UCLouvain

FACULTE DES SCIENCES

Faculté des sciences

Kernel Independent Component Analysis (ICA) for identification of structural innovations in multivariate GARCH models

Author: **Mathias Dah Fienon**
Supervisor: **Christian M. Hafner**
Reader: **Rainer von Sachs**
Academic year: 2023–2024
Master [120] en science des données, orientation statistique

Acknowledgments

First, I would like to thank my family for their continuous support during my studies. Hospice Dossou-Yovo and Teico Kadadji, thanks a lot, without you, this adventure would not be that easy. And of course, thank you, to all my other beloved ones who supported me.

Finally, I would like to thank Prof Christian Hafner who supervised this work and was always very kind and patient! The process was always very natural and our meetings really helped a lot to find which interesting path this work could take. Also, thank you Prof Rainer von Sachs, for reading this thesis and for all the courses you taught me during my Master's degree.

*

This dissertation has been prepared in partial fulfillment of the requirements for the master degree in data science (statistics orientation) delivered by the Université Catholique de Louvain (Louvain-la-Neuve, Belgium).

Contents

Contents

List of Abbreviations

GDP	...	Gross Domestic Product
PCA	...	Principal Components Analysis
ICA	...	Independente Components Analysis
KernelICA	...	Kernel Independent Components Analysis
GOOG	...	Google
AAPL	...	Apple
AMZN	...	Amazon
MSFT	...	Microsoft
DCC	...	Dynamic conditional correlation
CCC	...	Constant conditional correlation
VAR	...	Vector autoregressive
MGARCH	...	Multivariate generalized autoregressive conditional heteroskedasticity
AIC	...	Akaike information criteria
BIC	...	Bayesian information criteria
VIRF	...	Volatility impulse response function
GIRF	...	Generalized impulse response function

Notations

$\boldsymbol{a}$	Vector
$\boldsymbol{a}^\top$	Transposed vector
$\mathbf{A}$	Matrix (bold upper-case)
$\mathbf{A}^\top$	Transposed matrix
$\mathbf{A^{-1}}$	Inverted matrix
$\mathbb{E}(\mathbf{.})$	Expectations of $\mathbf{.}$
$\mathbb{V}\mathrm{ar}(\mathbf{.})$	Variance of $\mathbf{.}$

List of Figures

List of Tables

Introduction

With the globalization and the wake of a growing dependence across markets, enlightened by the financial and economic crisis, the volatility linkages between countries and markets has gained interest among financial regulators and policymarkers. In order to study this contagion and transmission of an unexpected change in one variable, usually called shock, on other variables, macro-economist and empirical financial experts have developed the structural models (Sims, 1980). In these same perspectives, multivariate GARCH models have been developed to model the conditional covariance matrix of multiple times series. Knowing this volatility matrix gives valuable information on the dynamics of the series and thus the financial assets.

Indeed, in such framework, it is important to distinguish correlation from causality as said Stock and Watson (2018), if the goal of the analysis is measuring the effects of exogenous interventions on the system.

Therefore, several identification criteria have been proposed [see among others Blanchard and Quah (1988), Lütkepohl, Staszewska-Bystrova, and Winker (2018), Rigobon (2003), etc...], among which the exploitation of the non-Gaussianity as sufficient statistical property for local identification instead of heteroskedasticity. The idea behind is to recover the structural dynamic as linear combinations of the residuals of the structural models as explained by Moneta and Pallante (2022). This is possible under the assumptions that the residuals are not just uncorrelated but also mutually statistical independent and thus by the means the data-driven techniques called the independent component analysis (ICA), one can retrieved these shocks. ICA techniques are flexible in the sense that they refrain from imposition of additional asumptions on the data or the economic relation between the variables (Maxand, 2020).

In this thesis, knowing the flexibility offers by ICA coupled with the performance of non linear ICA over simple ICA, we want to investigate potential benefits of using nonlinear ICA such as kernel independent component analysis (kernelICA) (Bach & Jordan, 2002) for identification. Hence, our thesis is structured as follows.

We start, in chapter 1, by a short literature review on structural models and the identification techniques used in empirical studies. This chapter ends will a note on multivariate GARCH models and some data-driven identification techniques.

Chapter 2 treats of fundamentals of independent component analysis, some of its algorithms and how it is used in data-driven structural shocks identification. It ends with a review of nonlinear independent component analysis proposed by Bach and Jordan (2002).

The third chapter is dedicated to the simulations to retrieve the performance of

kernelICA in recovering the structural shocks of given residuals obtained from a multivariate GARCH model. It also studies the impact of some key parameters on the algorithms performance.

The last chapter, we applied kernelICA to a set of financial assets indexes of the five high-tech leaders: Google, Apple, Meta (Facebook), Amazon and Microsoft, in order to study the structural dynamic under the behaviour of these 5 indexes. Afterward, we analyze the volatility impulse responses of a given shocks on the indexes with regards to the structural matrices obtained from various ICA algorithms and the identity matrix.

Chapter 1

Structural innovations and identification techniques

This first chapter starts by presenting vector autoregressive (VAR) model, when is it considered as structural; then presents the multivariate generalized autoregressive conditional heteroskedasticity (MGARCH) models, and deals with the literature review on identification techniques of the structural innovations of such models.

1.1 Vector autoregressive (VAR) model and identification

1.1.1 Vector autoregressive model

Vector autoregressive model commonly abbreviated as VAR is a widely used model for multivariate times series analysis (Kilian & Lütkepohl, 2017). It has become one of the major ways of extracting information about the macro-economy. It is used for quantifying impulse responses to macroeconomics shocks, for measuring the degree of uncertainty about the impulse responses or other quantities formed from them and for deciding on the contribution of different shocks to fluctuations and forecast errors through variance decomposition (Fry & Pagan, 2011). While capturing the contemporaneous and lagged relationships between macro-economic variables, it helps in understanding the economic mechanisms and interactions between variables often based on economic theory or empirical analysis. Thereforth, it finds interest in policy analysis and evaluation, in business cycle analysis and in economic shocks evaluation.

It consists of a system of autoregressive equations where each model variable is regressed on lags of its own passed values as well as lags of the other variables up to a prespecified maximum lag order, say p. Hence a VAR model with p lags is referred to as a VAR(p) model.

Let us consider as an illustrative example a 3-dimensional times series model denoted by $\boldsymbol{y}_t$ consisting of Belgium real Gross National Product (GNP) growth (r_t) , inflation rate (i_t) and employment rate (e_t), for $t = 1, \ldots, T$. Taking a monthly based series, a VAR(2)

model, as an example, is written as :

$$\boldsymbol{y}_t = \boldsymbol{c} + \mathbf{A}_1 \boldsymbol{y}_{t-1} + \mathbf{A}_2 \boldsymbol{y}_{t-2} + \boldsymbol{\varepsilon}_t \tag{1.1}$$

where $\boldsymbol{y}_t = \begin{pmatrix} r_t \\ i_t \\ e_t \end{pmatrix}$, $\boldsymbol{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$, $\mathbf{A}_i = \begin{bmatrix} a_{11,i} & a_{12,i} & a_{13,i} \\ a_{21,i} & a_{22,i} & a_{23,i} \\ a_{31,i} & a_{32,i} & a_{33,i} \end{bmatrix}$, $i = 1, 2$, and $\boldsymbol{\varepsilon}_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}$.

$$(1.1) \implies \begin{cases} r_t = c_1 + a_{11,1} r_{t-1} + a_{12,1} i_{t-1} + a_{13,1} e_{t-1} + a_{11,2} r_{t-2} + a_{12,2} i_{t-2} + a_{13,2} e_{t-2} + \varepsilon_{1t} \\ i_t = c_2 + a_{21,1} r_{t-1} + a_{22,1} i_{t-1} + a_{23,1} e_{t-1} + a_{21,2} r_{t-2} + a_{22,2} i_{t-2} + a_{23,2} e_{t-2} + \varepsilon_{2t} \\ e_t = c_3 + a_{31,1} r_{t-1} + a_{32,1} i_{t-1} + a_{33,1} e_{t-1} + a_{31,2} r_{t-2} + a_{32,2} i_{t-2} + a_{33,2} e_{t-2} + \varepsilon_{3t} \end{cases}$$

For convenience, the intercept $\boldsymbol{c}$ is considered not random (or not impactful since it can be recovered from the noise $\boldsymbol{\varepsilon}_t$), so one can suppress it. The zero mean innovations ε_{it}, $i = 1, 2, 3$, are serially uncorrelated if the maximum lag order has been chosen appropriately. Also given an information set consisting of the past values of all k ($k = 3$ in our case) model series, the vector of innovations $\boldsymbol{\varepsilon_t}$ is the unpredictable component of $\boldsymbol{y}_t$. Equation (1.1) is known as the reduced form.

The reduced form of order p can be viewed as representing data generated from a structural VAR(p) model :

$$\mathbf{B}_0 \boldsymbol{y}_t = \mathbf{B}_1 \boldsymbol{y}_{t-1} + \mathbf{B}_2 \boldsymbol{y}_{t-2} + \ldots + \mathbf{B}_p \boldsymbol{y}_{t-p} + \boldsymbol{w}_t \tag{1.2}$$

where

- $\boldsymbol{y}_t$ is a $K \times 1$ vector of the observed times series, $t = 1, \ldots, T$
- $\mathbf{B}_i$ with $i = 1, \ldots, p$ is a $K \times K$ matrix of the autoregressive coefficients
- $\mathbf{B}_0$ the $K \times K$ matrix that reflects the instantaneous relations among the model variables
- $\boldsymbol{w}_t$, the $K \times 1$ structural shocks vector of mean zero that is serially uncorrelated. Thus, its covariance matrix $\boldsymbol{\Sigma_w}$ is not only diagonal but also full rank such that the number of shocks coincides with the number of variables.

The $K \times K$ matrix $\mathbf{B}_0^{-1}$ captures the influence of each of the structural shocks on each of the model variables. The model (1.2) is then structural in the sense that the shocks are postulated to be mutually uncorrelated with each element of $\boldsymbol{w}_t$ having a distinct economic interpretation. This fact allows one to interpret movements in the data caused by any one element of $\boldsymbol{w}_t$ as being caused by that shocks.

Kilian and Lütkepohl (2017) say that the structural shocks in general are not directly observable, but under suitable conditions, they may be recovered from the reduced form representation of model (1.2) which is obtained by premultiplying both sides of the equation by $\mathbf{B}_0^{-1}$, resulting in the equation :

$$\boldsymbol{y}_t = A_1\boldsymbol{y}_{t-1} + \ldots + A_p\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t \tag{1.3}$$

where $\mathbf{A}_i = \mathbf{B}_0^{-1}\mathbf{B}_i$ and $\boldsymbol{\varepsilon}_t = \mathbf{B}_0^{-1}\boldsymbol{w}_t$.

The equation (1.3) can be estimated by unrestricted least-squares (LS) or by maximum likelihood (ML) estimation methods.

Once the reduced form (1.3) is given, all that is required to recover the structural model (1.2) is the structural impact multiplier matrix $\mathbf{B}_0^{-1}$ (or its inverse $\mathbf{B}_0$).

To estimate $\mathbf{B}_0$, additional restrictions on the data generating process is needed. These restrictions can be based on economic theory, institutional knowledge or other external constraints. And if, given these restrictions, the matrix $\mathbf{B}_0$ can be solved, we then say that the structural VAR model (1.2) parameters $(\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_p, \boldsymbol{\Sigma}_{\boldsymbol{w}})$ are identified; equivalently the structural shocks $\boldsymbol{w}_t = \mathbf{B}_0\boldsymbol{\varepsilon}_t$ are identified and suitable economic interpretation can be given to shocks. The problem of finding suitable economically credible restrictions on $\mathbf{B}_0^{-1}$ or $\mathbf{B}_0$ is the identification problem in structural VAR model analysis. And this has become a central endeavour in empirical macroeconomics studies.

These credible restrictions define the validity of the structural VAR model analysis, but finding a credible set of restrictions is very challenging. Also, depending on the identifying assumptions, the structural VAR may not be unique.

A good point to keep in mind is that with no clear economic interpretation of the elements of $\boldsymbol{w}_t$, the model (1.2) is not structural. Thus, no mutual correlation between elements of $\boldsymbol{w}_t$ is not sufficient. Indeed, VAR applications that overlooked this requirement of economical interpretation attracted strong criticism and stimulated the development of more explicit structural VAR models (Cooley & LeRoy, 1985).

The existence of a structural VAR model allows us to think of the variation in the data as being driven by the cumulative effects of economically interpretable structural shocks and the mapping from the reduced form allows to quantify these structural shocks. A key difference with a direct measures of exogenous shocks is that the structural shocks are usually unobserved and need not to be associated with any one VAR model variable in particular. This greatly increases the range of applications of the VAR approach, thus its interest.

Although finding credible restrictions is challenging, several approaches have been made available in the literature to tackle this task. These methods are the subject of the next section.

1.1.2 Structural shocks identification methods

Recall from the above section that the structural innovations or shocks are the $\boldsymbol{w}_t$ from the structural VAR model (1.2). These shocks are of mean zero and serially uncorrelated; also considered unconditionally homoskedastic, unless noted otherwise.

Expressing the reduced form (model 1.3) of this structural form and after consistently estimated its parameters, the question is to recover the structural representation which is actually governed by $\mathbf{B}_0$. By construction $\boldsymbol{\varepsilon}_t = \mathbf{B}_0^{-1}\boldsymbol{w}_t$. Hence, the symmetric variance-

covariance matrix $\boldsymbol{\Sigma}_{\varepsilon}$ of $\boldsymbol{\varepsilon_t}$ is given by:

$$\begin{aligned}\boldsymbol{\Sigma}_{\varepsilon} &= \mathbb{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{\top}) \\ &= \mathbb{E}[\mathbf{B}_0^{-1}\boldsymbol{w}_t\boldsymbol{w}_t^{\top}(\mathbf{B}_0^{-1})^{\top}] \\ &= \mathbf{B}_0^{-1}\mathbb{E}[\boldsymbol{w}_t\boldsymbol{w}_t^{\top}](\mathbf{B}_0^{-1})^{\top} \\ &= \mathbf{B}_0^{-1}\boldsymbol{\Sigma}_{\boldsymbol{w}}(\mathbf{B}_0^{-1})^{\top}. \end{aligned} \tag{1.4}$$

By assuming unit variance for $\boldsymbol{w}_t$ or by normalizing the variance to unit, one have $\boldsymbol{\Sigma}_w = \boldsymbol{I}_K$; thus $\boldsymbol{\Sigma}_{\varepsilon}$ is totally identifiable by $\mathbf{B}_0^{-1}$ (inverse of B_0):

$$\boldsymbol{\Sigma}_{\varepsilon} = \mathbf{B}_0^{-1}(\mathbf{B}_0^{-1})^{\top} \tag{1.5}$$

The system of nonlinear $K(K+1)/2$ independent equations of (1.5) can be solved using numerical methods provided that the number of parameters in $\mathbf{B}_0^{-1}$ does not exceed the number of independent equations. To achieve this, additional restrictions on selected elements of $\mathbf{B}_0^{-1}$ may be necessary. The most common restrictions is to force some particular elements of $\mathbf{B}_0^{-1}$ to zero [see (Sims, 1980), (Christiano, Eichenbaum, & Evans, 1999), (Kilian, 2009) and (Kilian & Lütkepohl, 2017)]. Forcing some particular elements of $\mathbf{B}_0^{-1}$ to zero means that one is imposing a certain structure within the variables based on prior knowledge of the relationships between the variables. Using prior knowledge to make restrictions may fault the analysis in purpose in such way that one is no longer measuring the actual structure within the variables. Nevertheless, proper restrictions based on some economical knowledge have been widely used to study the structural impacts on the variables of the model. In the following, we will discuss the principal restrictions imposed on the variables to perform the analysis and their respective limitations.

1.1.2.1 Short-Run and Long-Run identification techniques

We have stressed in the preceding that in purpose of rendering the identification task easier, a traditional approach is to consider some economical restrictions. Some of these restrictions come from the short-run knowledge of the dynamics of our economic system. Thus by judiciously imposing the restrictions, one can achieve identification easier. In other hands, these restrictions should have an economical interpretation. So, two main source of finding these restrictions are : economic theory and derived alternatives.

