.89	244.22	320.89	F(4,34)	2.65	0.46	0.28	0.21	
	2	1	277.06	230.70	325.06	F(4,31)	2.68	0.53	0.31	0.26	
	2	2	282.10	224.17	342.10	F(4,27)	2.73	0.42	0.70	0.24	
	3	1	271.77	216.71	328.77	F(4,28)	2.71	0.91	0.54	1.27	
	3	2	276.54	209.90	345.54	F(4,24)	2.78	0.57	1.12	0.33	
	3	3	278.94	200.70	359.94	F(4,20)	2.87	2.06	0.86	1.29	
	4	1	272.50	208.75	338.50	F(4,25)	2.76	0.69	0.77	1.27	
	4	2	276.45	201.11	354.45	F(4,21)	2.84	0.83	1.21	0.63	

	4	3	278.40	191.47	368.40	F(4,17	2.96	3.30	0.81	0.30	
BURUNDI	1	1	248.14	210.47	287.14	F(4,34)	2.65	2.29	0.41	0.87	
	2	1	243.14	196.77	291.14	F(4,31)	2.68	2.04	1.30	0.75	
	2	2	253.42	195.46	313.42	F(4,27)	2.73	1.12	2.44	0.62	
	3	1	243.46	188.40	300.46	F(4,28)	2.71	1.68	2.60	1.07	
	3	2	254.80	188.15	323.80	F(4,24)	2.78	0.78	2.96	2.55	
	3	3	257.09	178.86	338.09	F(4,20)	2.87	0.86	2.49	0.36	
	4	1	244.15	180.40	310.15	F(4,25)	2.76	0.57	2.43	1.59	
	4	2	253.57	178.23	331.57	F(4,21)	2.84	0.31	4.20	1.54	
	4	3	255.67	168.74	345.67	F(4,17	2.96	0.45	5.92	1.63	

Table 7: VARY Lag order results used in the analysis (p: lag order of domestic variable; q: lag order of foreign variable)

	RWANDA	KENYA	UGANDA	TANZANIA	BURUNDI
p	4	3	4	2	3
q	3	3	3	2	3

This table contains the lag orders of the domestic and foreign variables used in the analysis selected using the criterion with the highest value.

6.3.4 Cointegration Tests

The rank of the cointegrating space for each country is computed using Johansen's trace and maximal Eigen value statistics as set out in Pesaran et al. (2000) for models with weakly exogenous I(1) regressors. The final selection of the rank orders is determined by the trace statistic, which in small samples is known to have better power properties than the maximal Eigen value statistic.

Table 8: Detailed Cointegration Results for the Maximum Eigenvalue Statistic at the 5% Significance Level

Country	Rwanda	Kenya	Uganda	Tanzania	Burundi
Number of endogenous variables	3	4	3	3	3
Number of foreign variables	4	3	4	4	4
$r = 0$	72.2582	69.5551	55.0400	51.3710	59.8783
$r = 1$	59.7675	47.1868	34.5418	32.7366	21.1041
$r = 2$	27.9424	22.0356	14.3360	19.4847	16.1002
$r = 3$		20.2550			

Table 9: Detailed Cointegration Results for the Trace Statistic at the 5% Significance Level

Country	RWANDA	KENYA	UGANDA	TANZANIA	BURUNDI
Number of endogenous variables	3	4	3	3	3
Number of foreign (star) variables	4	3	4	4	4
$r = 0$	159.9681	159.0324	103.9178	103.5923	97.0826
$r = 1$	87.7099	89.4773	48.8778	52.2213	37.2044
$r = 2$	27.9424	42.2906	14.3360	19.4847	16.1002
$r = 3$		20.2550			

Table 10: Critical Values for Trace Statistic at the 5% Significance Level (MacKinnon, Haug, Michelis, 1999)

Country	RWANDA	KENYA	UGANDA	TANZANIA	BURUNDI
Number of endogenous variables	3	4	3	3	3
Number of foreign (star) variables	4	3	4	4	4
$r = 0$	71.56	91.81	71.56	71.56	71.56
$r = 1$	45.9	64.54	45.9	45.9	45.9
$r = 2$	23.63	41.03	23.63	23.63	23.63
$r = 3$		20.98			

Table 11: Cointegrating Relationships for the Individual VARY models

Country	Rwanda	Kenya	Uganda	Tanzania	Burundi
Number of cointegrating relations	3	3	2	2	1

6.3.5 Weak Exogeneity Tests

The main assumption underlying the estimation of the individual country VARY models is the weak exogeneity of the foreign variables. This assumption is compatible with a certain degree of weak dependence across u_{it} as discussed in Pesaran et al (2004). A formal test of this assumption for the country specific foreign variables and the observed global variables is carried out as described in Johansen (1992) andHarboet al. (1998). Testing for weak exogeneity involves the marginal model of the foreign variables.

The weak exogeneity test in this work contains the F statistics for testing the weak exogeneity of the foreign variables. The test statistics have been generated with the critical values at 5% level of significance and the given degrees of freedom as shown in table 12 below.

Table 12: Test for weak exogeneity at 5% significance level

Country	F test	Fcrit_0.05	infs	excs	ints	poil
Rwanda	F(3,6)	4.7571				
Kenya	F(3,18)	3.1599	5.9758	4.5338	0.6712	
Uganda	F(2,7)	4.7374				
Tanzania	F(2,33)	3.2849	0.1098	0.7379	1.2047	0.7937

Burundi	F(1,20)	4.3512	0.2617	3.2293	4.7369	0.1106

The weak exogeneity assumption is not rejected for most of the foreign variables, despite some exceptions. In particular, the assumption is rejected at the 5% significance level for Kenyan inflation. Therefore, given that only 1 out of 19 foreign variables fail to satisfy the weak exogeneity assumption, we consider these outcomes as acceptable, thereby justifying the estimation procedure of each country model in the GVAR.

6.3.6 Contemporaneous Effects of Foreign Variables on Domestic Counterparts

The contemporaneous effects of foreign variables on their domestic counterparts are provided together with t-ratios computed based on standard, as well as White and Newey-West adjusted variance matrices. These contemporaneous effects are given by the estimated coefficients on the contemporaneous foreign variables and can be interpreted as impact elasticities between domestic and foreign variables. They are particularly informative as regards the international linkages between the domestic and foreign variables. High elasticities between domestic and foreign variables imply strong co-movements between the two. In addition to these coefficient estimates, standard errors and t-values are also calculated. White's heteroskedasticity robust and Newey-West heteroskedasticity and autocorrelation consistent standard errors as well as the corresponding t-values are also computed. The results are listed in table 13 below.

Table 13: Contemporaneous effects of foreign variables on their domestic counterparts

		inf	exc	int
RWANDA	Coefficient	4.351×10^{-9}	0.003841	-0.0939
	Standard error	2.943×10^{-9}	0.032030	0.1419
	t-Ratio	1.4786474	0.119930	-0.6618
	White's Adjusted SE	8.013×10^{-10}	0.029187	0.1516
	t-Ratio	5.4303699	0.131615	-0.6194
	Newey-West's Adjusted SE	7.73×10^{-10}	0.026376	0.1467
	t-Ratio	5.6264517	0.145642	-0.6403
KENYA	Coefficient	6.76×10^{-10}	0.8312	-0.5008
	Standard error	8.95×10^{-10}	0.1945	0.2975
	t-Ratio	0.755774	4.2737	-1.6835
	White's Adjusted SE	5.44×10^{-10}	0.2737	0.2851
	t-Ratio	1.2439378	3.0365	-1.7566
	Newey-West's Adjusted SE	4.67×10^{-10}	0.2995	0.2848
	t-Ratio	1.4472997	2.7755	-1.7582
UGANDA	Coefficient	22100185	0.9492	-0.0659
	Standard error	44139876	0.1379	0.1725
	t-Ratio	0.5006853	6.8822	-0.3821
	White's Adjusted SE	27955489	0.1438	0.1406
	t-Ratio	0.7905491	6.5998	-0.4689
	Newey-West's Adjusted SE	29982063	0.1279	0.1373
	t-Ratio	0.7371136	7.4219	-0.4803
TANZANIA	Coefficient	1.197×10^{-8}	0.0552	0.0996
	Standard error	1.156×10^{-8}	0.0843	0.1318
	t-Ratio	1.0353234	0.6544	0.7558
	White's Adjusted SE	8.435×10^{-9}	0.0632	0.1128
	t-Ratio	1.4186784	0.8723	0.8828
	Newey-West's Adjusted SE	8.455×10^{-9}	0.0499	0.0849
	t-Ratio	1.4153838	1.1054	1.1733

BURUNDI	Coefficient	3.069×10^{-9}	0.0716	0.2428
	Standard error	3.869×10^{-9}	0.0901	0.2808
	t-Ratio	0.7933594	0.7942	0.8646
	White's Adjusted SE	7.026×10^{-10}	0.0726	0.2691
	t-Ratio	4.3679583	0.9858	0.9023
	Newey-West's Adjusted SE	7.23×10^{-10}	0.0533	0.2652
	t-Ratio	4.2428634	1.3419	0.9156

The results in table 13 indicate that, the impact elasticities of all the variables are statistically significant for all the countries. All the values are positive but lower than one. For a given country, impact elasticities lower than one indicate that the domestic variables do not overreact to a variation in the foreign variable of its trade partners, while an impact elasticity greater than one indicate that the domestic variables overreacts to a variation in the foreign variables of the corresponding trade partners. Moreover, these findings give us already some insights with respect to the dynamics of the GIRFs: there is no evidence of strong international linkages across countries.

6.3.7 Average Pairwise Cross-Section Correlations

One of the key assumptions of the GVAR modeling approach is that the 'idiosyncratic' shocks of the individual country models should be cross-sectionally 'weakly correlated'. So that $cov(y_{it}{}^{*}, u_{it}) \rightarrow 0, with\ N \rightarrow \infty$, and as a result the weak exogeneity of the foreign variables is ensured. Direct tests of weak exogeneity assumptions discussed earlier indirectly support the view that the idiosyncratic shocks could only be weakly correlated. The output in this section provides direct evidence on the extent to which this is likely to be true. The basic idea is that by conditioning the country-specific models on weakly exogenous foreign variables, viewed as proxies for the common unobserved global factors, it is reasonable to expect that the degree of correlation of the remaining shocks across countries/regions will be modest. These residual interdependencies could reflect, for example, policy and trade spillover effects.

A simple diagnostic of the extent to which the country specific-foreign variables have been effective in reducing the cross-section correlation of the variables in the GVAR model is provided by the average pairwise cross-section correlations for the levels and first differences of endogenous variables of the model, as well as those of associated residuals over the selected estimation period. These are computed as follows: for every country for each given variable, the pairwise correlation of that country with each of the remaining countries is computed and averaged across countries, Smith and Galesi (2011).

The results are reported in table 14 below

Table 14: Average pairwise cross-section correlations (variables and residuals)

Variable	Country	Levels	First differences	VECMY residuals
inf	RWANDA	0.229627	0.137225	0.143227
inf	KENYA	0.142006	0.044708	0.065463
inf	UGANDA	0.146446	0.196894	-0.047938
inf	TANZANIA	0.076313	0.005832	0.027205
inf	BURUNDI	0.238314	0.214816	0.171931
exc	RWANDA	0.156425	0.059393	-0.046310
exc	KENYA	0.217059	0.198116	-0.204706
exc	UGANDA	0.275157	0.236599	-0.038303
exc	TANZANIA	0.127288	0.081997	-0.003543
exc	BURUNDI	0.092029	0.110614	-0.031488
int	RWANDA	-0.00857	-0.026387	0.001811
int	KENYA	-0.00796	-0.029612	0.115923

int	UGANDA	-0.03486	-0.043307	0.008431
int	TANZANIA	0.020174	-0.046062	0.031504
int	BURUNDI	0.012512	0.024029	0.038705

Among the variables in levels, exchange rates appears to be the most correlated, with a maximum of 0.275157 for Uganda and a minimum of 0.092029 for Burundi. Moreover, with respect to variables in differences, we observe a fall in the degree of correlation. The VECMY residuals are obtained from the estimation of each VECMY* model, containing both the domestic and foreign variables. The VECMY residuals are generally weakly correlated and in some cases negatively weakly correlated for all the variables under study. This is a clear indication that the inclusion of the foreign variables in the country model estimation cleans the common factor among the variables, thereby yielding weakly correlated residuals. In this way, this condition allows us to simulate shocks which are mainly country-specific.

6.3.8 Structural Stability Tests

A number of structural stability tests similar to those considered by Stock and Watson (1996) are computed in order to detect the possibility of presence of breaks. Among them Ploberger and Kramer's (1992) maximal OLS cumulative sum (CUSUM) statistic, denoted by PK_{sup} and its mean square variant PK_{msq}. The PK_{sup}statistic is similar to the CUSUM test suggested by Brown et al. (1975), although the latter is based on recursive rather than OLS residuals. Also included are tests for parameter constancy against non-stationary alternatives proposed by Nyblom (1989) as well as sequential Wald type tests of a one-time structural change at an unknown change point. The latter include the Wald form of Quandt's (1960) likelihood ratio statistic (QLR), the

Mean Wald statistic (MW) of Hansen (1992) and Andrew's and Ploberger (1994), and the Andrew's and Ploberger (1994) Wald statistic based on the exponential average (APW). The heteroskedasticity-robust version of the Nyblom and sequential Wald tests is also presented.

The critical values of the tests are computed under the null of parameter stability, using the bootstrap samples obtained from the solution of the GVAR (p) model as described in Smith and Galesi (2011).

The results are tabulated in table 15 below.

Table 15: Structural Stability Tests: Statistics

Variables	inf	exc	int	poil
PK sup				
RWANDA	0.359932	0.904393	0.355965	0.695236
KENYA	0.536118	0.501499	0.451982	
UGANDA	0.706989	0.486239	0.464187	
TANZANIA	0.569243	0.454682	0.547295	
BURUNDI	0.565711	0.604124	0.796219	
PK msq				
RWANDA	0.02833	0.250192	0.02185	0.044807
KENYA	0.022784	0.054861	0.054056	
UGANDA	0.04612	0.040184	0.026889	
TANZANIA	0.034978	0.033828	0.057374	
BURUNDI	0.029425	0.050806	0.058843	
Nyblom				
RWANDA	0.886386	1.409229	1.017697	0.869498
KENYA	1.135407	2.024974	1.763694	
UGANDA	0.966677	1.216567	0.828311	
TANZANIA	2.717021	0.339499	0.548204	

BURUNDI	0.81393	0.867374	0.704142	
Robust Nyblom				1.392093
RWANDA	1.840368	1.657941	1.402177	
KENYA	1.546863	1.946568	2.066366	
UGANDA	1.587679	1.461125	1.237099	
TANZANIA	1.578382	1.110974	0.852562	
BURUNDI	1.37505	0.936735	0.67582	
QLR				14.24176
RWANDA	16.21293	16.04214	34.33437	
KENYA	39.15676	31.31936	67.75215	
UGANDA	120.7105	23.84811	15.85914	
TANZANIA	80.18234	11.82443	27.77492	
BURUNDI	27.07211	44.99649	22.3253	
Robust QLR				9.77673
RWANDA	18.43146	15.2405	12.89208	
KENYA	13.44944	16.67314	18.47729	
UGANDA	5.535181	13.84402	15.24781	
TANZANIA	15.85407	13.52576	10.44482	
BURUNDI	14.83571	12.08875	11.26889	
MW				3.277988
RWANDA	11.24115	9.970909	9.995908	
KENYA	28.46883	23.95472	43.88884	
UGANDA	16.39276	14.92725	10.54641	
TANZANIA	35.61475	5.024944	16.0033	
BURUNDI	16.82636	11.9782	14.62962	
Robust MW				5.317102
RWANDA	14.55056	10.41999	9.825153	
KENYA	9.452847	14.4076	12.92567	
UGANDA	4.617232	10.08207	9.441933	
TANZANIA	12.8078	4.131051	8.103186	
BURUNDI	11.78256	8.594726	9.315447	

APW RWANDA KENYA UGANDA TANZANIA BURUNDI	6.427152 17.25382 57.02306 37.38603 11.38738	6.141551 13.82969 10.20219 3.963395 19.17324	13.92634 31.87171 6.535157 11.60059 8.502685	3.944348
Robust APW RWANDA KENYA UGANDA TANZANIA BURUNDI	7.958227 5.280517 2.322421 7.104915 6.395804	6.077938 7.419622 5.857748 3.996355 4.55703	5.342661 7.200732 5.931122 4.270194 4.763835	2.946425

The structural break dates corresponding to the QLR statistic are given in the table below

Table 16: Break dates for structural stability tests

Variable / Country	inf	exc	int	poil
RWANDA	2003Q4	2004Q4	2004Q1	
KENYA	2008Q4	2007Q1	2005Q3	2003Q3
UGANDA	2003Q3	2007Q1	2005Q2	
TANZANIA	2004Q3	2006Q3	2005Q2	
BURUNDI	2006Q2	2003Q3	2006Q4	

The results reported in table 15 show that there is broad evidence in favor of the stability of the parameters. The main reason for the rejection seems to be breaks in the error variance as opposed to breaks in the parameter coefficients.

6.3.9 Eigen Values of the GVAR

The Eigen values of the GVAR model, denoted by $\lambda_{eig} = a \pm bi$, are computed as the eigen values of the companion matrix F given in equation 3.81 by solving the determinantal equation

$$\left|I_{kp}\lambda_{eig} - F\right| = 0$$

Their corresponding moduli are computed as $mod(\lambda_{eig}) = \sqrt{(a^2} + b^2)$. In the case of I(1), cointegrated variables the roots of the equation above should lie inside and at most on the unit circle.

The results indicate that our model is dynamically stable since the moduli of the 32 Eigen values of the F matrix in equation (3.19) are all on or within the unit circle. Specifically, the number of the Eigen values lying on the unit circle (i.e. the number of unitary roots) is 3. The stability of the GVAR model is another requirement as a sufficient condition to test for the validity of the GVAR model.

6.3.10 Country Specific Weights

Country specific weights are constructed using PPP-GDP figures. This is done by dividing the PPP-GDP figure of each country by the total sum across countries, such that the weights add up to one across the countries. These are used for computation of global shocks (i.e. shocks to a variable across all countries) in impulse response analysis and forecast error decompositions. The results are reported in table 17 below.

Table 17:Country Weights

Country	inf	exc	int	poil
KENYA	0.324806	0.324806	0.324806	1
RWANDA	0.191923	0.191923	0.191923	
UGANDA	0.186356	0.186356	0.186356	
TANZANIA	0.211879	0.211879	0.211879	
BURUNDI	0.085036	0.085036	0.085036	

6.3.11 Persistence Profiles

Persistence profiles (PPs) refer to the time profiles of the effects of system or variable-specific shocks on the cointegrating relations in the GVAR model, Pesaran and Shin (1996). PPs have a value of unity on impact, while they should tend to zero as the horizon $n \rightarrow \infty$, if the vector under the consideration is a valid cointegrating vector. They provide information on the speed with which the cointegrating relationship returns to their equilibrium states, that is, how long the system takes to revert to its steady-state path after being hit by a system wide shock.