Hence, by stating some short run assumptions based on prior knowledge one has about the structure of the economics in study and taking from economy theory, one can impose some coefficients of $\mathbf{B}_0^{-1}$ to 0. This point is best illustrated by example:

Consider our example of $\boldsymbol{y}_t$ series consisting of Belgium real Gross National Product (GNP) growth (r_t) , inflation rate (i_t) and employment rate (e_t), for $t = 1, \ldots, T$. Suppose that one recover the error $\boldsymbol{\varepsilon}_t$ by a given method (maximum likelihood for example). To recover the shocks $\boldsymbol{w}_t$ one postulated the following:

$$\begin{aligned}\boldsymbol{\varepsilon}_t &= \mathbf{A}\boldsymbol{w}_t \\ \begin{bmatrix}\varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3\end{bmatrix} &= \begin{bmatrix}a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3\end{bmatrix}\begin{bmatrix}w_1 \\ w_2 \\ w_3\end{bmatrix}\end{aligned} \tag{1.6}$$

$$(1.6) \implies \begin{cases} \varepsilon_1 = a_1 w_1 + a_2 w_2 + a_3 w_3 \\ \varepsilon_2 = b_1 w_1 + b_2 w_2 + b_3 w_3 \\ \varepsilon_3 = c_1 w_1 + c_2 w_2 + c_3 w_3 \end{cases}$$

with ε_1 the estimate error of r_t the GNP, ε_2 estimate error of i_t inflation rate and ε_3 estimate error of e_t employment rate.

Suppose now that based on knowledge of the dynamics of Belgium economy, an econometrician would like to study the structural shocks governing the dynamics of this economy. A priori, he knows that the inflation rate impacts the employment rate which in return impacts the GNP and not other way around. Hence, he must postulate the following :

$$\begin{cases} \varepsilon_1 = a_1 w_1 + a_2 w_2 + a_3 w_3 \\ \varepsilon_2 = b_1 w_1 \\ \varepsilon_3 = c_3 w_3 + c_2 w_2 \end{cases} \implies \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & 0 & 0 \\ 0 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}. \tag{1.7}$$

Thus, he sets some coefficients of $\mathbf{A}$ to 0. And furthermore, some ordering of the variable before estimating the errors, may lead him to render his $\mathbf{A}$ matrix triangular.

An important notion to take into account here is that such restrictions impose certain contemporaneous relationships and structural causality between variables in such way that the immediate effects of some shocks on the dynamic of the variables may not be uncovered. This is a one limit of the short-run restrictions approach. Another limitation of this approach is the fact that by imposing a structure in the model, one is not letting the data speaks of itself and reveals all available information one can extract from it. And by experience, one knows how economic dynamics can vary a lot through time. One more limitation and not the least is that short-run restrictions in VAR models does not guarantee a unique identification of the structural innovations. Thus, different sets of restrictions can lead to different interpretations and conclusions, rendering by the same occasion a subjective interpretation of the results of such approach.

On the other hand, to study the dynamic structure of the model in long-term towards equilibrium, one can proceed by using the long-run restrictions approach (Blanchard & Quah, 1988) which consist of the same scheme as in the short-run (setting some coefficients to 0 or the sums of some coefficients) but based the restrictions on the prior knowledge of the long-run dynamics of the economics. Thus, they provide insights into the direction and magnitude of the effects of shocks on the variables in long-term scenario. Also in this schema, one is imposing some structure in the model and not recovering the information that the data can convey. And as in the short-run scenario, limitation of such approach are obvious, and misspecification of the model can lead to fallacious interpretation of the results (Faust & Leeper, 1997). Another important limitation here is the fact that putting aside all other limitation, one need sufficiently long time series data to capture the equilibrium behavior. Inadequate data length will limit the reliability of the estimated long-run coefficients. Although the availability of financial times series data may not be such a huge problem, the amount of data needed to study a particular dynamics may not be available.

The contrast between using short-run or long-run has been studied and Lütkepohl et al. (2018) show that the impact multipliers are estimated more accurately when identification is based on short-run restrictions but if intermediate or longer propagation horizons are of interest, long-run multipliers are estimated much more precisely when using the long-run restrictions; this specifically holds for persistent processes with a root close to one [see Lütkepohl et al. (2018) for more detail].

1.1.2.2 Identification by sign restrictions

From the above subsection, we have emphasized the limitations of the short and long-run restrictions used in identification of the structural shocks. Based on that one may suggest to analyse the structural innovations by just imposing a sign restrictions which means that one is imposing a direction of the responses of variables to a particular shock [see Canova and De Nicolo (2002), Faust (1998), Hau and Rey (2004)]. That way, one is no longer interested in the magnitude of the effects of a particular shock on the variables but rather the directions impulsed by the shock on the dynamic of the variables. Such approach is known as the sign restrictions approach.

For example, taking back our econometrician, he may know that the inflation impact negatively the employment rate and the latest also impact negatively the GNP. Hence, the following sign restrictions will be imposed on the coefficients of the matrix $\mathbf{A}$:

$$\begin{bmatrix} + & - & - \\ + & + & + \\ + & - & + \end{bmatrix}$$

This method offers on the one hand a flexible and intuitive approach in the identification schema, and on the other hand the identification of multiple structural shocks simultaneously, making them suitable for exploring the effects of various innovations on the system. Fry and Pagan (2011) has shown that although this approach solves the structural identification problem by providing sufficient information to identify the structural parameters, it leaves the model identification problem unresolved as called by Preston (1978).

1.1.2.3 Identification based on data heteroskedasticity

Rigobon (2003) has approached the identification problem using the changes in the volatility of the variables for identifying structural VAR models. The assumption is that the presence of heteroskedasticity in the residuals can provide insights on identifying and interpreting the structural shocks and their impact on the structural changes. Heteroskedasticity can complement identifying restrictions based on economic theory or subject matter knowledge. The underlying idea is that if the variance of the structural shocks changes during the sample period and there is heterogeneity in the variance changes of different shocks,this feature can be used to distinguish (identify) the shocks (Lütkepohl, Meitz, Netšunajev, & Saikkonen, 2021).

1.1.2.4 Identification using external instruments

Another approach used for this structural shocks identification problem is using exogenous variables or external instruments to help identify and interpret the structural shocks or innovations in the model [see Stock and Watson (2018), and Mertens and Ravn (2013)]. This approach relies on the assumption that the exogenous variables (not included in the model specification) capture the unobserved factors that drive the structural changes in the system [see Canova and Ferroni (2022) and Miranda-Agrippino and Ricco (2023)].

These are some techniques used in empirical macro-economy to tackle the structural shocks identification problem in structural VAR models.

1.2 MGARCH models and identification

While structural VAR models find interest in analyzing policy and business, multivariate Generalized Autoregressive Conditional Heteroskedasticity (MGARCH) models emphasize volatility and correlation dynamics in asset returns, risk management, portfolio optimization and derivative pricing. How structural shocks of such models are identified and studied are the subject of the coming subsection.

1.2.1 MGARCH model

MGARCH model is an extension of univariate GARCH. Recall, a univariate GARCH model $y_t = \mu_t + \epsilon_t$ is such that :

- $\mu_t = \mathbb{E}(y_t|\mathcal{F}_{t-1}) = \mathbb{E}_{t-1}(y_t)$ the conditional expectation of y_t given $\mathcal{F}_{t-1}$ the past information set of y_t up to $(t-1)$
- ϵ_t the noise process.

In the particular case of GARCH models,

$$\epsilon_t = \sigma_t Z_t \tag{1.8}$$

where

- Z_t are independently and identically distributed (i.i.d.) random variables with mean 0 and variance 1,
- given $\mathcal{F}_{t-1}$, the conditional variance $\mathbb{V}\text{ar}(y_t|\mathcal{F}_{t-1}) = \sigma_t^2$ of y_t is in the form : $\sigma_t^2 = c + \sum_{j=1}^{p} \phi_j \sigma_{t-j}^2 + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}^2$. The model is thus said to have a GARCH(p,q) structure. Generally, GARCH(1,1) fits quite well the data.

Engle (1982) and Bollerslev (1986) introduced Eq. (1.8) to study the volatility phenomenon of y_t.

Similarly to the univariate case, to study the volatility phenomenon in the multivariate case of a $(m \times 1)$ vector process $\boldsymbol{y}_t$ presenting a conditional heteroskedasticity phenomenon, one can express $\boldsymbol{y}_t$ as :

$$\boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t$$

where :

- $\boldsymbol{\mu}_t = \mathbb{E}(\boldsymbol{y}_t|\boldsymbol{\mathcal{F}}_{t-1})$, the conditional expectation of $\boldsymbol{y}_t$ given the past information set $\boldsymbol{\mathcal{F}}_{t-1}$. Note here that $\boldsymbol{\mu}_t$ could be a VAR(p) model for example.
- $\boldsymbol{\varepsilon}_t$ the vector of error term (as in the univariate case) with conditional mean equal to zero and based on its heteroskedasticity, $\mathbb{V}\text{ar}(\boldsymbol{\varepsilon}_t|\boldsymbol{\mathcal{F}}_{t-1}) = \boldsymbol{H}_t$

By simple extension to Eq. (1.8), one has:

$$\boldsymbol{\varepsilon}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{Z}_t \tag{1.9}$$

where:

- $\boldsymbol{Z}_t$ are i.i.d. m-dimensional multivariate random vectors with mean $\mathbf{0}$ and identity covariance matrix $\boldsymbol{I}_m$ ($\boldsymbol{Z}_t \sim iid(\mathbf{0}, \boldsymbol{I}_m)$)
- $\boldsymbol{H}_t^{1/2}$ the matrix square root of $\boldsymbol{H}_t$ (the conditional covariance matrix of $\boldsymbol{y}_t$) obtained for example by eigenvalue decomposition.

Note :

$$\begin{aligned}\mathbb{V}\text{ar}(\boldsymbol{y}_t|\boldsymbol{\mathcal{F}}_{t-1}) = \mathbb{V}\text{ar}_{t-1}(\boldsymbol{y}_t) &= \mathbb{V}\text{ar}_{t-1}(\boldsymbol{\varepsilon}_t) \\ &= \boldsymbol{H}_t^{1/2}\underbrace{\mathbb{V}\text{ar}_{t-1}(\boldsymbol{Z}_t)}_{\mathbf{I}_m}(\boldsymbol{H}_t^{1/2})^\top \\ &= \boldsymbol{H}_t\end{aligned}$$

Various specification of $\boldsymbol{H}_t$ have been proposed in the literature, among which, one can distinguish : the VEC and DVEC models, the CCC models, the DCC models, the BEKK models and the factor models. [see Bauwens, Laurent, and Rombouts (2006)].

1.2.1.1 VEC and DVEC models

Starting by vectorizing the matrix (specially the half-vectorization) $\boldsymbol{H}_t$, Bollerslev, Engle, and Wooldridge (1988) proposed the following specification to catch the volatility in the errors :

$$vech(\boldsymbol{H}_t) = \boldsymbol{\Theta}_0 + \sum_{j=1}^{p}\Phi_j vech(\boldsymbol{H}_{t-j}) + \sum_{j=1}^{q}\boldsymbol{\Theta}_j vech(\varepsilon_{t-j}\varepsilon_{t-j}^\top) \tag{1.10}$$

where :

- each element of $\boldsymbol{H}_t$ is a function of lagged values of $\boldsymbol{H}_t$ and the squared and cross products of lagged variables or errors,
- $\boldsymbol{\Theta}_0$ is a $[m(m+1)/2] \times 1$ column vector
- each of the coefficient matrices $\boldsymbol{\Theta}_j$ and $\boldsymbol{\Phi}_j$, is a $[m(m+1)/2] \times [m(m+1)/2]$ matrix

Engle and Kroner (1995) shown that the VEC(p, q) model specified at (1.10) is stationary if and only if all eigen-values of the matrix $\sum_{j=1}^{p} \boldsymbol{\Phi}_j + \sum_{j=1}^{q} \boldsymbol{\Theta}_j$ are within the unit circle. In such case of unit root, the vector process $\boldsymbol{\varepsilon}_t$ is covariance stationary with unconditional covariance matrix given by $\boldsymbol{H}_t$ (Hafner & Herwartz, 2006).

The main puzzle of this VEC model is the number of parameters involved. And to solve that puzzle, Bollerslev et al. (1988) proposed a simpler representation by assuming the matrices $\boldsymbol{\Theta}_j$ and $\boldsymbol{\Phi}_j$ to be diagonal. Hence the model is known as the diagonal vector model (DVEC). In most cases, the VEC(1,1) is used and produce adequate results.

1.2.1.2 BEKK models

A frequent model of $\boldsymbol{H}_t$ used in empirical studies is the BEKK models named after the work of Baba, Engel, Kraft and Kroner. It presents the advantage of the positive definiteness of the $\boldsymbol{H}_t$ and its number of parameters grows linearly with the number of assets. In addition, the BEKK model is parsimonious and suitable for relatively large set of assets; and its a special case of the VEC model.

The BEKK(1,1) specification of Engle and Kroner (1995) is:

$$\boldsymbol{H}_t = \mathbf{C}\mathbf{C}^\top + \mathbf{A}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top\mathbf{A}^\top + \mathbf{B}^\top\boldsymbol{H}_{t-1}\mathbf{B}$$

where $\mathbf{C}$ is a lower triangular matrix, $\mathbf{A}$ and $\mathbf{B}$ are $m \times m$ matrices. Based on the symmetric parametrization of the model, $\boldsymbol{H}_t$ is almost surely positive definite provided that $\mathbf{C}\mathbf{C}^\top$ is positive definite.

1.2.1.3 Constant Conditional Correlation (CCC) models

In the same schema as in VEC model, in the purpose of reducing the number of parameters involved in the specification of the structure of the $\boldsymbol{H}_t$ matrix, Bollerslev (1990), proposed another specification known as the Constant Conditional Correlation (CCC) models. In this new schema, he proposed to render the conditional correlation matrix constant. Hence

$$\boldsymbol{H}_t = \boldsymbol{D}_t\mathbf{R}\boldsymbol{D}_t \tag{1.11}$$

where :

- $\mathbf{R}$ is a correlation matrix having ones on its diagonal and it is positive definite. Off its diagonal the ρ_{ij} is the constant conditional correlation between $\varepsilon_{i,t}$ and $\varepsilon_{j,t}$
- $\boldsymbol{D}_t = \text{diag}(\sigma_{1,1,t}^{1/2}, \ldots, \sigma_{m,m,t}^{1/2})$ with $\sigma_{i,j,t}^{1/2}$ the conditional standard deviation between components i and j at time t. These may come from univariate GARCH models : $\sigma_{i,i,t} = c_i + \phi_i\sigma_{i,i,t-1} + \theta_i\epsilon_{i,t-1}^2, i = 1, \ldots, m$

In such model, the number of parameters is $m(m+5)/2$. Although, the CCC model reduces the number of parameters, the constant assumption is too restrictive and in some case undesirable. Tse and Tsui (2002) proposed then a time-varying schema named the Dynamics Conditional Correlation (DCC) to circumvent the constant correlation restriction imposed by the CCC model.

1.2.1.4 Dynamic Conditional Correlation (DCC) models

In opposition to the CCC models were the correlations among asserts are assumed to be constant, the DCC model allows the correlation to change over time based on past information. Thus :

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t \tag{1.12}$$

with :

- $\boldsymbol{D}_t = \text{diag}(\sigma_{1,1,t}^{1/2}, \ldots, \sigma_{m,m,t}^{1/2})$ just as before
- $\boldsymbol{R}_t$ is a correlation matrix having ones on its diagonal and it is positive definite. Off its diagonal the ϕ_{ij} is the dynamic conditional correlation between $\varepsilon_{i,t}$ and $\varepsilon_{j,t}$

Often in practise, the CCC models are widely used due to their simplicity and reduced number of parameters since it only requires to estimate the conditional variances and then models all conditional correlations as a constant. Nevertheless, in a stating such that the number of assets are not quite huge, one can opt for the BEKK models where the positive definiteness of $\boldsymbol{H}_t$ is guaranteed.

Once the $\boldsymbol{H}_t$ is specified, one usually refer to maximum likelihood estimation to estimated the models parameters with a limited scope.

1.2.2 The identification problem at a glance

From all above, one can deduce that $\boldsymbol{H}_t^{1/2}$ is any $m \times m$ positive definite matrix such that the covariance matrix $\boldsymbol{H}_t$ of $\boldsymbol{y}_t$ is :

$$\boldsymbol{H}_t = \boldsymbol{H}_t^{1/2} \boldsymbol{H}_t^{1/2\top}. \tag{1.13}$$

Defining $\boldsymbol{H}_t^{1/2}$ that way is not quite informative without further assumptions. By introducing a rotation matrix $\mathbf{R}_\delta$ such that $\mathbf{R}_\delta \mathbf{R}_\delta^\top = \mathbf{I}_m$, where δ is the angle of rotation, one gets another representation $\widetilde{\boldsymbol{H}}_t^{1/2}$ of the $\boldsymbol{H}_t^{1/2}$ as pointed by Hafner, Herwartz, and Maxand (2022).

$$\widetilde{\boldsymbol{H}}_t^{1/2} := \boldsymbol{H}_t^{1/2} \boldsymbol{R}_\delta \implies \widetilde{\boldsymbol{H}}_t^{1/2} \widetilde{\boldsymbol{H}}_t^{1/2\top} = \boldsymbol{H}_t^{1/2} \boldsymbol{R}_\delta \boldsymbol{R}_\delta^\top \boldsymbol{H}_t^{1/2\top} = \boldsymbol{H}_t \tag{1.14}$$

Hence, $\boldsymbol{H}_t^{1/2}$ definition is not unique. Thus, it implies why the structural shocks identification problem of the model rises again. And thereforth, the identification of the structural shocks underlying the volatility and correlation dynamics of the model is an issue, although many approaches have been proposed to handle that. This is the subject of the next sections.