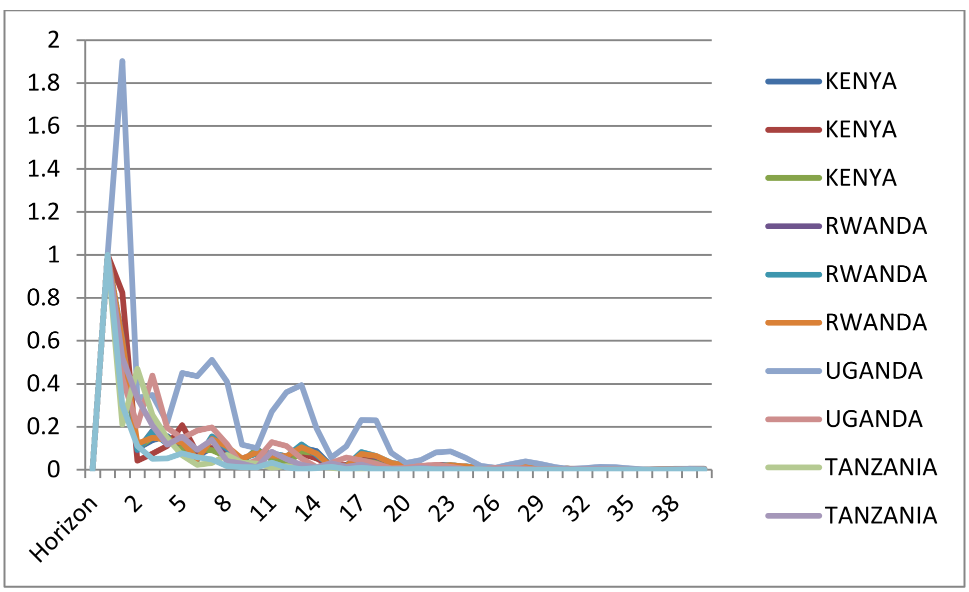


Figure 1: Persistence Profile of the Effect of System-Wide Shocks to the Cointegrating Relations of the GVAR Model

The figure above shows the absorption path of deviations from the steady state over a simulation of 40 quarters (10 years). In all cases we observe convergence towards the steady state, with the adjustment process being completed within the fourth year of the simulation.

6.3.12 Impulse Response Analysis

Impulse responses refer to the time profile of the effects of variable specific shocks or identified shocks (such as monetary policy or technology shocks, identified using a suitable economic theory) on the future states of a dynamical system and thus, on all the variables in the model. In this work different types of shocks are simulated. For instance, we simulate a negative global shock to a domestic variable, a shock to a global variable and a shock to domestic variables.

6.3.12.1 Generalized Impulse Response Functions

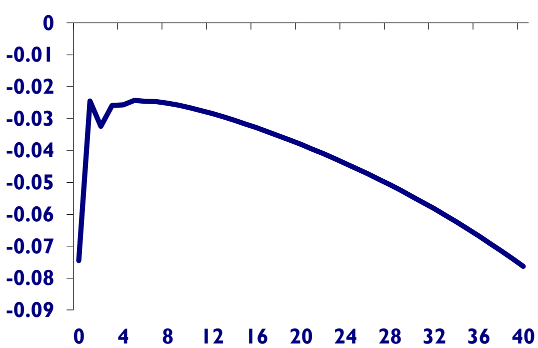
The impulse responses of shocks to specific variables considered for the GVAR model are the Generalized Impulse Response Functions (GIRFS), introduced in Koop et al. (1996) and adapted to VAR models in Pesaran and Shin (1998).

This relatively new approach differs in a number of ways from traditional Orthogonalized Impulse Responses (OIRs) in Sims (1980). First, it does not orthogonalize the residuals of the system, as it takes into account the historical correlations among the variables, summarized by the estimated variance-covariance matrix. For this reason, it does not require any a priori economic-based restrictions and its outcome is invariant to the ordering of the variables in the model. Second, since the shocks are not identified, the GIRFs cannot provide information about the causal relationships among the variables. This shortcoming limits the potential; applications of the GIRFs, especially for purposes of policy simulation. Nonetheless, GIRFs have a comparative advantage with respect to the traditional OIRs in the context of multi-country frameworks such as the GVAR model, Galesi and Sgherri (2009). In fact, they can provide interesting insights on how shocks internationally propagate, by unveiling potential linkages among different national economies. In addition, it is actually a difficult task to employ traditional OIRs in a GVAR, since there is no reasonable way to order the countries in the model.

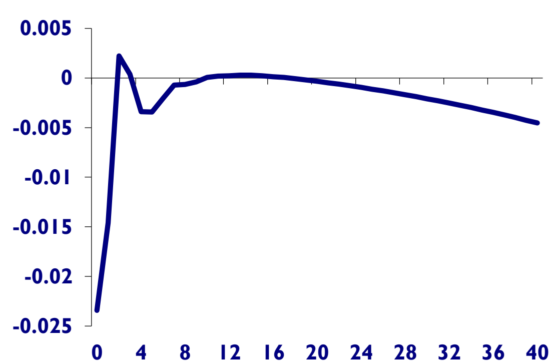
In our application, we analyze the dynamic properties of our GVAR model by simulating either a positive or a negative standard error shock to each country's variable. The scope of this simulation is to determine the degree of intercountry financial spillovers: in other words, we seek to analyze how each country responds to a specific shock.

For instance, the GIRFs associated to one standard error negative shock to Kenyan inflation on its partners' inflation are plotted in the figures below. For each region, the charts show the dynamic response of each variable over a time horizon of 10 years which has been used as our forecast horizon.

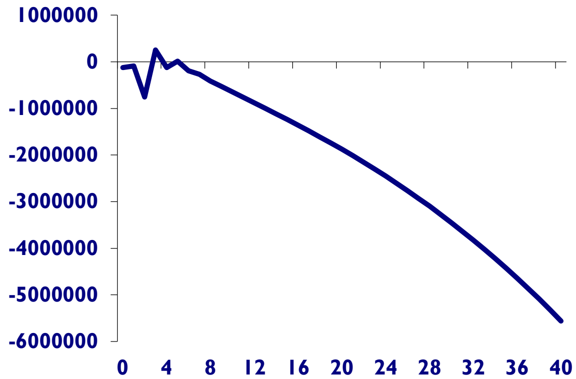
a) Kenya inflation



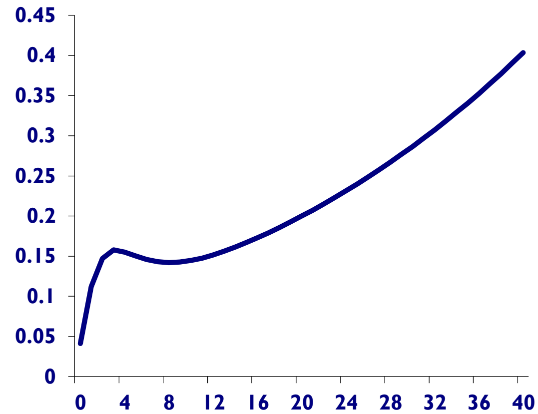
b) Rwanda inflation



c) Uganda inflation



d) Tanzania inflation



e) Burundi
inflation

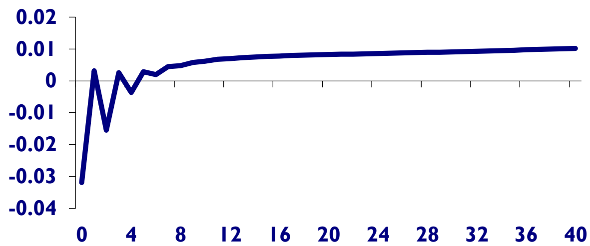
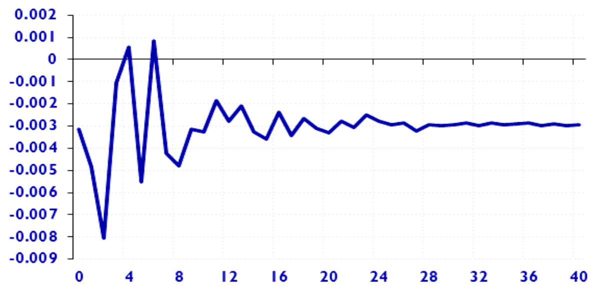


Figure 2: GIRFs; Response to one s.e. negative shock to Kenyan inflation

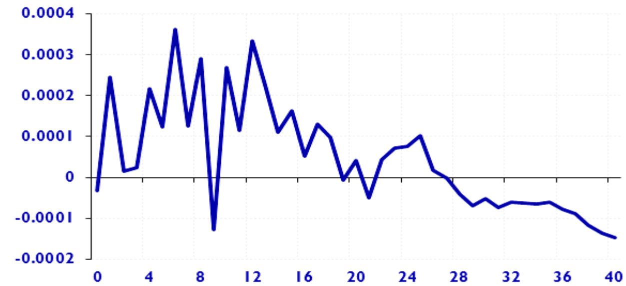
The graphs above indicate that Uganda and Tanzania have a significant response to a one standard error (s.e.) negative shock to Kenyan inflation as compared to Rwanda and Burundi. Rwanda and Burundi are only responding in the shortrun.

The graphs in figure 3, show the responses associated with exchange rates to one s.e shock to Kenyan inflation

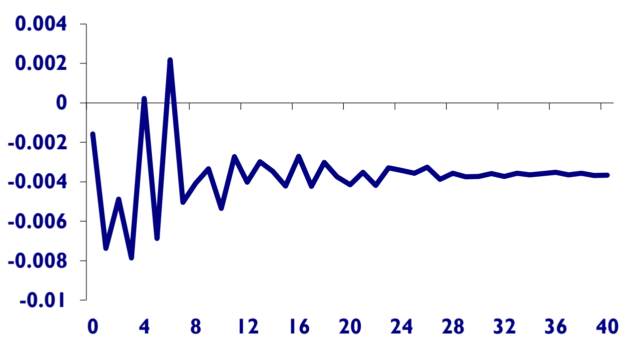
a) Kenya exchange rates



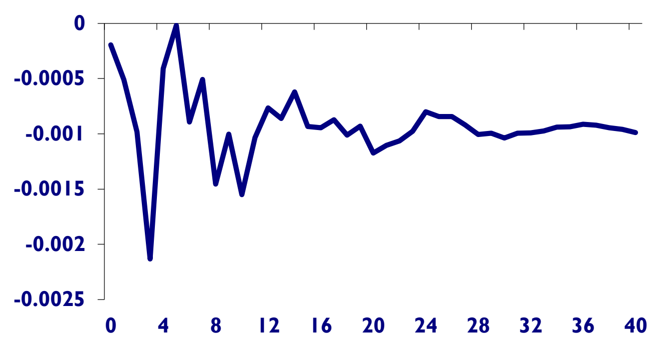
b) Rwanda exchange rates



c)Uganda exchange rates



d) Tanzania exchange rates



e)Burundi exchange rates

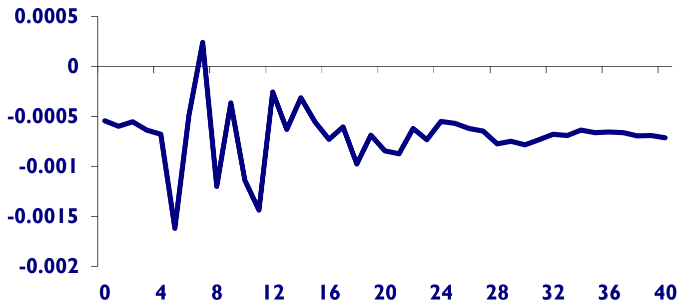
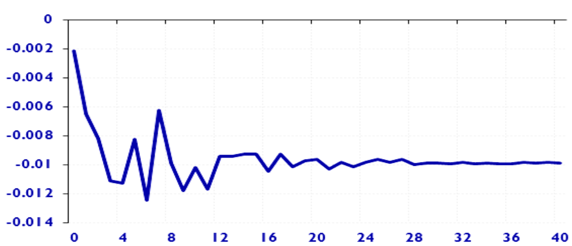