1.2.3 Structural shocks identification in MGARCH model

In purpose of fully specifying the dynamics of $\boldsymbol{\varepsilon}_t$ and not only its conditional variance structure, the Cholesky decomposition or the symmetric eigenvalue decomposition of $\boldsymbol{H}_t$

are used aside the parametric identification techniques identified in the above section. Hafner et al. (2022) stressed that Eq.(1.9) can be viewed as a structural scheme, the jth column of $\boldsymbol{H}_t^{1/2}$ formalizing how single shocks Z_{jt} in $\boldsymbol{Z}_t$ affects $\boldsymbol{y}_t$ and similarly the ith row of $\boldsymbol{H}_t^{1/2}$ shows the contribution of each shock in $\boldsymbol{Z}_t$ to the volatility received by a single time defined y_{it}.

One can then proceed by using the traditional approaches of shocks identification reveal in the VAR section above especially the data heteroskedasticity based methods. But in case of observed volatility phenomenon, some other approaches have proven their usefulness.

1.2.3.1 Identification by Cholesky decomposition

The Cholesky decomposition is a matrix factorization method that decomposes a symmetric, positive-definite matrix into the product of a lower triangular matrix and its transpose.

For $\boldsymbol{H}_t$ representing the covariance matrix and $\mathbf{L}$ be a lower triangular matrix with positive elements on its diagonal:

$$\boldsymbol{H}_t = \mathbf{L}\mathbf{L}^\top \tag{1.15}$$

Since the covariance matrix of the error term of the MGARCH model is symmetric and positive definite by construction, the Cholesky decomposition seem justified as the decomposition is unique. Hence it have been popular in identification of shocks related to market. Thus, the lower triangular matrix resulting from the Cholesky decomposition represents the structural impact matrix. Each row of the matrix represents the impact of the structural shocks on a particular variable. The elements of each row indicate the contemporaneous relationships between the variables and the shocks. The Cholesky decomposition provides a specific ordering of the shocks and by examining the structural impact matrix, one can identify the variables that are most affected by each structural shock.

1.2.3.2 Identification by eigenvalue decomposition

Also known as the spectral decomposition of matrix, eigenvalue decomposition also provide a hint in identification of the structural shocks by the transformation of the estimated conditional covariance matrix of the error term into a diagonal matrix of eigenvalues and an orthogonal matrix of eigenvectors.

If $\boldsymbol{H}_t$ represents the covariance matrix, the vector $\mathbf{v}$ of eigenvectors $v_1, \dots v_m$ and the vector $\boldsymbol{\lambda}$ of eigenvalues $\lambda_1, \dots, \lambda_m$ are the set of vectors and values such that

$$\boldsymbol{H}_t\mathbf{v} = \boldsymbol{\lambda}\mathbf{v} \tag{1.16}$$

The eigenvalues λ_i represent the variances or volatilities of the orthogonal shocks, while the eigenvectors represent the impact of each structural shock on the variables. The eigenvectors are also known as the structural impulse response functions. By examining the eigenvalues, you can identify the relative importance of each structural shock in explaining

the variabilities in the variables. Larger eigenvalues indicate higher contributions of the corresponding shocks to the overall volatility.

The identification of structural shocks through eigenvalue decomposition is based on the assumption that the shocks are uncorrelated (orthogonal) and have constant variances. This assumption may not always hold in practice, particularly in financial markets where correlations and volatilities can change over time.

1.2.3.3 Data-based identification approach

Although eigenvalue decomposition or Cholesky decomposition of $\boldsymbol{H}_t$ are used, we need to unravel some implication of such approaches. The eigenvalue decomposition implies a symmetry between the impact of shocks of one to another series, which might lacks economic justification in some contexts. On the other hand, the Cholesky decomposition presumed ordering of variables thus a linearity of impact, which might not let some flexibility in the models identification. Based on those limitation, Hafner et al. (2022) have proposed a moment based identification technique. Their took benefit from the non-normality of the innovation vector and from the mutual independence of its components to propose a moment based identification techniques.

1.3 Conclusion

Throughout this first chapter, we uncovered what are structural innovations in VAR models used mostly in macroeconomics analysis to study the dynamic between variables and in MGARCH model to study the volatility and correlation dynamic in financial assets analysis. We discussed the main techniques used in empirical studies to unveil the structural shocks and dynamics underlying the financial and macroeconomics fields. Although, these techniques are very powerful in there respective fields, they present some limitation in a sense that some restrictions or assumptions need to be settled before taking fully advantage of these methods. And with the present era of data flow and development of new techniques of multivariate data analysis, some techniques like the independent component has been proposed to uncover hidden latent variables in high-dimensional data. In the next chapter, we will see how the latter is used in structural shocks identification and how can we used a nonlinear form of it to do the same job.

Chapter 2

Independent component analysis in structural shocks identification

Here, we start by giving an overview of the fundamentals of independent component analysis; showing how it is used in empirical study to deal with the structural shocks identification. Afterward, in the second section, we provide hint about the nonlinear ICA called kernelICA and how this can help in recovery of the structural shocks in MGARCH models.

2.1 Independent component analysis

Also known as blind source separation (BSS) problem, ICA is the decomposition of an unknown mixture of non-Gaussian signals into its independent component signals (Cardoso, 1998). The best know example of BSS in the literature is the coktail-party problem of Cherry (1953). In the following, we discuss what is ICA and its usefulness in structural innovations identification in VAR or MGARCH models.

2.1.1 Overview

Independent components analysis (ICA) is a multivariate statistical technique that seeks to uncover hidden variables in high-dimensional data, and as such belongs to the class of latent variable models (Izenman, 2008). It linearly transforms a random vector into independent factors, called components, with a scale and order indeterminacy. All these under the assumptions that the observed data are mixtures of non-Gaussian and independent processes. ICA's schema is as follow :

Considering $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)^\top$ be a m-dimensional vector such that $\mathbb{E}(\boldsymbol{x}) = \mathbf{0}$ and $\mathbb{E}(\boldsymbol{x}\boldsymbol{x}^\top) = \boldsymbol{\Sigma}_{\boldsymbol{x}}$ assumed to be positive definite. And we also assumed that $\boldsymbol{x}$ is generated by a linear combination of r ($r \leq m$) independent components (for simplicity we assume $r = m$). That is

$$\boldsymbol{x} = \mathbf{A}\boldsymbol{s} \tag{2.1}$$

where

- $\mathbf{A}$ is an unknown $m \times m$ full rank matrix, called the *mixing matrix* in ICA jargon, with elements a_{ij} representing the effect of $\mathbf{s}_j$ on $\boldsymbol{x}$
- $\boldsymbol{s} = (s_1, \ldots, s_m)^\top$ is the vector of unobserved independent components.

The problem now is to estimate both $\mathbf{A}$ and $\boldsymbol{s}$ from the only observed $\boldsymbol{x}$. Hence, ICA looks for an $(m \times m)$ matrix $\mathbf{W} = \mathbf{A}^{-1}$, called the *unmixing matrix*, such that the components given by

$$\hat{s} = \mathbf{W}\boldsymbol{x} \tag{2.2}$$

are as independent as possible. To achieve that, ICA sets some crucial assumptions :

- the statistically independence of the components,
- the observed data can be represented as a linear combination of these components,
- non-Gaussian distributions of the independent components with the possible exception of one. Thus, ICA is meaningful only if the involved random variables up to one is non-Gaussian. Recall that for Gaussian random variables, independence is equivalent to uncorrelatedness. Therefore any decorrelated factors, such as the one obtained by PCA will yield independent components.

It's also to note that, in spite of previous assumptions, ICA can't determine either the sign or the order of the independent components. Hence, the components are said to be identified up to scale and order indeterminacy.

2.1.2 Estimation of independent components

ICA aims to obtain the independent latent factors as linear combinations of the data, as such, the methods used for estimating the ICs impose the restrictions that the rows of $\mathbf{W}$ are the directions that maximimze the independence of $\hat{\mathbf{s}}_t$.

Notwithstanding the various number of estimation algorithm available in the literature, we can state that the core different is based on the objective function optimized. This includes minimizing a high order cumulants, maximizing the mutual information of the outputs and so on. In the following, we will uncover the JADE and FastICA algorithm due to their popularity and efficiency in terms of computation optimization.

Before applying these algorithms, and as such in many multivariate analysis, it is useful to standardize the data vector $\boldsymbol{x}$. As such PCA is usually considered as a pre-processing step for ICA. Hence (2.2) becomes

$$\hat{s} = \mathbf{W}'\boldsymbol{z} \tag{2.3}$$

where $\mathbf{W}$' is the new unmixing matrix and $\boldsymbol{z}$ the m-dimensional vector of standardized data.

2.1.2.1 Joint Approximate Diagonalization of Eigen-matrices (JADE)

Developped by Cardoso and Souloumiac (1993) which estimates the ICs from (2.3) by maximizing their non-Gaussianity. Is needed to precise that, since under the non-Gaussianity assumption, the information provided by the covariance matrix of the data is not sufficient and higher-order information is needed, Cardoso and Souloumiac (1993) use the cumulants which are the coefficients of the Taylor series expansion of the logarithm of the characteristic function. JADE uses an iterative process of Jacobi rotations to solve the joint diagonalization of several fourth-order cumulants matrices. It is a very efficient algorithm in low dimensional problems, but when the dimension increases, it requires high computational cost.

2.1.2.2 Fast Fixed-Point Algorithm (FastICA)

Proposed by Hyvärinen (1999) and Hyvärinen and Oja (1997), it estimates (2.3) by maximizing their univariate kurtosis. Thus, it searches for the directions of projection that maximize the absolute value of the kurtosis of $\hat{\boldsymbol{s}}$. As kurtosis is very sensitive to outliers, Hyvärinen (1999) develops a more robust version of FastICA that measures the non-Gaussianity of the ICs by using an approximation of the negentropy instead of the kurtosis coefficient. Therefore, FastICA converges quickly in a few number of iterations.

Concretely, considering we observe m linear mixtures $x_1, x_2, \ldots, x_m$ (denoted $\boldsymbol{x}$) of m independent components $s_1, s_2, \ldots, s_m$ (denoted $\boldsymbol{s}$) such that

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \ldots + a_{im}s_m, \text{ for all } i \iff \boldsymbol{x} = \mathbf{A}\boldsymbol{s} \ ;$$

thus, the unmixing matrix $\mathbf{W} = \mathbf{A}^{-1}$ is such that $\boldsymbol{s} = \mathbf{W}\boldsymbol{x}$.

The objective in this FastICA algorithm, is to find $\mathbf{w}$ (given row of $\mathbf{W}$) that maximizes the approximation to the negentropy subject to the sphering constraint $\mathbb{E}\{(\mathbf{w}\boldsymbol{x})^2\} = \|\mathbf{w}\|^2 = 1$. This means that $\mathbf{w}$ is to be that direction that makes the density of the one-dimensional s as far away from the Gaussian density as possible. To achieve this, a family of approximations based on the maximum-entropy principles were developed :

$$J(x) = [\mathbb{E}[G(x)] - \mathbb{E}[G(v)]]^2 \tag{2.4}$$

where

- x is assumed to be of mean 0 and unit variance,
- v a Gaussian random variable of zero mean and unit variance,
- G some nonquadratic function. By choosing this function wisely, one can obtain a better approximations and in particular a not too fast growing G gives robust approximations to negentropy. The two nonquadratic function in Table 2.1 have proven useful. Note that for the `logcosh` density, $1 \leq \alpha \leq 2$.

Density	$G(x)$	$g(x) = G'(x)$	$g'(x) = G''(x)$
`logcosh`	$\frac{1}{\alpha}\texttt{logcosh}(\alpha x)$	$\texttt{tanh}(\alpha x)$	$\alpha(1 - \texttt{tanh}^2(\alpha x))$
`exp`	$-e^{-x^2/2}$	$xe^{-x^2/2}$	$(1 - y^2)e^{-x^2/2}$

Table 2.1: Nonquadratic density functions and their first and second derivatives

To find one direction $\mathbf{w}$ of the independent components, FastICA uses a Newton-Iteration scheme :

1. Center and whiten the data to obtain a mean zero and unit variance data
2. Choose G to be any nonquadratic density function with first and second partial derivatives g and g', respectively.
3. Choose an initial version of the m-vector $\mathbf{w}$ with unit norm and let $k = 0$
4. Let $\mathbf{w}_{k+1} \leftarrow \mathbb{E}(\mathbf{x}g(\mathbf{w}_k^\top \mathbf{x})) - \mathbf{w}_k \mathbb{E}(g'(\mathbf{w}_k^\top \mathbf{x}))$. In practice, the samples means are used.
5. Let $\mathbf{w}_{k+1} \leftarrow \mathbf{w}_{k+1}/\|\mathbf{w}_{k+1}\|$.
6. Iterate between steps 4 and 5. Stop when convergence is attained : $|\mathbf{w}_{k+1}^\top \mathbf{w}_k| \geq (1-\epsilon)$

Table 2.2: FastICA algorithm for determining a single source component

In the framework where, we are looking for a certain number of independent components, we need to run this FastICA the given certain number of times. Each time we estimate a different component, we need to orthogonalize the vectors $\mathbf{w}$ at each iteration by projecting the current solution $\mathbf{w}_k$ to the space orthogonal to the columns of the mixing matrix that were previously estimated. Two methods are available to address this : the deflation and the parallel (Izenman, 2008).

- `deflation` : the single component routine is run to find a new component, that is orthogonalized using the Gram-Schmidt method with respect to all previously found components, and then the resulting new component is normalized.
- `parallel` : the single component is carried out in parallel for each component to be extracted, and then a symmetric orthogonalisation is carried out on all components simultaneously.

Factually, the `deflation` method extracts independent components sequentially one-at-a-time, whereas the `parallel` method extracts all the independent components at once. Both algorithms are summarized in the Table 2.3

Table 2.3: Deflation and parallel algorithms for extracting multiple independent source component

---

`deflation algorithm`

1. Center and whiten the data to obtain a mean zero and unit variance data.
2. Decide the number m of IC to be extracted.
3. Choose G to be any nonquadratic density function with first and second partial derivatives g and g', respectively.
4. For $k = 1, 2, \ldots, m$,
   - randomly initialize the m-vector $\mathbf{w}_k$ with unit norm
   - Let $\mathbf{w}_{k+1} \leftarrow \mathbb{E}(\mathbf{x}g(\mathbf{w}_k^\top \mathbf{x})) - \mathbf{w}_k \mathbb{E}(g'(\mathbf{w}_k^\top \mathbf{x}))$ be the `FastICA` single component update for $\mathbf{w}_k$. In practice, the expectations are estimated using the sample averages.
   - Use the Gram-Schimdt algorithm to orthogonalize $\mathbf{w}_k$ with respect to the previously chosen $\mathbf{w}_1, \ldots, \mathbf{w}_{k-1}$:

     $$\mathbf{w}_k \leftarrow \mathbf{w}_k - \sum_{j=1}^{k-1} (\mathbf{w}_k^\top \mathbf{w}_j)\mathbf{w}_j.$$

   - Let $\mathbf{w}_k \leftarrow \mathbf{w}_k / \|\mathbf{w}_k\|$.
   - Iterate $\mathbf{w}_k$ until convergence : $|\mathbf{w}_{k-1}^\top \mathbf{w}_k| \geq (1 - \epsilon)$.
5. Set $k \leftarrow k + 1$. If $k \leq m$, return to step 4.

`parallel algorithm`

1. Center and whiten the data to obtain a mean zero and unit variance data.
2. Decide the number m of IC to be extracted and the G function to use.
3. Randomly initialize each m-vector $\mathbf{w}_1, \ldots, \mathbf{w}_m$ with unit norm. Let $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_m)^\top$.
4. Carry out a symmetric orthogonalization of $\mathbf{W}$ by $\mathbf{W} \leftarrow (\mathbf{W}\mathbf{W}^\top)^{-1/2}\mathbf{W}$.

5. For each $k = 1, \ldots, m$, let $\mathbf{w}_k \leftarrow \mathbb{E}(\mathbf{x}g(\mathbf{w}_k^\top \mathbf{x})) - \mathbf{w}_k \mathbb{E}(g'(\mathbf{w}_k^\top \mathbf{x}))$ be the `FastICA` single component update for $\mathbf{w}_k$.
6. Carry out another symmetric orthogonalization of $\mathbf{W}$.
7. If convergence has not occured, return to step 5.

Harpaz (2007) emphasizes in his paper that the advantage of estimating one IC at a time is in cases where not all of the components need to be estimated. Also the `FastICA` algorithms convergence is cubic whereas the other algorithms based on gradient descent methods are typically linear.