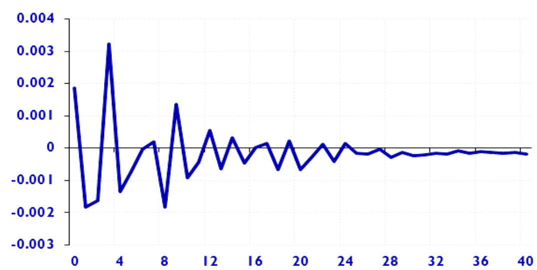
Figure 3: GIRFs; Response to one s.e shock to Kenyan inflation for exchange rates

The graphs indicate that there are strong fluctuations in GIRFs for Kenya, Uganda, Tanzania and Burundi in the shortrun but the trend stabilizes after 3 years. In the case of Rwanda, there are strong fluctuations for the first 3 years and a monotonic decrease in exchange rates' GIRFs in the longrun.

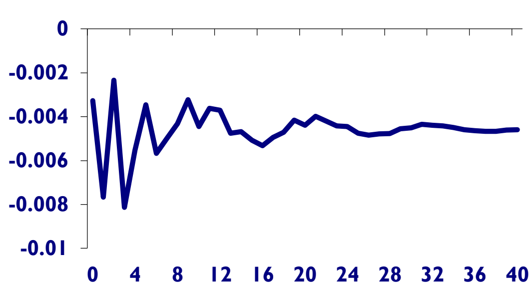
a) Kenya interest rates



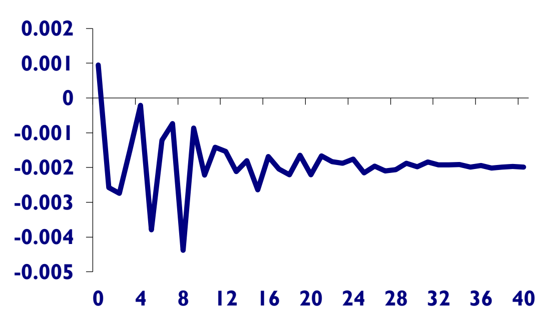
b) Rwanda interest rates



c)Uganda interest rates



d) Tanzania interest rates



e)Burundi interest rates

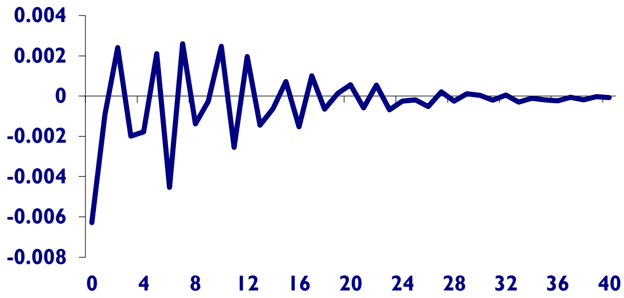
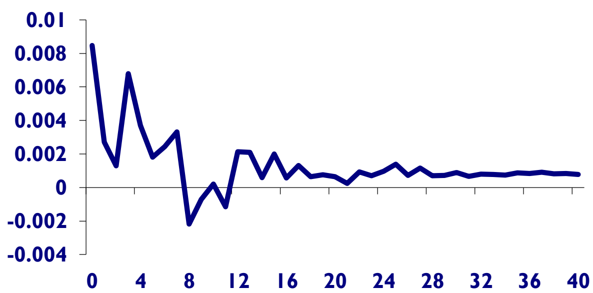


Figure 4: GIRFs; Response to one s.e shock to inflation for interest rates

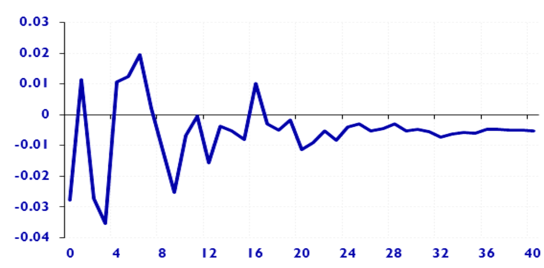
In this case, the graphs indicate that there are strong fluctuations in interest rates for the first three years but the trend stabilizes in the longrun. Moreover, there is a notable response that is observed for Kenya. The associated GIRFs monotonically decrease over the first year i.e. first four quarters but the trend thereafter is similar to the other countries.

Another form of shock simulated is a global shock to inflation. The results are represented in the graphs in figures 5, 6 and 7 for the stated variables.

a) Kenya inflation

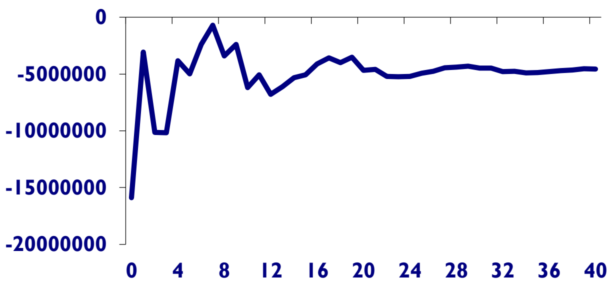


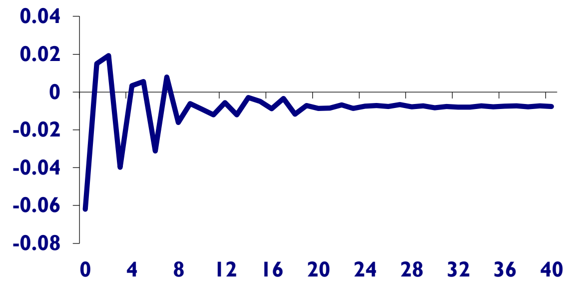
b) Rwanda inflation



c)Uganda inflation

d) Tanzania inflation





e)Burundi inflation

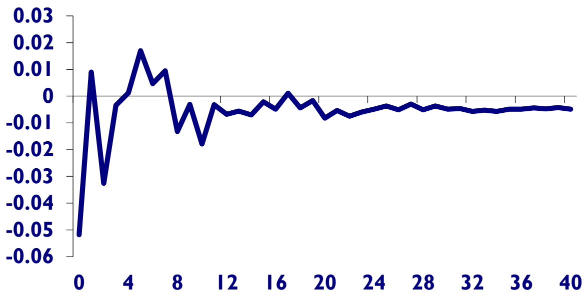
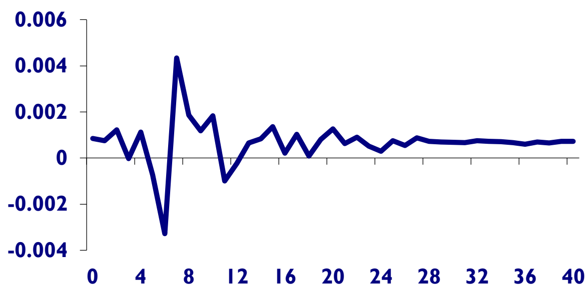


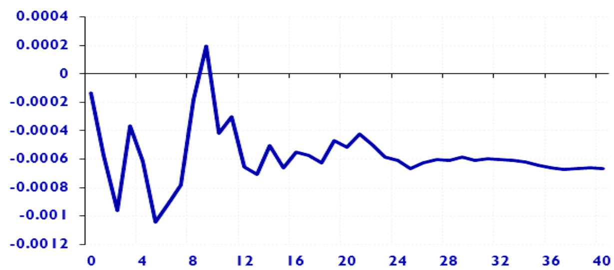
Figure 5: GIRFs; Response to one global s.e shock to inflation for inflation

The GIRFs for Kenyan inflation decreases for two years then stabilizes, while that of the other countries in the study keeps fluctuating for 2 years and then stabilizes.

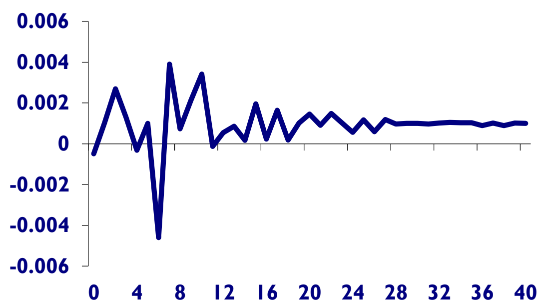
a) Kenya exchange rates

b) Rwanda exchange rates

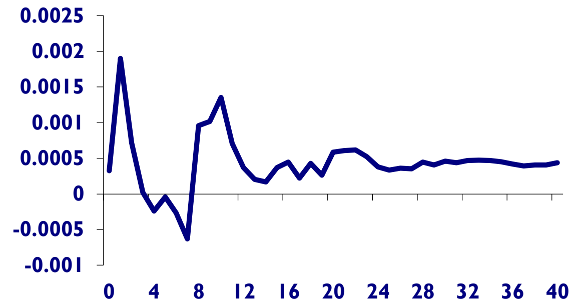




c)Uganda exchange rates



d) Tanzania exchange rates



e)Burundi exchange rates

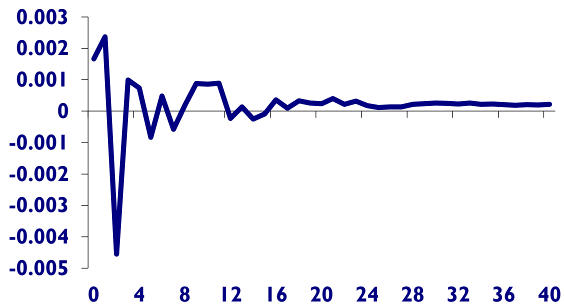
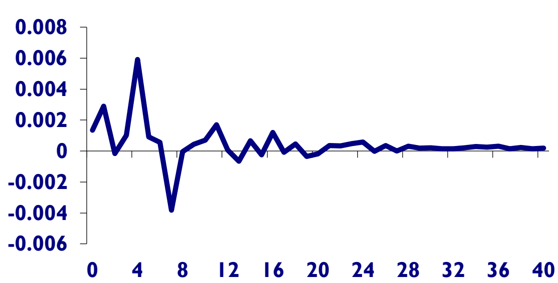


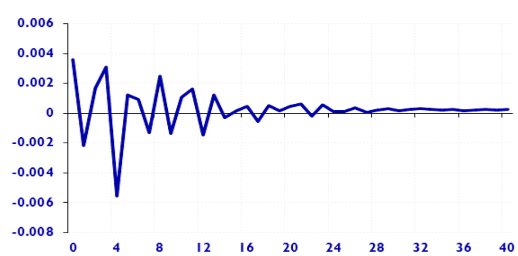
Figure 6:GIRFs; to one global s.e shock to inflation for exchange rates

For the case of exchange rates, there is a striking fluctuation in the GIRFs for all countries as shown in the graphs above. A similar trend is observed for the response in interest rates as shown in figure 7 below.

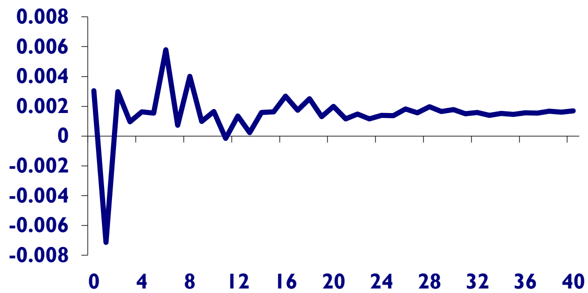
a) Kenya interest rates



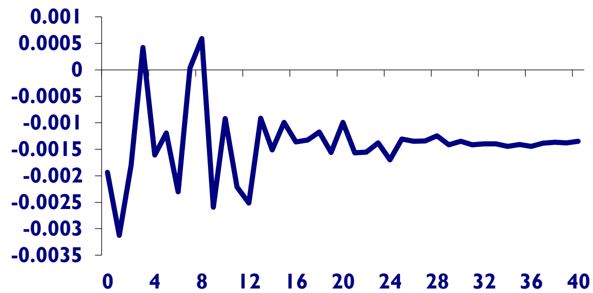
b) Rwanda interest rates



c)Uganda interest rates



d) Tanzania interest rates



e)Burundi interest rates

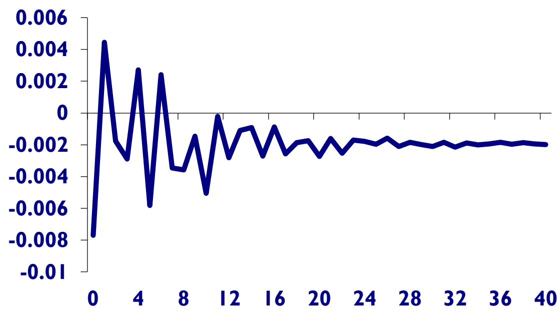


Figure 7: GIRFs; to one global s.e shock to inflation for interest rates

Other types of shocks simulated show similar trends to the ones discussed above, that is, sharp fluctuations in the short run-mostly 2 to 4 years and then stabilization in the long run.

6.3.12.2 Generalized Forecast Error Variance Decompositions (GFEVDs)

Traditionally the forecast error variance decomposition of a VAR model is performed on a set of orthogonalised shocks, whereby the contribution of the j^{th}orthogonalised innovation to the mean square error of the n-step ahead forecast of the model is calculated. In the case of the GVAR, the shocks across countries, that is u_{it} and $u_{st}$for $i \neq s$, are not orthogonal. In fact, there is evidence that on average, the shocks across countries are positively correlated, Smith and Galesi (2011). The standard application of the orthogonalised FEVD to the GVAR model is therefore not valid.

Results of the GFEVDs for a one negative shock on Kenyan exchange rates are reported in table 18 below.

Following a shock to the Kenyan exchange rates, we observe that among the Kenyan variables, exchange rates explain most of the forecast error variance in the short run. However, the relative contribution of exchange rates decreases over time, while the opposite is for the other Kenyan and non-Kenyan variables. Hence, we observe that if a shock is simulated, the variable which explains most of the variance of the shock in the short-term is the variable in which that shock is injected. On the contrary, in the longer term, the other domestic variables gain increasing relevance.

From a global perspective, we generally observe the same dynamic behavior just highlighted in the Kenyan case: the variable in which the shock is injected explains most of the forecast error variance for all countries over the short run; its relative importance decreases over time, while the opposite is observed for the rest of the variables.

Table 18**:**GFEVD; proportion of N-step ahead forecast Error variance of Kenyan exchange rates

Quarter		0	5	10	15	20	25	30	35	40
Kenya	Inf	0.017	0.108	0.128	0.141	0.160	0.177	0.0194	0.209	0.223
Kenya	Exc	0.577	0.304	0.254	0.238	0.228	0.219	0.210	0.203	0.196
Kenya	Int	0.001	0.126	0.108	0.101	0.097	0.093	0.090	0.087	0.084
Rwanda	Inf	0.010	0.013	0.013	0.013	0.014	0.014	0.015	0.015	0.016
Rwanda	Exc	0.012	0.010	0.009	0.009	0.009	0.009	0.008	0.008	0.008
Rwanda	Int	0.002	0.017	0.019	0.018	0.018	0.018	0.018	0.018	0.018
Uganda	Inf	0.001	0.004	0.028	0.029	0.029	0.029	0.029	0.030	0.030
Uganda	Exc	0.050	0.090	0.093	0.087	0.084	0.082	0.079	0.077	0.075

Uganda	Int	0.011	0.029	0.030	0.048	0.049	0.056	0.059	0.063	0.067
Tanzania	Inf	0.009	0.031	0.031	0.029	0.029	0.028	0.027	0.027	0.026
Tanzania	Exc	0.020	0.080	0.065	0.062	0.062	0.062	0.061	0.060	0.060
Tanzania	Int	0.005	0.011	0.012	0.014	0.013	0.014	0.014	0.014	0.014
Burundi	Inf	0.005	0.005	0.006	0.006	0.006	0.006	0.006	0.005	0.005
Burundi	Exc	0.007	0.014	0.014	0.014	0.015	0.015	0.016	0.016	0.016
Burundi	int	0.001	0.010	0.010	0.011	0.013	0.014	0.015	0.016	0.017

6.4 Performance of the Bayesian GVAR (B-GVAR) Forecasts Relative to the Benchmarks

In order to accomplish this objective, we have summarized the Root Mean Square Forecast Errors (RMSFEs) of all the benchmark models for all the variables. Recall that the selected benchmarks are random walk (RW) model with and without a drift and a univariate AR (1) model with and without a drift. The researchers have also included the standard GVAR model put forward in Dees et al. (2007) which is the standard model employed in developing our literature. The results are reported in table 19 below.

Table 19: relative forecasting performance, One-Quarter-Ahead: RMSFEs (Average across countries)

	B-GVAR	GVAR	AR (1)	AR (1) trend	RW	RW with a drift
Inflation	0.5088	0.5903	0.5321	0.5870	0.5531	0.5820
Exchange rates	0.5864	0.6020	0.7890	0.7156	0.6157	0.6075

Interest rates	0.2574	0.2601	0.1900	0.2481	0.1730	0.1901

We note that none of the benchmark models do better than the B-GVAR with a Minnesota prior in forecasting inflation and exchange rates on a One-Quarter-Ahead basis. The situation however is different in forecasting interest rates. In this case, the simple benchmark models tend to outperform the B-GVAR and the standard GVAR models.

Moreover, a notable result in this study is the poor performance of the standard GVAR model as proposed in Dees et al. (2007) over the simple models.

CHAPTER 7

Summary, Conclusions and Recommendations

7.1 Introduction

This chapter presents summary of findings, conclusions and recommendations for further research.

7.2 Summary of Major Findings

The researcher is aimed at studying financial interlinkages in five East African countries using the BMA Global Vector Autoregressive modeling approach. the specific objectives of the study were to estimate the VAR models for Kenya, Uganda, Tanzania, Rwanda and Burundi; to stack the VAR models obtained in order to obtain East Africa GVAR models; to determine the dynamic properties of the GVAR models obtained; to estimate forecasts in inflation rates, exchange rates and interest rates and to evaluate the performance of the Bayesian GVAR model obtained over the random walk with and without a drift and the univariate autoregressive models with and without a drift.

The researcher used trade weights in order to capture the foreign interlinkages and found out that the vast majority of the trade weights are granular, i.e. they are not too close to one. The largest weights were observed towards Uganda and Tanzania. This is one of the sufficient conditions of the validity of the GVAR methodology as discussed in PSW (2004a).

The study also used the AIC, SBC and log-likelihood statistics in order to select the lag orders $(p_i \text{ and } q_i)$ of the domestic and foreign variables respectively of the individual country models.

The lag order results used in the analysis is (4, 3) for Rwanda, (3, 3) for Kenya, (4, 3) for Uganda, (2, 2) for Tanzania and (3, 3) for Burundi. These lag orders were selected using the criterion with the highest value.

Another major finding is that the weak exogeneity assumption is not rejected for most of the foreign variables. In particular, the assumption is rejected at the 5% significance level for Kenyan inflation. Given that only 1 out of 19 foreign variables fail to satisfy the weak exogeneity assumption, this outcome justifies the estimation procedure of each country model in the GVAR model.