2.1.3 ICA for structural shocks identification

ICA have been used in unveiling the structure of multivariate financial time series by Back and Weigend (1997) and many others in the literature. In the context of assert returns, it can help to identify and separate the latent sources or components of volatility from the observed data. To achieve that, it's used essentially on the residuals since these residuals capture the unexplained variability in the data after accounting for the conditional volatility modeled by the MGARCH model specified. Hence, a typical roadmap for the use of ICA in structural shocks identification is :

1. gather the data and ensuring its stationary,
2. fit an adequate MGARCH model and retrieve the conditional variances between the variables,
3. get the residuals from the fitted MGARCH models and apply the ICA to uncover the latent sources of volatility,
4. from the identified sources of variability, study the volatility impulse response, for a given shocks on a considered horizon,
5. perform predictions if necessary.

2.2 Kernel independent component analysis

Previous section show that Independent Component Analysis (ICA) is a powerful statistical technique that aims to find statistically independent components from observed data, allowing for the representation of complex data as a linear combination of these independent sources. However, in certain real-world applications, which turns to usually be the case, data may not be this linear combination of independent sources, leading to limitations in traditional ICA methods. To address this issue, kernel Independent Component Analysis (kernelICA) emerged as an extension of ICA, leveraging the concept of kernel methods to overcome the linearity constraint. Hence, this section will treat of what we can gain from these kernel approach of ICA in structural shocks identification.

2.2.1 Overview

Developped by Bach and Jordan (2002), the main differences with the classical ICA is that they propose to search for the contrast function in a reproducing kernel Hilbert space (RKHS), then perform its optimization. This approach reduces to finding the eigenvalues and eigenvectors of a certain matrix derived from a kernelized version of canonical variates analysis (CVA) [see Izenman (2008)].

Bach and Jordan proposed two ICA contrast functions. The first one is based on the $\mathcal{F}$-correlation which can be obtained by computing the first canonical correlation in a reproducing kernel Hilbert space. The second is based on computing the entire canonical correlation analysis spectrum known as the "generalized variance".

2.2.2 KernelICA and structural shocks identification

As specified previously, kernelICA is an extension of ICA to the kernel space. In the following, we will first recap the essentials of kernel methods and study how these can help in structural innovations identification using kernelICA developped by Bach and Jordan (2002).

2.2.2.1 Kernel methods

Kernel methods use basis functions ϕ to map the input data from the input space $\mathcal{I}$ into a different space (higher dimension) called the feature space $\mathcal{F}$. The purpose of this mapping is to uncover a linear relationship between the input data, when in the input space, the linear relationship is not quite obvious.

After this mapping, simple known algorithms can be trained on the new feature space, instead of the input space, which can result in an increase in the performance of the models.

To avoid explicitly mapping the input data into the features space through the mapping function ϕ, the input data are represented via their pairwise similarity values comprising a kernel matrix $\mathbf{K}$. $\mathbf{K}$ is a $m \times m$ kernel matrix defined as :

$$\mathbf{K} = \begin{pmatrix} K(x_1, x_1) & K(x_1, x_2) & \dots & K(x_1, x_m) \\ K(x_2, x_1) & K(x_2, x_2) & \dots & K(x_2, x_m) \\ \vdots & \vdots & \ddots & \vdots \\ K(x_m, x_1) & K(x_m, x_2) & \dots & K(x_m, x_m) \end{pmatrix} \tag{2.5}$$

where $K : \mathcal{I} \times \mathcal{I} \to \mathbb{R}$ is a kernel function on any two points in input space, which should satisfy the condition

$$K(x_i, x_j) = \phi(x_i)^\top \phi(x_j) \tag{2.6}$$

Once the kernel matrix has been computed, we no longer even need the input points x_i, as all operations involving only dot products in the feature space can be performed over the $m \times m$ kernel matrix $\mathbf{K}$.

2.2.2.2 The $\mathcal{F}$-correlation

Considering x_1 and x_2 as random variable in $\mathbb{R}^n$, K_1 and K_2 be Mercer kernels with features maps Φ_1 and Φ_2 and features space $\mathcal{F}_1, \mathcal{F}_2$, the canonical correlation $\rho_{\mathcal{F}}$ between $\Phi_1(x_1)$ and $\Phi_2(x_2)$ defined by :

$$\rho_{\mathcal{F}} = \max_{(f_1,f_2)\in\mathcal{F}_1\times\mathcal{F}_2} \texttt{corr}(\langle\Phi_1(x_1), f_1\rangle, \langle\Phi_2(x_2), f_2\rangle)$$

is equal to

$$\rho_{\mathcal{F}} = \max_{(f_1,f_2)\in\mathcal{F}_1\times\mathcal{F}_2} \texttt{corr}(f_1(x_1), f_2(x_2)) \tag{2.7}$$

While so far, we have defined $\mathcal{F}$-correlation in terms of a population expectation, we generally do not have access to the population but rather to a finite sample. Thus we need to develop an empirical estimate of the $\mathcal{F}$-correlation. Thus, the $\hat{\rho_{\mathcal{F}}}$-correlation is defined by :

$$\widehat{\rho_{\mathcal{F}}} = \max_{(f_1,f_2)\in\mathcal{F}_1\times\mathcal{F}_2} \frac{\widehat{\texttt{cov}}(f_1(x_1), f_2(x_2))}{(\widehat{\texttt{var}}\{f_1(x_1)\})^{1/2}(\widehat{\texttt{var}}\{f_2(x_2)\})^{1/2}} \tag{2.8}$$

where by definition

$$\begin{aligned} \widehat{\texttt{cov}}(f_1(x_1), f_2(x_2)) &= n^{-1}\boldsymbol{\alpha_1}^\top\mathbf{K}_1\mathbf{K}_2\boldsymbol{\alpha_2}, \\ \widehat{\texttt{var}}\{f_1(x_1)\} &= n^{-1}\boldsymbol{\alpha_1}^\top\mathbf{K}_1^2\boldsymbol{\alpha_1}, \\ \widehat{\texttt{var}}\{f_2(x_2)\} &= n^{-1}\boldsymbol{\alpha_2}^\top\mathbf{K}_2^2\boldsymbol{\alpha_2}, \end{aligned} \tag{2.9}$$

$\boldsymbol{\alpha_1} = (\alpha_{11}, \ldots, \alpha_{1n})^\top$, $\boldsymbol{\alpha_2} = (\alpha_{21}, \ldots, \alpha_{2n})^\top$, $\mathbf{K}_1$ and $\mathbf{K}_2$ the $(n \times n)$ Gram matrices associated with x_1 and x_2.

So,

$$\widehat{\rho_{\mathcal{F}}} = \max_{(f_1,f_2)\in\mathcal{F}_1\times\mathcal{F}_2} \frac{\boldsymbol{\alpha_1}^\top\mathbf{K}_1\mathbf{K}_2\boldsymbol{\alpha_2}}{(\boldsymbol{\alpha_1}^\top\mathbf{K}_1^2\boldsymbol{\alpha_1})^{1/2}(\boldsymbol{\alpha_2}^\top\mathbf{K}_2^2\boldsymbol{\alpha_2})^{1/2}} \tag{2.10}$$

Differentiating (2.10) with respect to $\boldsymbol{\alpha_1}$ and $\boldsymbol{\alpha_2}$ and setting the result to 0 yields the generalized eigen-equation :

$$\mathcal{K}\boldsymbol{\alpha} = \lambda\mathcal{D}\boldsymbol{\alpha} \tag{2.11}$$

where $\mathcal{K} = \begin{pmatrix} \mathbf{0} & \mathbf{K}_1\mathbf{K}_2 \\ \mathbf{K}_2\mathbf{K}_1 & \mathbf{0} \end{pmatrix}$, $\mathcal{D} = \begin{pmatrix} \mathbf{K}_1^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_2^2 \end{pmatrix}$, $\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{pmatrix}$

Out of this eigen-equation, the problem is that $\mathcal{D}$ will be singular due to the centering of data which renders the Gram matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ singular. Also, in feature space, all pairs of "kernelized canonical variates" will be perfectly correlated. To circumvent these issues, Bach and Jordan developed a *regularized* form that provides a useful estimator.

The regularized form is in the same spirit as ridge regression by penalizing the RKHS norm of f_1 and f_2 by a same small constant $\kappa > 0$. Hence the regularized version is :

$$\widehat{\rho_{\mathcal{F}}} = \max_{(f_1,f_2)\in\mathcal{F}_1\times\mathcal{F}_2} \frac{\boldsymbol{\alpha_1}^\top\mathbf{K}_1\mathbf{K}_2\boldsymbol{\alpha_2}}{(\boldsymbol{\alpha_1}^\top(\mathbf{K}_1 + \kappa\mathbf{I}_n)^2\boldsymbol{\alpha_1})^{1/2}(\boldsymbol{\alpha_2}^\top(\mathbf{K}_2 + \kappa\mathbf{I}_n)^2\boldsymbol{\alpha_2})^{1/2}} \tag{2.12}$$

Differentiating (2.12) with respect to $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ and setting the results equal zero yields the following in matrix form :

$$\mathcal{K}\boldsymbol{\alpha} = \lambda\mathcal{D}_\kappa\boldsymbol{\alpha} \tag{2.13}$$

where $\mathcal{K}$ is as in (2.11) and

$$\mathcal{D} = \begin{pmatrix} (\mathbf{K}_1 + +\kappa\mathbf{I}_n)^2 & \mathbf{0} \\ \mathbf{0} & (\mathbf{K}_2 + \kappa\mathbf{I}_n)^2 \end{pmatrix} \tag{2.14}$$

For optimization purpose, Bach and Jordan (2002) suggest $\kappa = n \times 10^{-2}$ if $n \leq 1000$, otherwise $\kappa = n \times 10^{-3}$; where Leurgans, Moyeed, and Silverman (1993) consider using cross-validation to determine a good choice of κ.

This framework can be extended to more than 2 variables which is straightforward. Thus, denoting $\mathcal{K}_\kappa$ the $nm \times nm$ matrix whose blocks are $(\mathcal{K}_{ij}) = \mathbf{K}_i\mathbf{K}_j$, for $i \neq j$ and $(\mathcal{K}_{ii}) = (\mathbf{K}_i + \kappa\mathbf{I}_n)^2$, and $\mathcal{D}_\kappa$ the $nm \times nm$ block-diagonal matrix with blocks $(\mathbf{K}_i + \kappa\mathbf{I}_n)^2$, we obtain the following generalized eigenvalue problem :

$$\mathcal{K}_\kappa\boldsymbol{\alpha} = \lambda\mathcal{D}_\kappa\boldsymbol{\alpha} \tag{2.15}$$

where

$$\mathcal{K}_\kappa = \begin{pmatrix} (\mathbf{K}_1 + \kappa\mathbf{I}_n)^2 & \mathbf{K}_1\mathbf{K}_2 & \dots & \mathbf{K}_1\mathbf{K}_m \\ \mathbf{K}_2\mathbf{K}_1 & (\mathbf{K}_2 + \kappa\mathbf{I}_n)^2 & \dots & \mathbf{K}_2\mathbf{K}_m \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{K}_m\mathbf{K}_1 & \mathbf{K}_m\mathbf{K}_2 & \dots & (\mathbf{K}_m + \kappa\mathbf{I}_n)^2 \end{pmatrix} \tag{2.16}$$

is the $(nm \times nm)$ covariance matrix of the m vectors $\mathbf{x}_i, i = 1, \dots, m$;

$$\mathcal{D}_\kappa = \begin{pmatrix} (\mathbf{K}_1 + \kappa\mathbf{I}_n)^2 & 0 & \dots & 0 \\ 0 & (\mathbf{K}_2 + \kappa\mathbf{I}_n)^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & (\mathbf{K}_m + \kappa\mathbf{I}_n)^2 \end{pmatrix} \tag{2.17}$$

the $(nm \times nm)$ block-diagonal matrix of the individual covariance matrices, and

$$\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_m \end{pmatrix} \tag{2.18}$$

the nm-vector of eigenvalues.

The minimal eigenvalue of this problem (2.15) will be denoted $\lambda_{min}(\mathcal{K}_\kappa)$.

From (2.15), Bach and Jordan (2002) suggest two possible objective functions for ICA :

- $\hat{I}_{\lambda_\mathcal{F}}(\mathbf{K}_1, \dots, \mathbf{K}_m) = -\frac{1}{2}\texttt{log}(\lambda_{min}(\mathcal{K}_\kappa))$ for *kernelized canonical correlation analysis* known as kernelICA-KCCA algorithm

- $\hat{I}_{\delta_{\mathcal{F}}}(\mathbf{K}_1, \ldots, \mathbf{K}_m) = -\frac{1}{2}\texttt{log}(\texttt{det}(\mathcal{K}_\kappa)/\texttt{det}(\mathcal{D}_\kappa))$ for *kernel generalized variance* algorithm known as kernelICA-KGV

Both objective functions are functions of the Gram matrices $\mathbf{K}_1, \ldots, \mathbf{K}_m$ through the unmixing matrix $\mathbf{W}$ and therefore, can be optimized with respect to that matrix.

Obsiously, the huge dimension of the matrices ($nm \times nm$) is an issue for computation purposes. This is handle by Bach and Jordan (2002) by using a low-rank approximations to the m Gram matrices through incomplete Cholesky decomposition. These kernelICA algorithms are known to outperformed all classical ICA methods developped in section 2.1.

2.2.2.3 Structural Shocks identification through kernelICA

As of classical ICA methods in section 2.1.3, in order to separate the shocks and the associated responses completely, ICA methods uniquely identify the structural matrix under non-Gaussianity, kernelICA also, can be used to recover the independent components of a given estimated residuals of an observed MGARCH data as closely as possible to the true shocks; since kernelICA performs better than classical ICA. Hence the steps are the same as defined in section 2.1.3:

1. gather the data and ensuring its stationary,
2. fit an adequate MGARCH model and retrieve the conditional variances between the variables,
3. get the residuals from the fitted MGARCH models and apply the kernelICA to uncover the latent sources of volatility,
4. from the identified sources of variability, study the volatility impulse response,
5. perform predictions if needed.

2.3 Conclusion

This chapter started by presenting the ICA purpose, its algorithms and then moved to kernelICA and how it can help recover structural shocks innovations in MGARCH model. The next chapter will start a simulation study to see how the kernelICA performs in controlled environment. Next, we will move to a real data application and end will the impulse response analysis.

Chapter 3

Simulation study

The objective here in this simulation case, is to compare the performance of simple ICA with the kernelICA in a controlled framework. Hence, we start by simulating the residuals of a MGARCH model for which, we know the structural shocks implication in the error term. Then, by applying the ICA and the kernelICA, we will try to see through a given loss statistics which one perform well in recovering the pre-defined shocks.

3.1 Data generation process

Heading the objective of comparing the performance of ICA and kernelICA on simulated data, we will first define the framework of the data to use.

Therefore, let us recall the error term of a MGARCH model and the ICA model in the following table and emphasize their similarities.

Model	expression	conditons
error term	$\boldsymbol{\varepsilon}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{Z}_t$	$\boldsymbol{Z}_t \sim iid(\mathbf{0}, \boldsymbol{I}_m)$ $\boldsymbol{H}_t^{1/2}(\boldsymbol{H}_t^{1/2})^\top := \boldsymbol{H}_t$
ICA	$\boldsymbol{x} = \mathbf{A}\boldsymbol{s}$	$\mathbf{A}$ is the unknown mixing matrix $\boldsymbol{s}$ unobserved ICs

Table 3.1: MGARCH and ICA comparison

One can make the following identification :

- $\boldsymbol{\varepsilon}_t$ is the observed data (e.g. financial returns as it will be the case in our next chapter) like the $\boldsymbol{x}$,

- $\boldsymbol{H}_t^{1/2}$ can be seen as the $\mathbf{A}$ matrix called *mixing matrix*,
- $\boldsymbol{Z}_t$ as the structural shocks to be identified as the ICs $\boldsymbol{s}$.

3.1.1 Simulation design

To design the simulation, we need to generate the residuals of our fitted MGARCH model ($m = 3$ for simplicity) and retrieve the $\boldsymbol{H}_t^{1/2}$ matrix which will be rotated according to different angle to obtain a new matrix $\tilde{\boldsymbol{H}}_t^{1/2}$ that aims the same function as stated in section 1.2.2. Inspired by the work of Hafner et al. (2022), we used the same rotation matrix design :

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\delta_1) & -\sin(\delta_1) \\ 0 & \sin(\delta_1) & \cos(\delta_1) \end{bmatrix} \times \begin{bmatrix} \cos(\delta_2) & 0 & -\sin(\delta_2) \\ 0 & 1 & 0 \\ \sin(\delta_2) & 0 & \cos(\delta_2) \end{bmatrix} \times \begin{bmatrix} \cos(\delta_3) & -\sin(\delta_3) & 0 \\ \sin(\delta_3) & \cos(\delta_3) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.1}$$

and also, we assume that the $\boldsymbol{H}_t^{1/2}$ matrix is known. Thus the residuals to generate will appear to be :

$$\tilde{\varepsilon}_t = \mathbf{R}_\delta \boldsymbol{Z}_t := \boldsymbol{H}_t^{-1/2} \varepsilon_t \tag{3.2}$$

where :

- $\boldsymbol{Z}_t$ are generated independently from a univariate student's t-distribution and then standardized. This way of generating $\boldsymbol{Z}_t$ will insure independence,
- $\varepsilon_t := \boldsymbol{H}_t^{1/2} \mathbf{R}_\delta \boldsymbol{Z}_t$ are the observed residuals.