This study has also found out that exchange rates appears to be the most correlated variable with a maximum of 0.275157 for Uganda and a minimum of 0.092029 for Burundi. With respect to the variables in differences, there is a fall in the degree of correlation. The VECMY* residuals are generally weakly correlated for all the variables under study. This outcome is a clear indication that the inclusion of the foreign variables in the countries models estimation cleans out the common factor among the variables, thereby yielding weakly correlated residuals and hence allowing simulation of shocks which are mainly country specific.

The study also finds out that the GVAR model obtained is dynamically stable since the moduli of the Eigen values of the F matrix in equation 3.81 are all on or within the unit circle. This is another sufficient condition for testing the validity of the GVAR model.

In terms of shock simulation, the researchers found out that most countries respond to shocks in the short run, mostly 2-4 years but thereafter stabilizes, though there are notable exceptions. For instance, a response to a negative shock to Kenyan inflation as discussed in chapter 4 above.

In the case of forecast performance analysis, there are mixed results but the B-GVAR model with a Minnesota prior seems to outdo the simple models in most cases. The simple models perform better for interest rates in this case. However, there is a notable out come in this study whereby the standard GVAR model employed in our literature performs poorly in forecasting over the simple models. This outcome however conforms to most of the previous research such as Cuaresma et al (2014), Dees et al. (2007) and others.

7.3 Conclusion

From the findings of the study a B-GVAR model with a Minnesota prior is a worthy model since it improves forecasting in most of the variables. Moreover, the validity of the GVAR methodology is confirmed by the stability conditions and the 'granular' conditions of the trade weights being satisfied. The weak exogeneity assumption is not rejected and this further justifies the estimation procedure of each country model in the GVAR model.

7.4 Suggestions for Further Study

While implementing the Bayesian approach we had assumed a Minnesota prior for the VAR coefficients. This choice however ignored other priors which have been suggested in literature. It is necessary that future research should focus on comparing the performance of these priors in forecasting performance. Future research should also focus on the performance of B-GVAR models in forecasting over other simple Bayesian models, for instance Bayesian VARs and others.

Although the endogeneity of the foreign variables was taken into account in the empirical implementation of the GVAR model, it was ignored when theoretically estimating the model.

Future research should focus in estimating the GVAR models taking the endogeneity of the foreign variables into account.

REFERENCES

Alogoskoufis G. and Smith R. (1991). On error correction models: specification, interpretation, estimation, Journal of Economic Surveys, vol. 5, pg. 97-128.

Andrews D. W. K and Ploberger W. (1994). Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative, Econometrica, vol. 62, pg. 1383-1414.

Barnard G. A. (1963).New methods of quality control.Journal of Royal statistics and social services, A.

Bates J. M. and Granger C. W. J. (1969). The combination of forecasts.Journal of operational research quarterly. Vol. 20, pg. 451-468.

Bauwens L. and Rombouts J. V. K. (2007).Bayesian clustering of many GARCH models. Econometric reviews. Available online at http://www.tandfonline.com/bi/lecr20. January, 2014.

Blake A. and Mumtaz H. (2012). Applied Bayesian Econometrics for Central Bankers, centre for Central Banking Studies technical Handbook-No. 4. Bank of England.

Brown R. L., Durbin J. and Evans J. M. (1975). Techniques for testing the Constancy of Regression Relationships Over Time (with discussion), Journal of the Royal Statistical Society B, vol. 37, pg. 149-163.

Chatfield C. (1995). Model uncertainty, data mining and statistical inference (with discussion).Journal of Royal statistics and social services, vol. 158, pg. 419-466.

Clements M. P. and Hendry D. F. (1998).Forecasting Economic time series. Cambridge University Press, Cambridge,UK.

Cuaresma C. J., Feldkircher M. and Huber F. (2014).Forecasting with Bayesian Global Vector Autoregressive models: A Comparison of Priors.Working Paper Series No. 189, Oesterrechische National Bank.

deFinetti, B. (1972). Probability, Induction and Statistics.The Art of Guessing, Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York.

Dees, Stephane, Fillipo di Mauro, Pesaran H. M. and Smith L. V. (2007).Exploring the international linkages of the euro area: a global VAR analysis, Journal of Applied Econometrics, vol. 22, No. 1, pg. 1-38.

Draper D. (1995).Assessment and propagation of model uncertainty.Journal of royal statistics and social services, vol. 57, pg. 45-97.

Edgerton, H. A. and Kolbe, L. E. (1936).The method of minimum variation for the combination of criteria.Psychometrika, vol. 1, pg. 183-187.

Elliot G., Rothenberg T. J. and Stock J. H. (1996). Efficient Tests for an Autoregressive Unit Root, Econometrica, vol. 64, pg. 813-836.

Garcia R. and Perron P. (1996). An analysis of the real interest rate under regime shifts. Review of economics and Statistics, vol. 78, pg. 111-125.

Galesi A. and Sgherri S. (2009). Regional financial spillovers across Europe: a global VAR analysis. IMF working paper 09/23.

Geisser S. (1965). Bayes approach for combining correlated estimates. Journal of American statistical association, vol. 60, no.310, pg. 602-607.

Geisser, S. and Eddy, W. F. (1979).A predictive approach to model selection.Journal of the AmericanStatistical Association, vol. 74, pg.153–160.

Hansen B. E. (1992). Tests for Parameter Instability in Regressions with I(1) Processes, Journal of Business and Economic Statistics, vol. 10, pg. 321-336.

Hansen B. E. (2007). Least squares model averaging. Econometrica, Economic society, vol. 75, No. 4, pg. 1175-1189.

Harbor I., Johansen S., Nielsen B. and Rahbek A. (1998). Asymptotic Inference on Cointegrating Rank in Partial Systems, Journal of Business and Economic statistics, vol. 16, pg. 388-399.

Hjort N. L. and Claeskens G. (2003).Frequentist model averaging estimators.Statistical research report.University of Oslo.

Hoeting J. A. Madigan D., Raftery A. E. and Volinsky C. T. (1999). Bayesian model averaging: a tutorial. Journal of statistical Science, vol. 14, No. 4, pg.382-417.

Horn R. A. and Johnson C. R. (1985). Matrix Analysis. Cambridge: Cambridge University Press.

Johansen S. (1992). Cointegration in Partial Systems and the Efficiency of the Single-Equation Analysis, Journal of Econometrics, vol. 52, pg 231-254.

Johnston J. and Dinardo J. (1996), Econometrics Methods, 4th Ed. McGraw Hill, Newyork, USA.

Kass R. E. and Raftery A. E (1995).Bayes Factors. Journal of American Statistics Association, vol. 90, pg. 773-795.

Koop G., Pesarn M. H. and Potter S. M. (1996). Impulse Response Analysis in Nonlinear Multivariate Models, Journal of Econometrics, vol. 74, pg. 199-147.

Leybourne S., Kim T.-H. andNewbold P. (2005). Examination of Some more Powerful modifications of the Dickey-Fuller Test, Journal of Time Series Analysis, vol. 26, pg. 355-369.

Mann H. B. and Wald A. (1943).On the Statistical Treatment of Linear Stochastic Difference Equations, Econometrica, vol. 11, 173-220.

Miller A. J. (1990). Subset selection in regression.Chapman and Hall. London.

Mutl J. (2006). Dynamic Panel Data Models with Spatially Correlated Disturbances.PhD Thesis, University of Maryland, College Park.

Newey W. K. and West D. K. (1987). A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica, vol. 55, 703-708.

Nyblom J. (1989). Testing for the Constancy of Parameters over Time, Journal of the American Statistical Association, vol. 84, pg. 223-230.

Pantula S. G., Gonzalea-Farias G. and Fuller W. A. (1994). A Comparisson of Unit-Root Test Criteria, Journal of Business and Economic Statistics, vol. 12, pg. 449-459.

Park H. J. and Fuller W. A. (1995). Alternative Estimators and Unit Root Tests for the Autoregressive Process, Journal of Time Series Analysis, vol. 16, pg. 415-429.

Pesaran H. M. and Smith R. (2006).Macro-economic modeling with a global perspective. Manchester School, University of Manchester, vol. 74, pg. 24-49.

Pesaran M. H., Schuermann T. and Smith L. V. (2008).Forecasting economic and financial variables with global VARs. Federal Reserve Bank of New York Staff Reports. Staff Report No. 317

Pesaran, M. H., Schuermann T. and Weiner S. M. (2004). Modeling regional interdependencies using a global correcting macro econometric model, Journal of Business and Economics Statistics, vol. 22, No. 2, pg. 129-162.

Pesaran M. H., Schuermann T. and Weiner S. M. (2002).Modeling Regional Interdependecies using a Global Error-Correcting Macroeconometric Model.Wharton Working Paper Series, 01-38-B.

Pesaran M. H., Shin Y. and Smith R. (2000). Structural Analysis of Vector Error Correction Models with Exogenous I(1) Variables, Journal of Econometrics, vol. 97, pg. 293-343.

Pesaran M. H. and Shin Y. (1998). Generalized Impulse Response Analysis in Linear Multivariate Models, Economics Letters, vol. 58, pg. 17-29.

Pesaran M. H. and Shin Y. (1996). Cointegration and the Speed of Convergence to equilibrium, Journal of Econometrics, vol. 71, pg. 117-143.

Ploberger W. and Kramer W. (1992).The CUSUM test with OLS residuals, Econometrica, vol. 60, pg.271-286.

Quandt R. (1960). Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes, Journal of the American Statistical Association, vol. 55, pg. 324-330.

Raftery A. E (1995).Bayesian model selection in social research (with discussion). In sociological and methodology 1995 (P. V. Marsden, ed), pg 111-195, Blackwell, Cambridge, MA.

Raftery A. E, Madigan D. and Volinsky C. T. (1997).Bayesian Model averaging for linear regression models.Journal of American statistical association, vol. 92, pg. 179-191.

Rao C. R. (1973). Linear Statistical Inference and its Applications, 2nd Edition, John Wiley, New York.

Roberts H. V. (1995). Probabilistic prediction.Journal of American statistical association, vol. 60, pg. 50-62.

Sims C. (1980). Macroeconomic and Reality, Econometrica, vol. 48, pg. 1-48.

Smith L. V. and Galesi A. (2011). GVAR Toolbox 1.1, www-cfap.jbs.cam.ac.uk/research/gvartoolbox/index.html. June, 2014.