By assuming in this simulation that the $\boldsymbol{H}_t^{1/2}$ is known, we thus generate $T + 100 = 1100, 2100, 4100$ data point using (3.2), and discard the first 100 draws afterwards to avoid initialization effects. The rotation angle vector will be $\delta = (.10, .25, .40)\pi$. This rotation angle aim to rotate the true $\boldsymbol{H}_t^{1/2}$ to obtain a new matrix (called the *mixing matrix* in ICA jargon). The possible $\tilde{\boldsymbol{H}}_t^{1/2}$ are :

- $\tilde{\boldsymbol{H}}_t^{1/2} = \boldsymbol{H}_t^{1/2} \boldsymbol{R}_\delta$, the clockwise rotation of $\boldsymbol{H}_t^{1/2}$, which is the true known mixing matrix,
- $\tilde{\boldsymbol{H}}_t^{1/2} = \boldsymbol{H}_t^{1/2} \mathbf{R}_{q\delta}$, with $q = 1.010, 1.025$ and 1.050. This insure slight clockwise rotation from the initial rotation,
- $\tilde{\boldsymbol{H}}_t^{1/2} = \boldsymbol{H}_t^{1/2} \mathbf{R}_{q^{-1}\delta}$. This insure 'insufficient' clockwise rotations.

For the simulation, we generate $\tilde{\varepsilon}_t := \mathbf{R}_\delta \boldsymbol{Z}_t$ for $m = 3$ assets and then apply ICA algorithm to retrieve on one hand, the estimated components which are the structural shocks and on other hand, the estimated unmixing matrix $\widehat{\tilde{\boldsymbol{H}}_t^{1/2}}$. Once the latest are obtained, we measure their similarities with the true and known $\tilde{\boldsymbol{H}}_t^{1/2}$ using the following metric introduced by

Amari, Cichocki, and Yang (1995) and used by Bach and Jordan (2002) to evaluate the performance of an ICA algorithm:

$$d(\mathbf{A},\mathbf{B}) = \frac{1}{2m}\sum_{i=1}^{m}\left(\frac{\sum_{j=1}^{m}|a_{ij}|}{\texttt{max}_j|a_{ij}|}-1\right)+\frac{1}{2m}\sum_{j=1}^{m}\left(\frac{\sum_{i=1}^{m}|a_{ij}|}{\texttt{max}_i|a_{ij}|}-1\right) \tag{3.3}$$

where $a_{ij} = (\mathbf{A}\mathbf{B}^{-1})_{ij}$ with $\mathbf{A} = \tilde{\boldsymbol{H}}_t^{1/2}$ and $\mathbf{B} = \widehat{\tilde{\boldsymbol{H}}_t^{1/2}}$. This metric, invariant to permutation and scaling which we refer to now on «Amari distance error », is used to choose between algorithm, since smaller the metric better is the algorithm in recovering the true unmixing matrix.

Also, for mapping the data in the features space in the case of kernelICA, we will use the Gaussian radial basis kernel as this insure the mapping to a high dimensional space where data can be linearly separable. Bach and Jordan (2002) recommend that the scale parameter σ be $1/2$ when $n > 1000$ and 1 if $n \leq 1000$. In our case here, since all the sample size are larger than 1000, we will use $1/2$ as the scale parameter.

Finally, we vary the degree of freedom of the student's t-distribution and replicate the simulation 500 times and measure the proportion of times the kernelICA performs better than the simple FastICA through amari distance error between the known mixing matrix $\boldsymbol{H}_t^{1/2}\mathbf{R}_\delta$ and the retrieved one.

3.2 Simulation's results analysis

3.2.1 Simulation purpose

To better understand the following, some aspect of our simulation need to be clearly enlightened. From the previous section, we know the mixing matrix which is the $\boldsymbol{H}_t^{1/2}\mathbf{R}_\delta$. In the ICA schema, choosing the right non-quadratic function is crucial, hence, we will explore the two candidates function and retrieve the one that perform better in recovering the components. Thus, two main comparisons will be made:

1. Classical ICA with `exp` function as non-quadratic function denoted (ICA_exp) and `logcosh` versus kernelICA-kgv and kernelICA-kcca. This comparison will provide us the optimal algorithm regarding our data.

2. kernelICA-kgv versus kernelICA-kcca to see which one of these two performed better in recovering the sources; the best one will be our favorite choice in the application on real data in next chapter.

These comparisons will be based on proportion of times one method overperformed its contrast one over the 500 simulations and on different setting according to a slight rotation q of the angle, the sample size n, the degree of freedom *df* of the distribution (supposed here to be student's t-distribution).

Out of these comparisons, what we are hoping is that the kernel methods will overperformed the simple ICA methods. Thus, the kernelICA will be used on real data to get the structural shocks.

3.2.2 Simulation results analysis

Here, we start by diving into the overall performances of our algorithms. By contrasting classical ICA and kernelICA, we have the following tables. For each value of given parameters (q, df, n), the tables present the number of times over 500 replications each considered ICA methods outperformed its fellow competitors.

3.2.2.1 Clockwise rotation

The tables (3.2 to 3.5) show the performance of the methods when a clockwise rotation in made according to a certain angle as defined in the simulation design.

	$\boldsymbol{Z}_t$	ICA		kernelICA	
df	sample size n	`exp`	`logcosh`	`kgv`	`kcca`
5	1000	28	30	224	218
	2000	26	29	229	216
	4000	13	16	239	232
10	1000	37	33	233	197
	2000	30	25	267	178
	4000	21	18	245	216
15	1000	73	64	207	156
	2000	48	42	223	187
	4000	21	24	263	192

Table 3.2: Performance of the methods according to amari distance error when q = 1

	$\boldsymbol{Z}_t$	ICA		kernelICA	
df	sample size n	`exp`	`logcosh`	`kgv`	`kcca`
5	1000	29	29	242	200
	2000	33	25	240	202
	4000	21	20	221	238
10	1000	46	42	222	190
	2000	18	27	245	210
	4000	17	15	263	203
15	1000	81	82	182	155
	2000	52	48	221	179
	4000	24	26	225	225

Table 3.3: Performance of the methods according to amari distance error when q = 1.01

	$\boldsymbol{Z}_t$	ICA		kernelICA	
df	sample size n	`exp`	`logcosh`	`kgv`	`kcca`
5	1000	24	27	255	194
	2000	24	30	231	215
	4000	27	17	226	230
10	1000	53	45	218	184
	2000	33	15	254	198
	4000	30	20	242	208
15	1000	71	67	185	177
	2000	49	33	242	176
	4000	32	20	254	194

Table 3.4: Performance of the methods according to amari distance error when q = 1.025

	$\boldsymbol{Z}_t$	ICA		kernelICA	
df	sample size n	`exp`	`logcosh`	`kgv`	`kcca`
5	1000	24	23	231	222
	2000	24	22	241	213
	4000	20	20	227	233
10	1000	43	38	220	199
	2000	21	19	267	193
	4000	18	21	246	215
15	1000	81	70	181	168
	2000	36	37	234	193
	4000	23	21	242	214

Table 3.5: Performance of the methods according to amari distance error when q = 1.05

As expected, the kernelICA algorithms (kgv and kcca) in more than 90% of the cases, return a smaller amari error when the rotation is done in clockwise angle. Therefore, in retrieving an unmixing matrix close to the true provided one, the kernelICA alogorithms are working quite well.

Let us check what is the output in case of anti-clockwise rotation or insufficient rotation.

3.2.2.2 Anti-clockwise rotation

Below tables (3.6 to 3.8) outlined the anti-clockwise rotation results. And also, kernelICA methods are far better than the classical ICA.

	$\boldsymbol{Z}_t$	ICA		kernelICA	
df	sample size n	`exp`	`logcosh`	`kgv`	`kcca`
5	1000	29	35	229	207
	2000	32	24	237	207
	4000	18	20	220	242
10	1000	54	42	219	185
	2000	23	33	255	189
	4000	15	14	265	206
15	1000	81	80	184	155
	2000	58	39	228	175
	4000	24	29	220	227

Table 3.6: Performance of the methods according to amari distance error when q = 0.99

	$\boldsymbol{Z}_t$	ICA		kernelICA	
df	sample size n	`exp`	`logcosh`	`kgv`	`kcca`
5	1000	26	32	246	196
	2000	21	30	232	217
	4000	19	12	220	249
10	1000	65	45	208	182
	2000	35	15	255	195
	4000	28	17	240	215
15	1000	69	81	188	162
	2000	38	38	232	192
	4000	31	20	243	206

Table 3.7: Performance of the methods according to amari distance error when q = 0.9756

	$\boldsymbol{Z}_t$	ICA		kernelICA	
df	sample size n	`exp`	`logcosh`	`kgv`	`kcca`
5	1000	32	32	222	214
	2000	28	28	240	204
	4000	17	23	233	227
10	1000	42	53	216	189
	2000	28	25	251	196
	4000	23	15	244	218
15	1000	76	83	168	173
	2000	41	52	218	189
	4000	25	27	238	210

Table 3.8: Performance of the methods according to amari distance error when q = 0.9523

3.2.2.3 Comparison between kernelICA-kgv and kernelICA-kcca performance

Out of these tables, although kernelICA algorithm's results are tight, kgv algorithm works better than kcca (as shown by figure 3.1 and 3.2). This suggest the use of kgv algorithm in retrieving structural shocks in our next chapter where we will be working on real data.

Indeed, taking a global picture shows that kernelICA methods are way better than simple ICA, let us now take a closer look of the situation according to our parameters : angle rotation q, student's t distribution's degree of freedom df and sample size n.

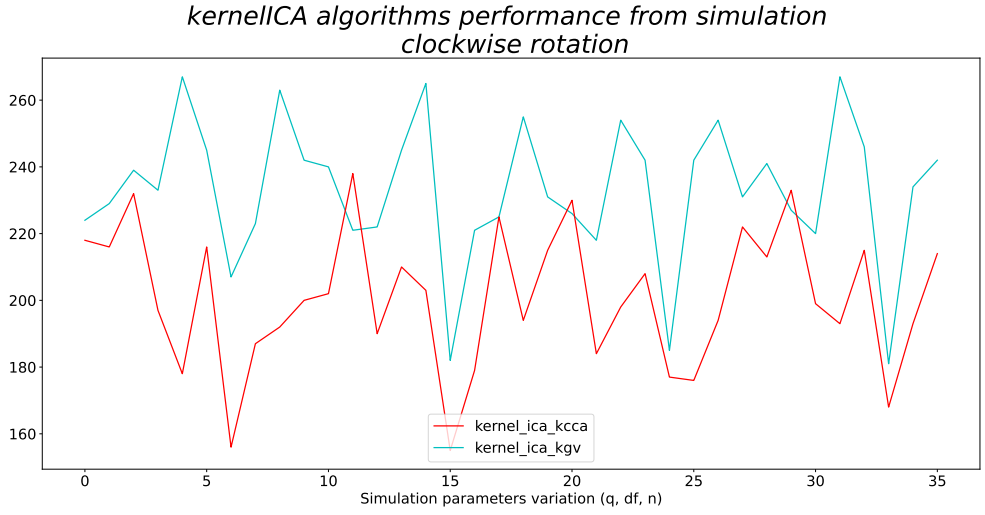


Figure 3.1: KernelICA algorithms performance in clockwise rotation scenario.

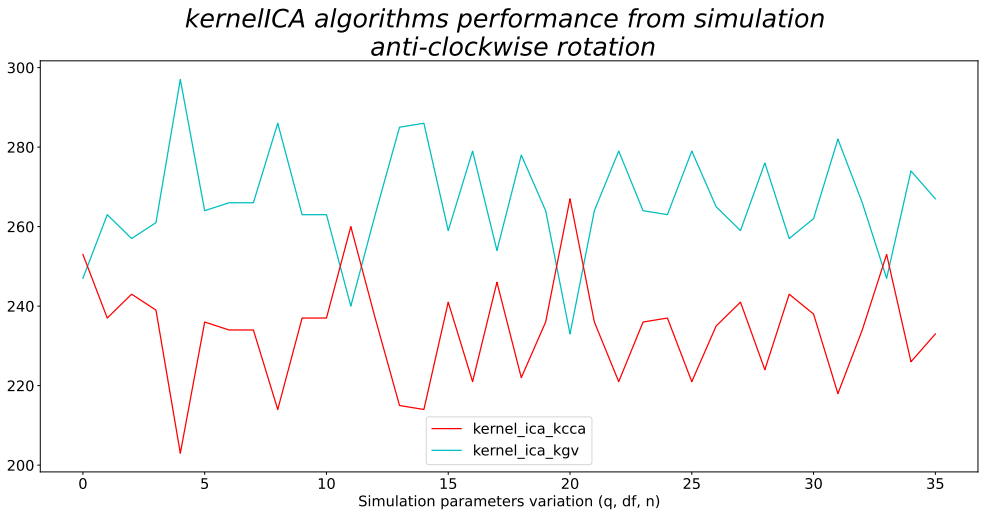


Figure 3.2: KernelICA algorithms performance in anti-clockwise rotation scenario.

3.2.2.4 Angle rotation influences

What we can see from these figures (3.3 and 3.4) is that over 500 replications, each ICA algorithm's mean and median are still tight together for the four values of q. Inducing a first idea that the rotation angle seems to not affect that much the estimations of the methods.

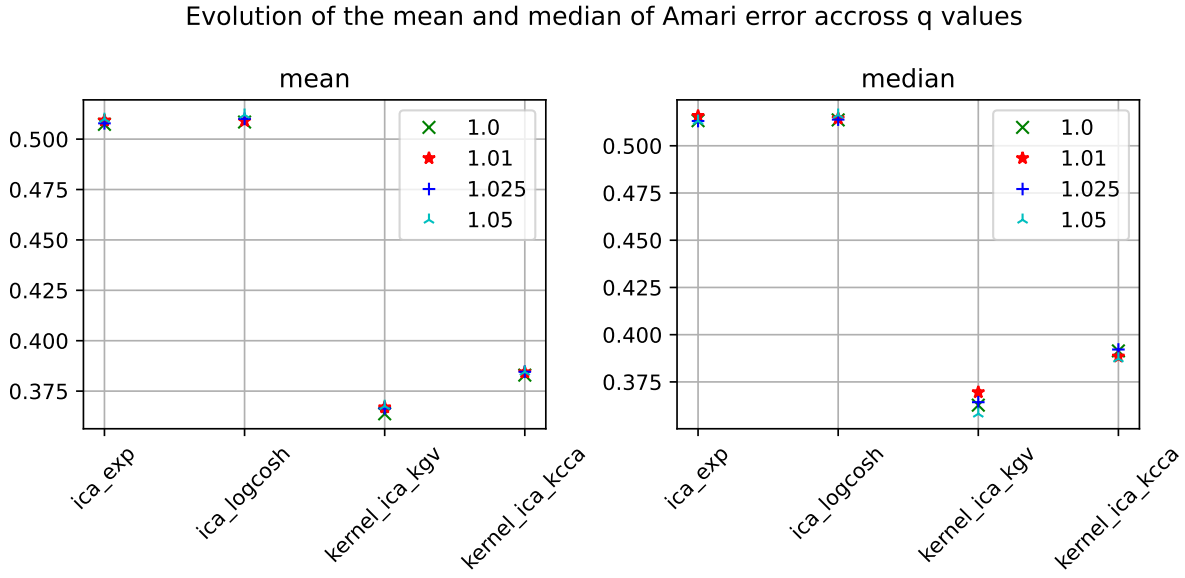


Figure 3.3: Amari error's mean and median for the angle q values (clockwise case)