Stigler G. (1973). General economic conditions and national elections. The American economic reiew, vol. 63, issue 2, pg. 160-167.

Stock J. H and Watson M. W. (1996).Evidence on structural instability in macroeconomic time series relations.Journal of business and economics statistics, vol. 14, pg. 11-30.

Timmermann A. (2006). An evaluation of the world economic outlook forecasts.IMF working paper 06/59.

White H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, Econometrica, 48, 817-838.

Zellner A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias, Journal of the American Statistical Association, vol. 57, pg. 348-368.

Appendices

A: Vectors and Matrices

A1) The Kronecker product

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

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{bmatrix}$$

is the Kronecker product or direct product of A and B.

Rules:-Assuming suitable dimensions for the matrices

1) $A \otimes B \neq B \otimes A$
2) $(A \otimes B)' = A' \otimes B'$
3) $A \otimes (B + C) = (A \otimes B) + (A \otimes C)$
4) $(A \otimes B)(C \otimes D) = AC \otimes BD$
5) If A and B are invertible, then $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
6) If A and B are square matrices with eigen values $\lambda_A \; and \; \lambda_B$ respectively and associated eigen vectors $v_A \; and \; v_B$ then $\lambda_A \lambda_B$ is an eigen value of $A \otimes B$ with eigen vector $v_A \otimes v_B$
7) If A and B are $(m \times m)$ and $(n \times n)$ square matrices respectively, then $|A \otimes B| = |A|^n |B|^m$
8) If A and B are square matrices $tr\,(A \otimes B) = tr(A)tr(B)$

A2) Thevec and vech operators and related matrices

- **the operators**

Let $A = (a_1, \dots, a_n)$ be an $(m \times n)$ matrix with $(m \times 1)$columns a_i. The vec operator transforms A into an$(mn \times 1)$ vector by stacking the columns, that is

$$vec\ (A) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

Rules:-let A, B and C be matrices with appropriate dimensions then

1) $vec\ (A + B) = vec(A) + vec(B)$
2) $vec(ABC) = (C' \otimes A)vec(B)$
3) $vec\ (AB) = (I \otimes A)vec(B) = (B' \otimes I)vec(A)$
4) $vec(ABC) = (I \otimes AB)vec\ (C) = (C'B' \otimes I)vec(A)$
5) $vec(B')vec(A) = tr(BA) = tr(AB) = vec(A')'vec(B)$
6) $tr(ABC) = vec(A')'(C' \otimes I)vec(B)$

$$= vec(A')'(I \otimes B)vec\ (C)$$

$$= vec\ (B')'(A' \otimes I)vec(C)$$

$$= vec\ (B')'(I \otimes C)vec\ (A)$$

$$= vec(C')'(B' \otimes I)vec(A)$$

$$= vec(C')'(I \otimes A)vec\ (B)$$

The vech operator stacks the elements on and below the main diagonal of a square matrix.

B: Stochastic convergence and asymptotic distributions

B1) Convergence in probability and in distribution

Let $x_{1,}, x_2, \ldots$ or $\{x_T\}, T = 1,2, \ldots$ be a sequence of scalar random variables which are all defined on a common probability space $(\Omega, \mathcal{F}, \mathrm{Pr})$. The sequence $\{x_T\}$ converges in probability to the random variable x (which is also defined on $(\Omega, \mathcal{F}, \mathrm{Pr})$) if for every $\varepsilon > 0$

$$\lim_{T \to \infty} \mathrm{Pr}(|x_T - x| > \varepsilon) = 0$$

Or equivalently

$$\lim_{T \to \infty} \mathrm{Pr}(|x_T - x| < \varepsilon) = 1$$

This type of stochastic convergence is abbreviated as

$plim x_T = x$ or $x_T \xrightarrow{p} x$

The sequence $\{x_T\}$ converges almost sure (a.s) or with probability one to the random variable x if for every $\varepsilon > 0$

$$\mathrm{Pr}(\lim_{T \to \infty} |x_T - x| < \varepsilon) = 1$$

This type of convergence is often written as $x_T \xrightarrow{a.s} x$ and is sometimes called strong convergence.

Denoting the distribution functions of x_T and x by F_T and F respectively, the sequence $\{x_T\}$ is said to converge in distribution or weakly or in law to x, if for all real numbers c for which F is continous

$$\lim_{T \to \infty} F_T(c) = F(c)$$

This type of convergence is abbreviated as $x_T \xrightarrow{d} x$.

These concepts of stochastic convergence can be extended to sequences of random vectors (multivariate random variables)

Suppose $\{x_T = (x_{1T}, \dots, x_{KT})'\}, T = 1,2, \dots$ is a sequence of K-dimensional random vectors and $x = (x_1, \dots, x_K)'$ is a K-dimensional random vector. Then

$plim x_T = x$ or $x_T \xrightarrow{p} x$ if $plim x_{kT} = x_k$ for $k = 1, \dots, K$

$x_T \xrightarrow{a.s} x$ if $x_{kT} \xrightarrow{d} x_k$ for $k = 1, \dots, K$

$x_T \xrightarrow{d} x$ if $lim F_T(c) = F(c)$ for all continuity points of F.

In this case F_T and F are the joint distribution functions of x_T and x respectively.

B2) Law of Large Numbers (LLN) and Central Limit Theorems (CLT)

Suppose $\{x_t\}, t = 1,2, \dots$ is a sequence of zero mean random variables and let Ω_t be an information set available at time t which includes at least $\{x_1, \dots, x_t\}$ and possibly other random variables. The sequence $\{x_t\}$ is said to be a martingale difference sequence with respect to the sequence Ω_t if $E(x_t/\Omega_{t-1}) = 0$ for all $t = 2,3, \dots$

It is simply referred to as a martingale difference sequence if $E(x_t) = 0$ for $t = 1,2, \dots$ and $E(x_t/x_{t-1}, \dots, x_1) = 0 \; for \; t = 2,3, \dots$

More generally, a sequence $\{x_t\}$ of K-dimensional vector random variables satisfying $E(x_t) = 0$ for all t and $E(x_t/x_{t-1}, \ldots, x_1) = 0 \ for \ t = 2,3, \ldots$ is a vector martingale difference sequence.

Proposition B1: law of large numbers (LLN)

1) LLN for martingale differences

   Let $\{x_t\}$ be a strict stationary martingale difference sequence with $E|x_t| < \infty$ t=1,2,…, then $\bar{x}_T \overset{p}{\to} 0$

2) LLN for martingale difference arrays

   Let $\{x_t\}$ be a martingale difference array with $E\left|x_{T,t}\right|^{1+\varepsilon} \leq c < \infty \ for \ all \ t \ and \ T$ for some $\varepsilon > 0$ and a finite constant c. Then

$$\bar{x}_T \coloneqq T^{-1} \sum_{t=1}^{T} x_{T,t} \overset{p}{\to} 0$$

Proposition B2: Central Limit Theorem (CLT)

1) CLT for martingale difference arrays

   Let $\left\{x_{T,t} = (x_{1T,t}, \ldots, x_{kT,t)})'\right\}$ be a K-dimensional martingale difference array with covariance matrices

   $E\left(x_{T,t}x'_{T,t}\right) = x_{Tt}$such that $T^{-1}\sum_{t=1}^{T}\sum_{Tt} \to \sum, \ where \ \sum is \ positive \ definite$

   Moreover suppose that $T^{-1}\sum_{t=1}^{T} x_{T,t}x'_{T,t} \overset{p}{\to} \sum and \ E(x_{iT,t}x_{jT,t}x_{kT,t}x_{lT,t}) < \infty$ for all $t \ and \ T$ and all $1 \leq i,j,k,l \leq K$. Then

$$\sqrt{T}\bar{x}_T \overset{d}{\to} N(0, \sum)$$

2) CLT for stationary processes

Let $x_t = \mu + \sum_{j=0}^{\infty} \Phi_j u_{t-j}$ be a K-dimensional stationary stochastic process with $E(x_t) = \mu < \infty, \sum_{j=0}^{\infty} \left\| \Phi_j \right\| < \infty$ and $u_t \sim (0, \sum_u)$ iid white noise. Then

$$\sqrt{T}(\bar{x}_T - \mu) \xrightarrow{d} N(0, \sum_{j=-\infty}^{\infty} \Gamma_x(j))$$

where

$$\Gamma_x(j) \coloneqq E\left[(x_t - \mu)(x_{t-j} - \mu)'\right]$$

C: STOCHASTIC INEQUALITIES

Chebyshev

Let X be a non-negative random variable with finite mean μ_X and finite variance σ_X^2. Then, for any $\varepsilon \in \boldsymbol{R}, \varepsilon > 0$

$$P(|X - \mu_X|) > \sqrt{\frac{\sigma_X^2}{\varepsilon}} \leq \varepsilon$$

Cauchy-Schwartz

For p=q=2, we have

$$E(|XY|) \leq \sqrt{E|X|^2}\sqrt{E|Y|^2}$$

Minkowski

If for some $p \geq 1, E(|X|^p) < \infty \; and \; E(|Y|^p) < \infty$, then

$$E(|X + Y|) \leq [E(|X|^p)]^{\frac{1}{p}}[E(|Y|^p)]^{\frac{1}{p}}$$

D: STOCHASTIC CONVERGENCE AND ASYMPTOTIC DISTRIBUTIONS

D1: Concepts of stochastic convergence

Let $x_{11}, x_2, \ldots$ or $\{x_T\}, T = 1,2, \ldots$ be a sequence of scalar random variables which are all defined on a common probability space $(\Omega, \mathcal{F}, Pr)$. The sequence $\{x_T\}$ converges in probability to the random variable x (which is also defined on $(\Omega, \mathcal{F}, Pr)$ if for every $\in > 0$

$$\lim_{T\to\infty} Pr(|x_T - x| > \in) = 0$$

Or equivalently

$$\lim_{T\to\infty} Pr(|x_T - x| < \in) = 1$$

This type of stochastic convergence is abbreviated as

$$plim x_T = x \ or \ x_T \overset{p}{\to} x$$

The limit x may be a fixed, non-stochastic real number which is then regarded as a degenerate random variable that takes on one particular value with probability one.

The sequence $\{x_T\}$ converges almost surely (a.s) or with probability one to the random variable x if for every $\in > 0$

$$Pr\left(\lim_{T\to\infty} |x_T - x| < \in\right) = 1$$

This type of stochastic convergence is often written as $x_T \overset{a.s}{\to} x$ and is sometimes known as strong convergence.

The sequence $\{x_T\}$ converges in quadratic mean or mean square error to x, written briefly as $x_T \overset{q.m}{\rightarrow} x$, if

$$\lim_{T\to\infty} E(x_T - x)^2 = 0$$

This type of convergence requires that the mean and variance of the $x_T's$ and$x$ exist.

Denoting the distribution functions of $x_T \ and \ x$by $F_T \ and \ F, respectively$, the sequence $\{x_T\}$ is said to converge in distribution or weakly or in law to x, if for all real numbers c for which F is continuous

$$\lim_{T\to\infty} F_T(c) = F(c)$$

This type of convergence is abbreviated as $x_T \overset{d}{\rightarrow} x$.

All these concepts of stochastic convergence can be extended to sequences of random vectors (multivariate random variables). Suppose $\{x_T = (x_{1T}, \dots, x_{KT})'\}, T = 1,2, \dots$, is a sequence of K-dimensional random vectors and $x = (x_1, \dots, x_K)'$ is a K-dimensional random vector. Then the following dimensions are used

$$plim x_T = x \ or \ x_T \overset{p}{\rightarrow} x \ if \ plim x_{kT} = x_k \ for \ k = 1, \dots, K$$

$$x_T \overset{a.s}{\rightarrow} x \text{ if } x_{kT} \overset{a.s}{\rightarrow} x_k \ for \ k = 1, \dots, K$$

$$x_T \overset{q.m}{\rightarrow} x \ if \ lim E[(x_T - x)'(x_T - x)] = 0$$

$x_T \overset{d}{\rightarrow} x \ if \ lim F_T(c) = F(c) for \ all \ continuity \ points \ of \ F.$

Here F_T and F are the joint distribution functions of x_T $and\ x\ respectively.$

Proposition D1: convergence properties of sequences of random variables

Suppose $\{x_T\}$ is a sequence of K-dimensional random variables. Then the following relations hold:

1. $x_T \overset{a.s}{\to} x \Rightarrow x_T \overset{p}{\to} x \Rightarrow x_T \overset{d}{\to} x$
2. $x_T \overset{q.m}{\to} x \Rightarrow x_T \overset{p}{\to} x \Rightarrow x_T \overset{d}{\to} x$
3. If x is a fixed non-stochastic vector, then

$$x_T \overset{q.m}{\to} x \Leftrightarrow [limE(x_T) = x\ and\ limE\{(x_T - Ex_T)'(x_T - Ex_T)\} = 0]$$

4. If x is a fixed, non-stochastic random vector, then $x_T \overset{p}{\to} x \Rightarrow x_T \overset{d}{\to} x$
5. Slutsky's theorem

   If $g: R^K \to R^m$ is a continuous function, then $x_T \overset{p}{\to} x \Rightarrow g(x_T) \overset{p}{\to} g(x)[plimg(x_T) = g(plimx_T)]$

   $x_T \overset{d}{\to} x \Rightarrow g(x_T) \overset{d}{\to} g(x)andx_T \overset{a.s}{\to} x \Rightarrow g(x_T) \overset{a.s}{\to} g(x)$

Proposition D2:properties of convergence in probability and in distribution

Suppose $\{x_T\}$ and $\{y_T\}$ are sequences of $K \times 1$ random vectors, $\{A_T\}$ is a sequence of $K \times K$ random matrices, x is a fixed $K \times K$ matrix

1. If $plimx_T, plimy_T\ and\ plimA_T$ exist, then

   a) $plim(x_T \pm y_T) = plimx_T \pm plimy_T$

   b) $plim(c'x_T) = c'(plimx_T)$

c) $plimx_T'^{y_T} = (plimx_T)'(plimy_T)$

d) $plimA_Tx_T = plim(A_T)plim(x_T)$

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

3. If $x_T \overset{d}{\rightarrow} x$ and $plimy_T = c$, then

a) $x_T \pm y_T \overset{d}{\rightarrow} x \pm c$

b) $y_T' x_T \overset{d}{\rightarrow} c'x$

4. If $x_T \overset{d}{\rightarrow} x$ and $plimA_T = A,\ then\ A_Tx_T \overset{d}{\rightarrow} Ax$

5. If $x_T \overset{d}{\rightarrow} x$ and $plimA_T = 0, then\ plimA_Tx_T = 0$

Proposition D3: weak laws of large numbers

1. Khinchen's theorem (Rao, 1973 p112)

Let$\{x_t\}$ be asequence of iid random variables with $E(x_t) = \mu < \infty$. Then

$$x_T := \frac{1}{T}\sum_{t=1}^{T} x_T \overset{p}{\rightarrow} \mu$$

2. Let $\{x_t\}$ be a sequence of independent random variables with $E(x_t) = \mu < \infty$ and $E|x_t|^{1+\epsilon} \leq c < \infty\ (t = 1,2,\dots)\, for\ some\ \ \epsilon > 0$ and a finite constant c. Then $\bar{x}_T \overset{p}{\rightarrow} \mu$

3. Chebyshev's theorem (Rao, 1973 p112)

Let $\{x_t\}$ be a sequence of uncorrelated random variables with $E(x_t) = \mu < \infty$ and $limE(\bar{x}_T - \mu)^2 = 0\ then\ \bar{x}_T \overset{p}{\rightarrow} \mu$

4. Corollary to Chebyshev's theorem

Let $\{x_t\}$ be a sequence of independent random variables with $E(x_t) = \mu < \infty var(x_t) \leq$ $c < \infty\ (t = 1,2,\ldots) for\ some\ finite\ constant\ c, then\ \bar{x}_T \xrightarrow{p} \mu$.

E: WISHARTS DISTRIBUTION

E1: The Wishart Distribution

The Wishart distribution is the multivariate expansion of the gamma distribution, although most statisticians use the Wishart distribution in the special case of integer degrees of distribution, in which case it simplifies to a multivariate generalization of the χ^2 distribution. As the χ^2 distribution describes the sums of squares of n draws from a multivariate normal distribution, the Wishart distribution represent the sum of squares (and cross products) of n draws from a multivariate normal distribution.

Let $\mathrm{S} \sim \mathrm{Wish}_{\mathrm{p}}(\sum, \mathrm{v})$, where $\sum$ denotes a positive definite scale matrix (which can be thought of as a variance/covariance matrix from a multivariate normal distribution), v is the parameter that denotes the degrees of freedom and p indicates the dimensions of S, i.e., $\mathrm{S} \epsilon \mathbb{R}^{\mathrm{p} \times \mathrm{p}}$. Then S is positive definite with pdf

$$f(S) = \frac{|S|^{\frac{v-p-1}{2}}}{2^{\frac{vp}{2}} |\sum \quad|^{\frac{v}{2}} \Gamma_p(\frac{v}{2})} exp\left[-\frac{1}{2} tr(\sum^{-1} S)\right]$$

Where $\Gamma_p(x) = \pi^{\frac{1}{2}(\frac{p}{2})} \prod_{j=1}^{p} \Gamma[x + (1-j)/2]$ is the multivariate generalization of the gamma function Γ.

E2: The Inverse-Wishart distribution

The Inverse-Wishart distribution is the multivariate extension of the Inverse-gamma distribution (or similar to the Wishart distribution; the inverse χ^2 distribution in the case of the integers degrees of freedom).

Let $\mathrm{T} \sim \mathrm{InvWish}_{\mathrm{p}}(\psi, \mathrm{m})$ where ψ denotes a positive definite scale matrix (which can be thought of as sums of square matrix from a multivariate normal distribution), m is the parameter that denotes the degrees of freedom and p indicates the dimensions of T, i.e., $\mathrm{T} \epsilon \mathbb{R}^{\mathrm{p} \times \mathrm{p}}$. then T is positive definite with pdf

$$f(T) = \frac{\psi^{\frac{\mathrm{m}}{2}}}{|\mathrm{T}|^{\frac{\mathrm{m}+\mathrm{p}+1}{2}} 2^{\frac{\mathrm{mp}}{2}} \Gamma_{\mathrm{p}}\left(\frac{\mathrm{m}}{2}\right)} \exp\left[-\frac{1}{2} \mathrm{tr}(\psi \mathrm{T}^{-1})\right]$$

Where $\Gamma_p(x)$ is as defined in appendix E1.

The choice of the Wishart priors in this work is motivated by the fact that it does not bound σ_i away from zero and increases sampling efficiency considerably.

F: TABLE OF EIGEN VALUES

Table 20: Eigen values of the GVAR model and corresponding moduli

Eigenvalues of the GVAR Model in Descending Order	Corresponding Moduli
1.02981416203722 +0.00000000000000i	1.029814162
1.00000000000000 -0.00000000000001i	1
1.00000000000000 +0.00000000000001i	1
0.73825198141311 -0.00000000000000i	0.738251981
0.41446466318635 -0.07890768783217i	0.652142445
0.41446466318635 +0.07890768783217i	0.652142445
0.15171280897435 +0.14459097421581i	0.493157551
0.15171280897435 -0.14459097421581i	0.421909209
0.03466968718702 -0.40465695338521i	0.421909209
0.03466968718702 +0.40465695338521i	0.406139431
0.00969975457916 +0.00000000000000i	0.406139431
0	0.358265217
0	0.209578926
0	0.209578926
0	0.171419581
0	0.104784157
0	0.104784157
0	0.020272797
0	0.009699755
0	0
0	0
0	0
0	0
0	0
-0.02027279690868 -0.00000000000000i	0
-0.10463456254011 -0.00559714000765i	0
-0.10463456254011 +0.00559714000765i	0
-0.17141958056471 +0.00000000000000i	0
-0.22349000598944 -0.61265160196674i	0
-0.22349000598944 +0.61265160196674i	0
-0.35826521686838 +0.00000000000000i	0
-0.49315755115142 +0.00000000000000i	0