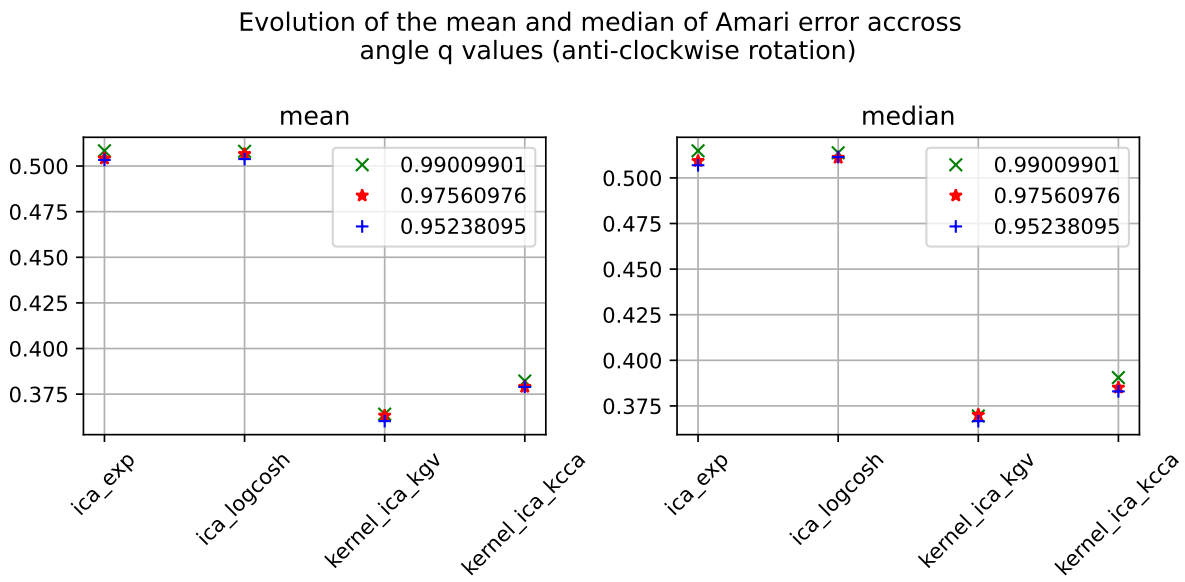


Figure 3.4: Amari error's mean and median for the angle q values (anti-clockwise case)

3.2.2.5 Student's t-distribution's degree of freedom influences

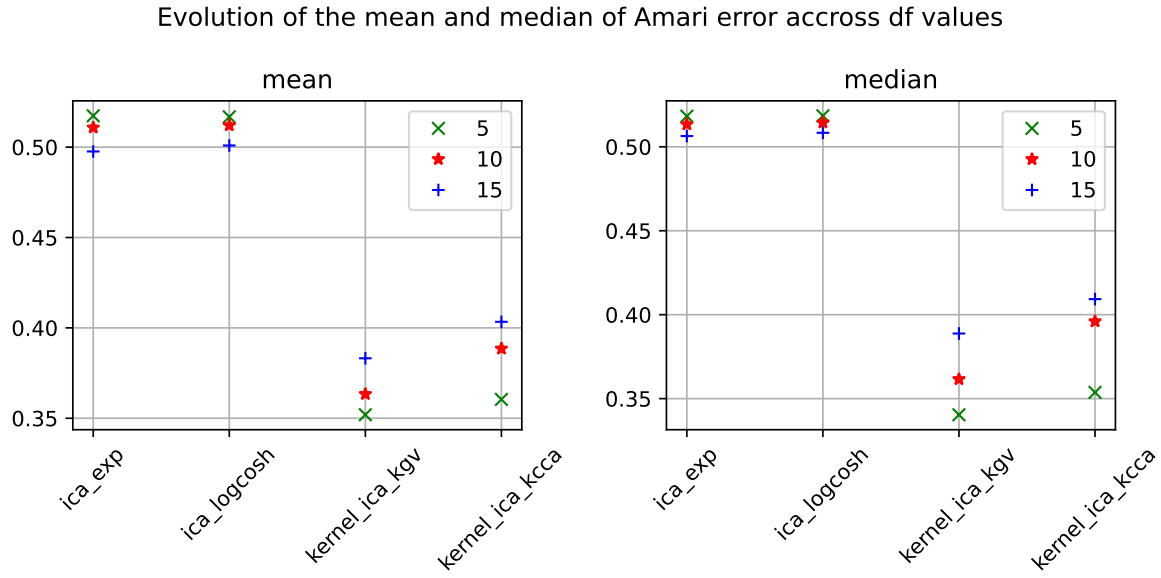


Figure 3.5: Amari error's mean and median for the student's t-distribution's degree of freedom df values (clockwise case)

What we can see from these figures (3.5 and 3.6) is over 500 replications, each ICA algorithm's mean and median are not still tight together for the three values of df, especially

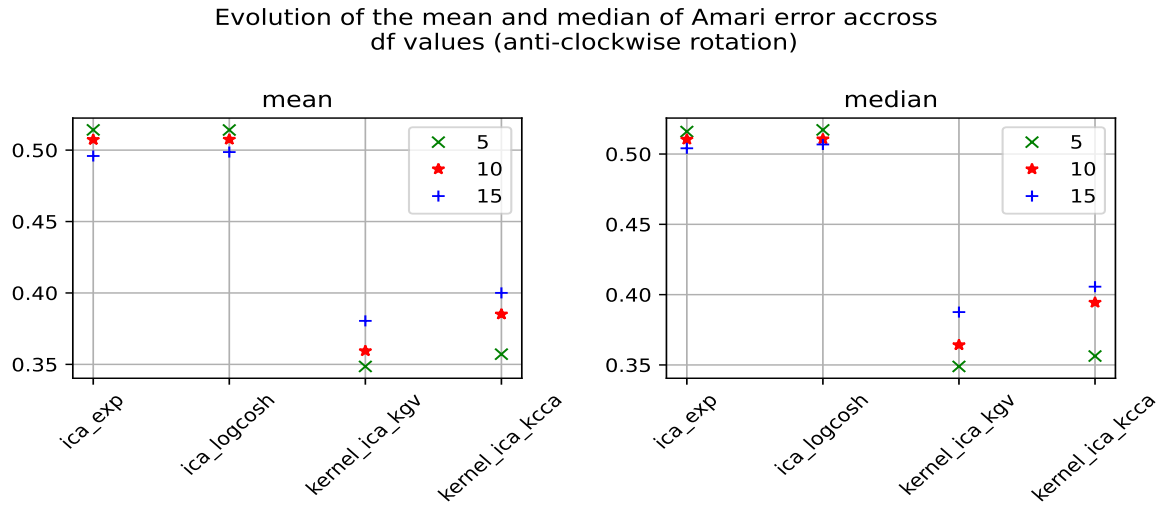


Figure 3.6: Amari error's mean and median for the student's t-distribution's degree of freedom df values (anti-clockwise case)

the kernelICA algorithms. It is observed that, when the degree of freedom increase, the kernelICA algorithms tends to estimate unmixing matrix that are becoming far from the true one. Nevertheless, this discrepancy is not that much, suggesting that the kernelICA algorithms still do a good job in recovering the sources.

Let us not forget that, by increasing the degree of freedom, we are likely to be close to the Gaussian distribution; where the condition of no-Gaussianity of the distribution will not be guaranteed, inducing the behaviour observed in the figures 3.5 and 3.6.

3.2.2.6 Sample size influences

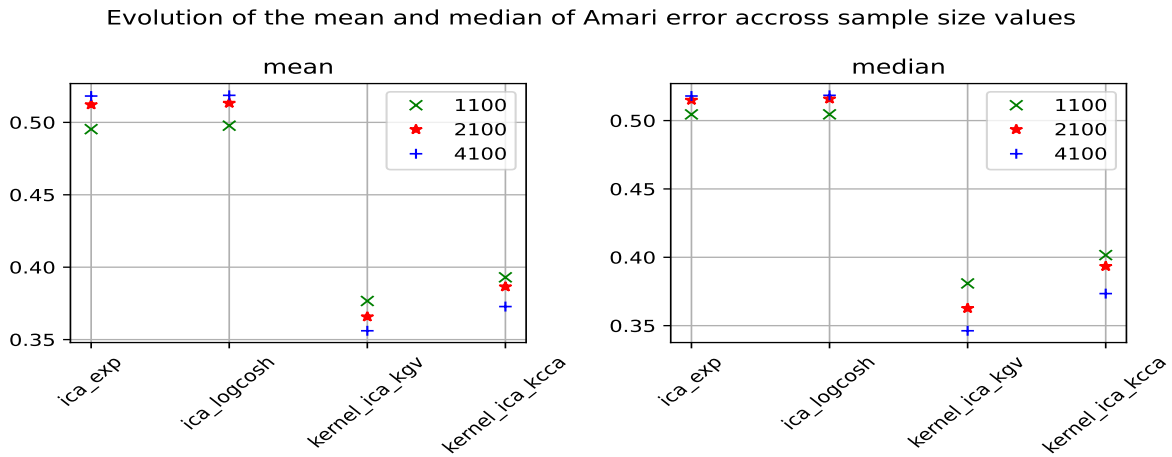


Figure 3.7: Amari error's mean and median for the sample sizes values (clockwise case)

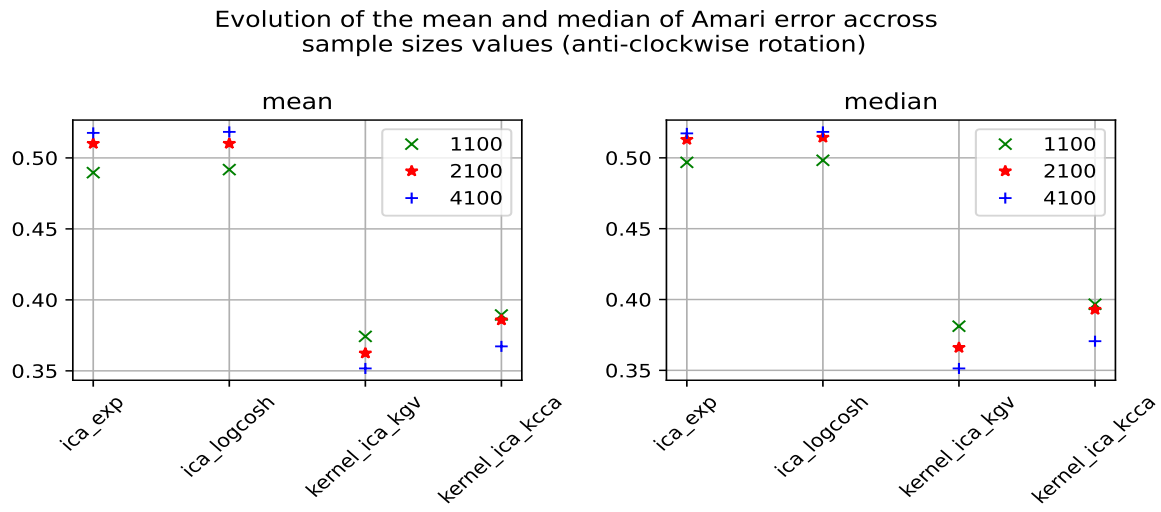


Figure 3.8: Amari error's mean and median for the sample sizes values (anti-clockwise case)

From figure 3.7 and 3.8 below, as in the case of the rotation angle cases, the sample sizes seem to not affect the results of our considered algorithms. Thus, the mean and median are still closed one to another over the 500 replications.

3.3 Conclusion

In this chapter, we have settled our simulation framework and find out that the kernelICA could be used in unveiling the ICs from a given sources as outlined before. However, the distribution of the sources may play a crucial role, although these kernelICA algorithm are working fine. Also, from the performance of the kgv algorithm, one can expect kernelICA to be useful in unveiling the structural shocks through the unmixing matrix estimated from the residuals of a given MGARCH model. This is the scope of our next chapter, where we will work with real data.

Chapter 4

Application on real data

This last chapter presents a ause case of kernelICA on real data. Hence, it will treat of MGARCH model specification, its parameters estimations, retrieving the residuals. Out of the residuals, we will retrieve the structural shocks by kernelICA and study the volatility impulse response function of a given shock.

4.1 The data

While a lot of market indexes exist to collectively track the developments of a sector, there are a handful of companies that lead the pack in terms of sector influence. FAANG, GAFAM, and BATX are acronyms for influential companies that have a large impact on their sectors, and the market as a whole. Studying these indexes and their structural dynamic appears like a perfect case where applying ICA methods will be helpful. Our interest is set upon the high-tech sectors especially the GAFAM indexes.

GAFAM is made up of Alphabet Inc.(GOOG), Apple Inc. (AAPL), Meta (META) (formerly Facebook), Amazon Inc. (AMZN), and Microsoft Corp. (MSFT). All five companies are listed on the NASDAQ stock exchange. Studying GAFAM stock exchange appears to be a great opportunity to unveil the structural shocks underlie these indexes on the market even though the dynamic of such markets is complex and many factors play key role.

The time horizon we are interested about is the past eight years. Hence, we collect the daily adjusted closing prices from october 31st 2015 to october 31st 2023, thus 2012 days are of interest. The use of the *adjusted closing prices* is due to the fact that they account for possible corporate actions, such as stock splits (Lewinson, 2022).

4.1.1 Exploration of the GAFAM indexes

The figure 4.1 and A.1 in Appendix A shows the evolution of the stocks and their log-returns during the time horizon considered.

At first sight, it is difficult to visually identify any trend or seasonality from the data. However, we can see that the series are not stationary. This is also confirmed by the

Figure 4.1: GAFAM's stock values from 2015-10-31 to 2023-10-31

Dicker fuller test applied on each of them (see table A.1 in Appendix A for full results). Moreover, one can suspect a causality effect between the series around middle 2018, end first quarter 2020 and 2022, since a change in the trend is observed for all the series. This will be studied through the granger causality test in the next subsection. Also, without any other assumption, the drop around middle 2020 may be caused by the covid-19 pandemic, however this needs to be investigated.

By looking at the log-returns plots (see Appendix A.1), we observe a stationary trend and moreover, one can suspect a volatility clustering phenomenon and leptokurtic distribution, characteristic of financial assets. Table 4.1 below summarized the principal statistics of the log-returns, and the last two columns present the Jarque-Bera normality test results.

As usually the case in financial assets analysis, the log-returns are stationary. Due to the interesting properties of the log-returns in financial assets analysis (Lewinson, 2022), we will study from now on the log-returns instead of the observed data. For further checking, Appendix A.2 presents a descriptive statistics of the log-returns and a causality test study on the observed data.

	min	med.	avg	std	max	skew.	kurt.	test.stat	test.pval.
META	-0.306	0.001	0.001	0.025	0.209	-1.523	30.599	64603.044	0.0
AAPL	-0.138	0.001	0.001	0.019	0.113	-0.228	8.599	2643.809	0.0
GOOG	-0.118	0.001	0.001	0.018	0.099	-0.279	7.714	1887.884	0.0
MSFT	-0.159	0.001	0.001	0.018	0.133	-0.227	10.678	4957.118	0.0
AMZN	-0.151	0.001	0.001	0.021	0.127	-0.049	7.945	2049.771	0.0

Table 4.1: Summary statistics and normality test results of log-returns data

4.2 Fitted model

Here, we will fit a BEKK(1,1) to the log-returns data to retrieve the standardized residuals. The BEKKs package developed by Fülle, Lange, Hafner, and Herwartz (2023) provides us suitable function to estimate this model. This package provides two ways of estimating the model : the diagonal BEKK and the scalar BEKK model of Ding and Engle (2001). The latter allows faster but less flexible estimation in higher dimensions. We thus fit these two models and select the best one according to AIC criteria. The results[1] of the fitted models are presented in the table 4.2 below .

	BEKK(1,1)	
	diag.	scale
Log-Likelihood	28645.97	28577.77
AIC	-57272.95	-57040.53
BIC	-57167.85	-57016.44

Table 4.2: BEKK's AIC and BIC

Based on the results, the diagonal BEKK seems to provided suitable estimation. We then retrieved the standardized residuals $\boldsymbol{\varepsilon}_t = \hat{\boldsymbol{H}}_t^{-1/2}\boldsymbol{y}_t$ which qq-plots are provided in Figure A.2 of Appendix A.3. Needless to add that these residuals are whiten, however the mean vector and covariance matrix is provided in Appendix A.3.

From the Figure A.2 of Appendix A.3 coupled with the Table 4.3, none of the standardized residuals has a normal distribution. This aligns with our assumption in the simulation study here we work with data from student's t-distribution.

These residuals are now passed as observed series to the kernelICA algorithms to unveil the structural shocks and the structural impact multiplier matrix.

[1]Check Table B.1 in Appendix B.1 for estimated matrices

	min	med.	avg	std	max	skew.	kurt.	test.stat	test.pval.
META	-14.834	0.039	0.022	1.046	6.646	-2.361	36.486	95826.368	0.0
AAPL	-5.296	0.054	0.052	1.02	7.653	0.351	7.69	1884.801	0.0
GOOG	-9.083	0.009	0.014	1.023	9.514	0.31	18.433	19989.441	0.0
MSFT	-5.458	0.011	0.049	0.995	8.866	0.75	10.959	5495.596	0.0
AMZN	-8.807	-0.042	-0.006	1.016	9.157	0.178	18.45	20012.246	0.0

Table 4.3: Summary statistics and normality test results of the standardized residuals

4.3 Structural shocks identification

We have estimated the models and selected the best that fits to our data, now let us retrieve the residuals and apply the kernelICA on them to uncover the latent components and the unmixing matrices. Indeed, our interest is the kernelICA method, still, we will expand the analysis to both classical and non linear ICA.

The structural impact multiplier matrices obtained in both classical and kernel ICA are summarized in the table below.

FastICA										
		exp.					logcosh			
$\begin{pmatrix} -0.262 & -0.101 & 0.941 & -0.003 & 0.189 \\ 0.559 & 0.522 & 0.234 & -0.588 & -0.119 \\ -0.155 & 0.206 & 0.163 & 0.284 & -0.909 \\ 0.760 & -0.244 & 0.182 & 0.573 & 0.026 \\ -0.128 & 0.784 & -0.021 & 0.495 & 0.351 \end{pmatrix}$					$\begin{pmatrix} -0.050 & 0.096 & 0.259 & 0.199 & -0.939 \\ 0.686 & -0.572 & -0.350 & -0.166 & -0.227 \\ -0.350 & -0.190 & 0.145 & -0.893 & -0.150 \\ -0.345 & 0.218 & -0.889 & -0.021 & -0.209 \\ -0.534 & -0.761 & 0.005 & 0.367 & 0.030 \end{pmatrix}$				
kernelICA										
		kgv					kcca			
$\begin{pmatrix} 0.093 & -0.090 & 0.120 & 0.096 & 0.969 \\ 0.273 & -0.150 & 0.875 & -0.335 & -0.115 \\ -0.199 & 0.144 & -0.253 & -0.933 & 0.129 \\ 0.396 & 0.900 & 0.024 & 0.020 & 0.021 \\ 0.798 & -0.323 & -0.339 & -0.142 & -0.033 \end{pmatrix}$					$\begin{pmatrix} -0.687 & -0.198 & -0.278 & -0.614 & 0.006 \\ 0.458 & -0.694 & -0.161 & -0.241 & -0.350 \\ -0.338 & 0.081 & -0.162 & 0.418 & -0.807 \\ 0.073 & 0.079 & -0.890 & 0.283 & 0.289 \\ -0.340 & -0.658 & 0.199 & 0.568 & 0.336 \end{pmatrix}$				

Table 4.4: Structural impact multiplier matrices obtained from ICA algorithms

The components obtained from the kernelICA-kgv algorithm are shown on the Figure A.3 of Appendix A.4.1. Need to recall that the so identified components can not be assigned to a given market index since ICA algorithms suffer scale and order indeterminancy.

In sum, the log-returns at a given date t can be seen as :

$$\boldsymbol{y}_t = \hat{\boldsymbol{H}}_t^{1/2} \mathbf{R}_\delta \boldsymbol{Z}_t \tag{4.1}$$

where :

- $\boldsymbol{y}_t$ is the log-returns,
- $\hat{\boldsymbol{H}}_t^{1/2}$ is the square root of the conditional covariance matrix of the $\boldsymbol{y}_t$. In the case of the BEKK package used, the $\hat{\boldsymbol{H}}_t^{1/2}$ matrix is obtained by eigenvalues decomposition,
- $\mathbf{R}_\delta$ is the structural impact multiplier matrix obtained by kernelICA-kgv algorithms,
- and $\boldsymbol{Z}_t$ are the structural shocks. Indeed, unobserved but still govern the different movement observed in the GAFAM's indexes values from 2015 to 2023.

The distribution of these shocks are presented in the Figure A.4 of Appendix A.4, but the summarized information are provided in the table 4.5 below.

	min	med.	avg	std	max	skew.	kurt.	test.stat.	test.pval
comp.1	-8.362	-0.039	0.0	1.012	9.151	0.273	18.991	21451.371	0.0
comp.2	-9.744	-0.006	0.0	1.006	9.579	-0.311	18.286	19612.458	0.0
comp.3	-6.243	0.018	0.0	0.995	6.254	-0.269	8.14	2238.202	0.0
comp.4	-5.162	0.005	0.0	1.017	5.222	-0.076	5.41	488.574	0.0
comp.5	-17.555	-0.001	0.0	1.068	7.872	-4.12	71.834	402708.197	0.0

Table 4.5: Summary statistics and normality test results of retrieved components

4.4 Volatility impulse response analysis

As one of the most important applications of MGARCH models (Bauwens et al., 2006), the analysis of responses of volatility to shocks remains a way to study the behavior of volatility. So, this subsection tackles the dynamic impact of shocks on volatility using our structural shocks provided by kernelICA-kgv.

Conversely to generalized impulse response function (GIRF), VIRF does not trace the dynamic effects of a shock on the conditional mean but on the conditional covariance matrix. Thus, Hafner and Herwartz (2006) define VIRF as the difference between the expected h-step-ahead covariance conditioning on the shock $\mathbf{Z}_t$ and past information $\mathcal{F}_{t-1}$ and the expected h-step-ahead covariance conditioning on past information $\mathcal{F}_{t-1}$ only, which means :

$$\mathrm{V}_{t+h}(\mathbf{Z}_t) = \mathbb{E}[\mathrm{vech}(\boldsymbol{H}_{t+h})|\mathbf{Z}_t, \mathcal{F}_{t-1}] - \mathbb{E}[\mathrm{vech}(\boldsymbol{H}_{t+h})|\mathcal{F}_{t-1}] \tag{4.2}$$

Fengler and Polivka (2022) derive this for a BEKK(1,1) model and get :

$$\mathrm{V}_{t+h}(\mathbf{Z}_t, \eta) = (\tilde{\mathbf{A}}_1 + \tilde{\mathbf{B}}_1)^{h-1}\tilde{\mathbf{A}}_1\mathbf{D}_n^+ \left(\boldsymbol{H}_t^{1/2} \otimes \boldsymbol{H}_t^{1/2}\right)\mathbf{D}_n\mathrm{vech}(\mathbf{R}_\delta\mathbf{Z}_t\mathbf{Z}_t^\top\mathbf{R}_\delta^\top - \mathbf{I}_n) \; (h \geq 1) \tag{4.3}$$

$$\implies \mathrm{V}_{t+h}(\mathbf{Z}_t, \eta) = (\tilde{\mathbf{A}}_1 + \tilde{\mathbf{B}}_1)\mathrm{V}_{t+h-1}(\mathbf{Z}_t) \quad (h \geq 2) \tag{4.4}$$

where

- $\mathrm{V}_{t+h}(\mathbf{Z}_t, \eta)$ is the VIRF given an initial shock $\mathbf{Z}_t$ and BEKK model parameters $\eta = \{\mathbf{A}, \mathbf{B}, \ldots\}$,
- $\mathbf{D}_n$ is the duplicate matrix and $\mathbf{D}_n^+$ its Moore-Penrose inverse,
- $\tilde{\mathbf{A}}_1 = \mathbf{D}_n^+(\mathbf{A}_1 \otimes \mathbf{A}_1)^\top\mathbf{D}_n$ with $\mathbf{A}_1$ the $\mathbf{A}$ matrix in the BEKK specification of $\boldsymbol{H}_t$,
- $\tilde{\mathbf{B}}_1 = \mathbf{D}_n^+(\mathbf{B}_1 \otimes B_1)^\top\mathbf{D}_n$ with $\mathbf{B}_1$ the $\mathbf{B}$ matrix in the BEKK specification of $\boldsymbol{H}_t$,
- $\mathbf{R}_\delta$ and $\mathbf{Z}_t$, the rotation matrix and the given shocks respectively,
- and $\otimes$ the Kronecker product

Based on the formulas above (4.3 and 4.4), we can compute the VIRF of our new technique and compare it to the result of the classical VIRF (where $\mathbf{R}_\delta = \mathbf{I}_n$) and the responses obtained for the 4 others ICA algorithms (exploiting of the structural impact multiplier matrix provided by each of them).

Therefore, we consider 4 different scenarii : a given shock of 6 points on each stock index, *ceteris paribus*, on 2023-10-13 and study the responses of the (co-)variances of the indexes to these shocks on $h = 60$ days horizon.

Looking at the trends of the estimated VIRF, although the magnitude of the effects are not the same, we notice that :

- considering the structural shock on META (Figure 4.2), this shock tends to induce a strong, persistent and slow decay volatility on AAPL, MSFT and GOOG indexes (co)-variances, when we consider the structural impact multiplier matrix provided by the kcca algorithm or the simple case where the structural impact multiplier matrix is supposed to be identity. Conversely, for FastICA and kgv structural matrices, the impact seems not that strong, and tends to vanish 30 to 40 days after. So, a clear difference is noticed between the the identity matrix case and the case of structural impact multiplier matrices provides by ICA algorithms.

- for the shock on AMZN, as expected, the volabitility induced on META is also slow decaying. When considering the structural matrices, two group can be set : group of high volatility with matrices provided by kgv and FastICA-logcosh algorithms and the group of low volatility gathering the 3 others algorithms. Hence, as expected from the simulation, the kgv structural matrix exhibit a different from the identity matrix.

Something that may be subject to analysis, is the fact that in all five shocks configuration, the VIRF on the variance of the AMZN indexes tend to be persistent for some given structural matrices. Studying deeper the dynamic under this high tech market may suggest some hint of explanation.

Indeed, the impacts of the shocks we considered, are function of the structural multiplier matrices and the indexes considered. Even though the trend are seemingly close, it is not quiet evident to draw an overall conclusion on the behavior of the VIRF based on the structural matrices but one thing to consider is that in all of the cases, the kgv algorithm response is quiet different from the identity matrix response. This showcases the particularity of the kgv algorithms in structural shocks identification as observed in the simulation study above.

In order to validate the previous observation, we consider this time a negative shock. Recall that by definition, the impulse response is an even function of the shock, as opposed to an odd function in the traditional analysis meaning $\mathrm{V}_{t+h}(\mathbf{Z}_t) = \mathrm{V}_{t+h}(-\mathbf{Z}_t)$. But here, we consider the same shock as previously but with negative sign. The results are displayed in Appendix B.2. In general, the difference between the structural matrices is observed and the responses are shifted and mixed; showcasing the impact of the structural impact matrices on the volatility. However, the clear separation between kgv and the identity matrices responses is still observed as previously.

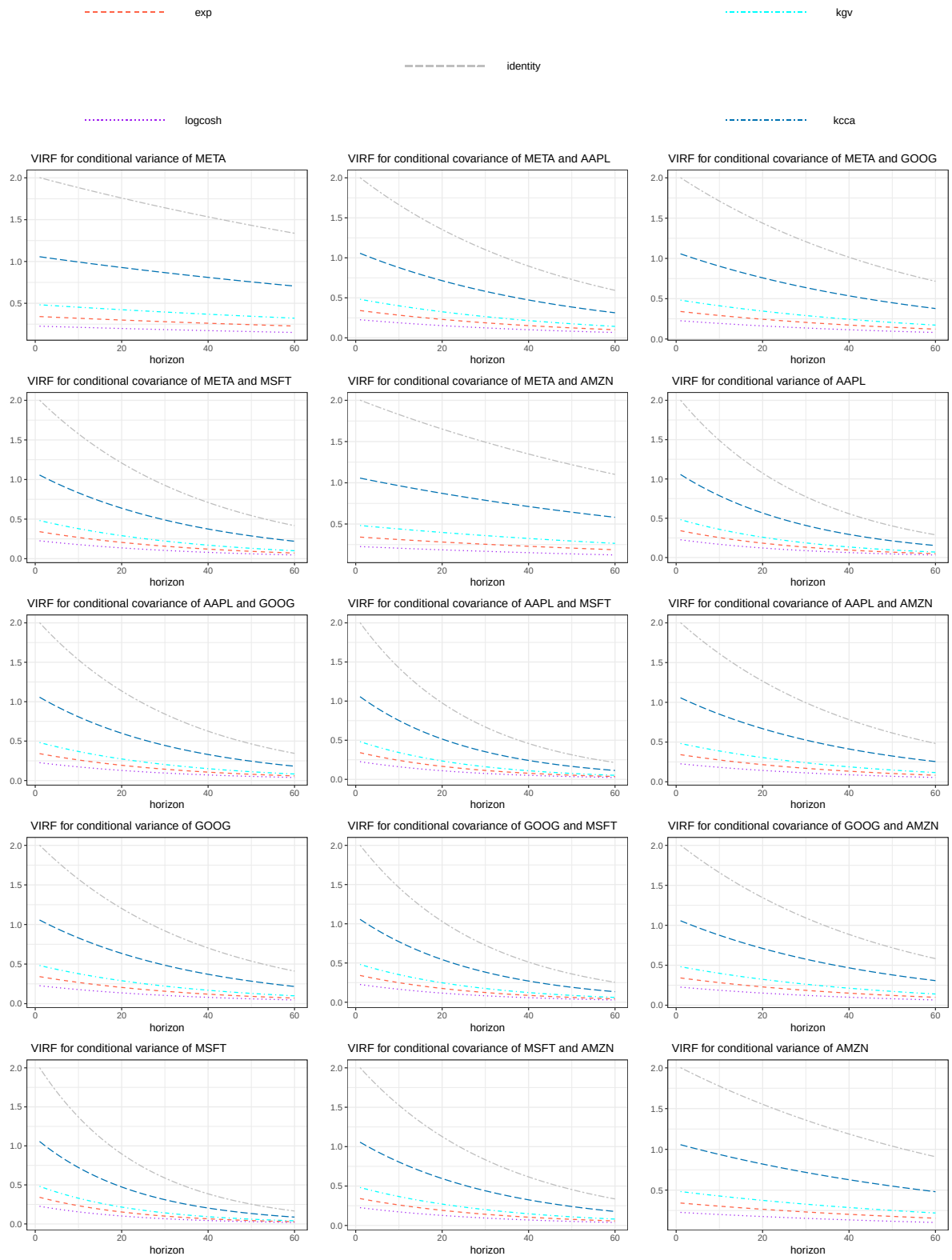


Figure 4.2: Plots of the estimated VIRF of 6 points shock on 2023-10-13 on META

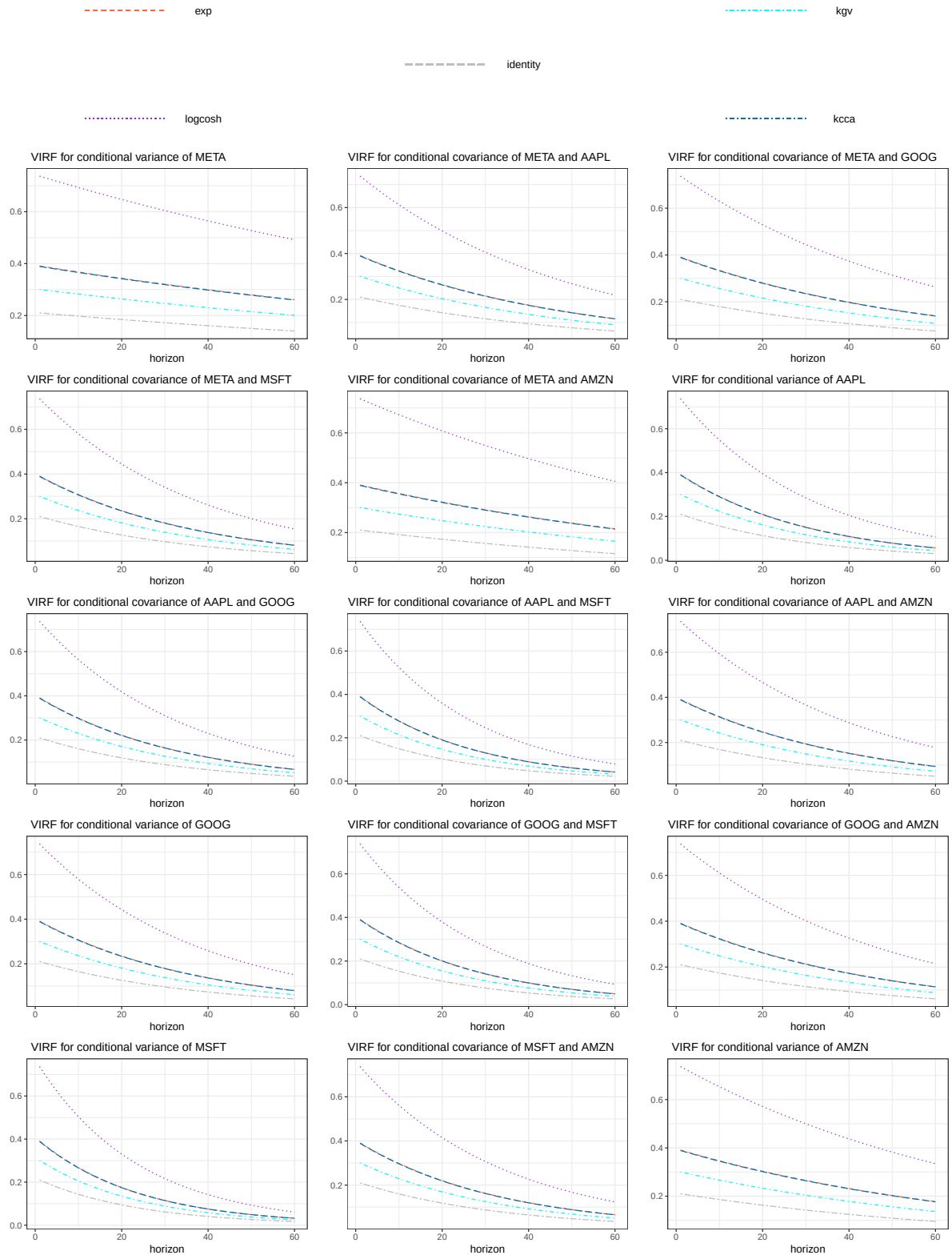


Figure 4.3: Plots of the estimated VIRF of 6 points shock on 2023-10-13 on APPL

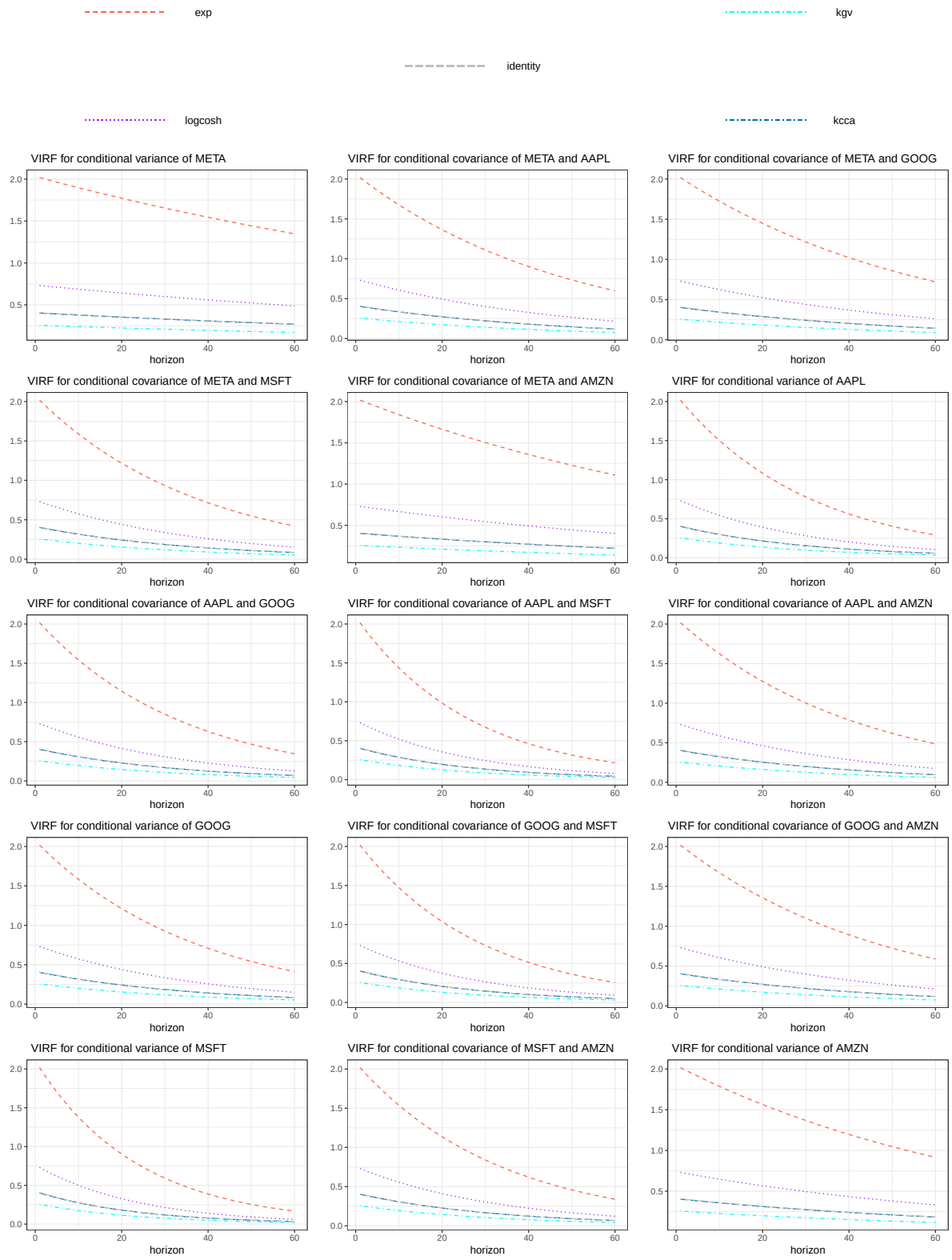


Figure 4.4: Plots of the estimated VIRF of 6 points shock on 2023-10-13 on GOOG

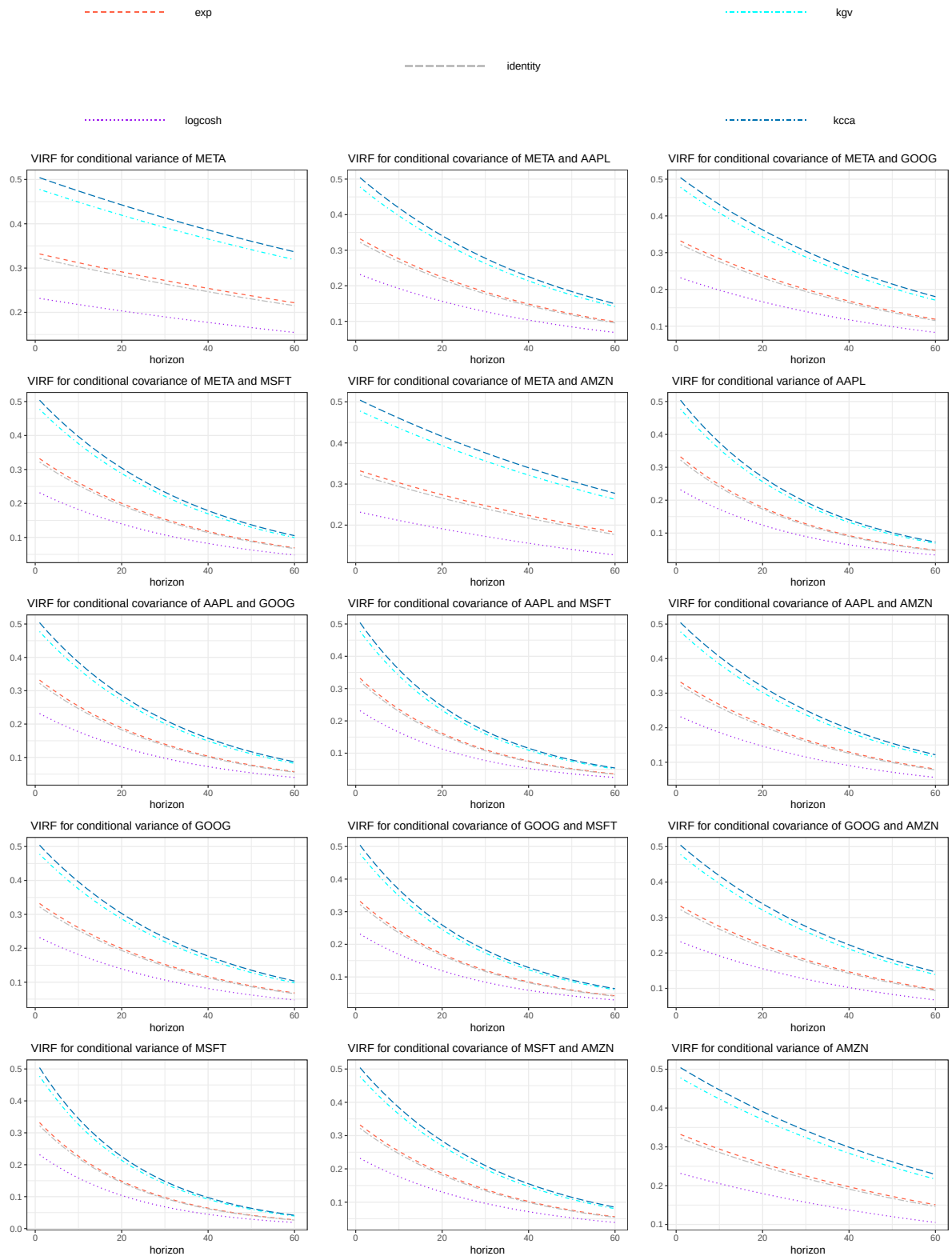


Figure 4.5: Plots of the estimated VIRF of 6 points shock on 2023-10-13 on MSFT

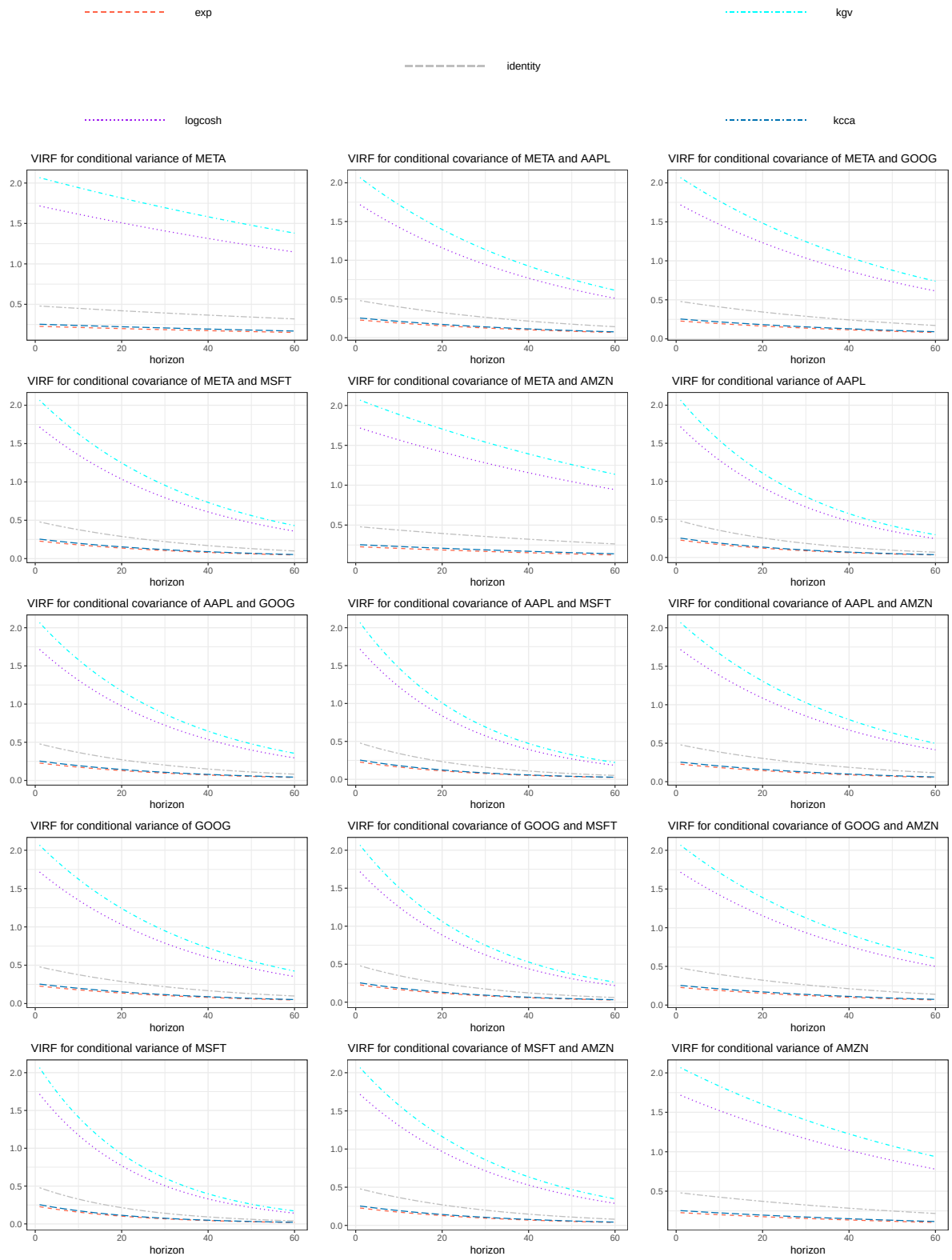


Figure 4.6: Plots of the estimated VIRF of 6 points shock on 2023-10-13 on AMZN

4.5 Conclusion

Throughout this application on GAFAM indexes, we have explored how kernelICA can be used in unveiling the structural shocks and the structural multiplier matrix. Also, we looked at the volatility impulse responses induced by a shock through the structural multiplier matrices provided by the ICA algorithms we have explored.

Although it is not done here, a good thing would be to check the case of an asymmetric shocks on VIRF. In addition, it may be interesting to study the significance of the VIRF induced by our shocks using the formula provided by Fengler and Polivka (2022) about the distribution of the estimated VIRF. Finally, as said, the dynamic under such market is quiet complex and a deeper analysis can unveil some important features.

Conclusion & Discussion

In this thesis, we look at the potential benefits of using nonlinear independent components analysis developed by Bach and Jordan (2002) in structural shocks identification. Firstly, we present what is structural shocks and methods available in the literature to recover this shocks. Then, we present independent component analysis and the kernelICA of Bach and Jordan and how it could be used in unveiling the structural shocks. Once the literature has been set up, we dive into a simulation study where we study the performance of ICA algorithms in recovering structural shocks, out of what the kgv algorithms is the high performer.

Finally, we apply the kernelICA methods to a real financial data from yahoo finance: the GAFAM stock indexes. The structural shocks and multiplier matrices obtained from the different algorithms have been studied. The latter is essential in capturing the influence of each of the structural shocks on each of the model variables and also on the VIRF of historical or empirical shocks. Out of this analysis, we have fought that, the kgv algorithm as suggested by the simulation, provided a suitable structural impulse multiplier and its impacts on the VIRF is different from the identity matrix impact. Hence, kernelICA would be a good approach in retrieving the structural dynamic of financial assets based on the data and not on any theoretical knowledge. Nevertheless, a combination of empirical knowledge and data driven identification should be used to not only identify the structure of the shocks but also have a suitable economical interpretation to these shocks and their impact of the assets.

Although this first analysis provided a hint about the usefulness of kernelICA in structural shocks analysis, the statistical significance of the retrieved structural impact multiplier matrices have not been analyzed here. Fengler and Polivka (2022) and Hafner and Herwartz (2006) proposed ways to analyse this.

Bibliography

Amari, S.-i., Cichocki, A., & Yang, H. (1995). A new learning algorithm for blind signal separation. *Advances in neural information processing systems*, *8*.

Bach, F. R., & Jordan, M. I. (2002). Kernel independent component analysis. *Journal of machine learning research*, *3*(Jul), 1–48.

Back, A. D., & Weigend, A. S. (1997). A first application of independent component analysis to extracting structure from stock returns. *International journal of neural systems*, *8*(04), 473–484.

Bauwens, L., Laurent, S., & Rombouts, J. V. (2006). Multivariate garch models: a survey. *Journal of Applied Econometrics*, *21*(1), 79–109.

Blanchard, O. J., & Quah, D. (1988). *The dynamic effects of aggregate demand and supply disturbances.* National Bureau of Economic Research Cambridge, Mass., USA.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307-327. Retrieved from `https://www.sciencedirect.com/science/article/pii/0304407686900631`

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *The review of economics and statistics*, 498–505.

Bollerslev, T., Engle, R. F., & Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of political Economy*, *96*(1), 116–131.

Canova, F., & De Nicolo, G. (2002). Monetary disturbances matter for business fluctuations in the g-7. *Journal of Monetary Economics*, *49*(6), 1131–1159.

Canova, F., & Ferroni, F. (2022). Mind the gap! stylized dynamic facts and structural models. *American Economic Journal: Macroeconomics*, *14*(4), 104–135.

Cardoso, J.-F. (1998). Blind signal separation: statistical principles. *Proceedings of the IEEE*, *86*(10), 2009–2025.

Cardoso, J.-F., & Souloumiac, A. (1993). Blind beamforming for non-gaussian signals. In *Iee proceedings f (radar and signal processing)* (Vol. 140, pp. 362–370).

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, *25*(5), 975–979.

Christiano, L. J., Eichenbaum, M., & Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? *Handbook of macroeconomics*, *1*, 65–148.

Cooley, T. F., & LeRoy, S. F. (1985). Atheoretical macroeconometrics: a critique. *Journal of Monetary Economics*, *16*(3), 283–308.

Ding, Z., & Engle, R. F. (2001). Large scale conditional covariance matrix modeling, estimation and testing.

Bibliography

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, *50*(4), 987–1007. Retrieved 2023-07-12, from `http://www.jstor.org/stable/1912773`

Engle, R. F., & Kroner, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric theory*, *11*(1), 122–150.

Faust, J. (1998). The robustness of identified var conclusions about money. In *Carnegie-rochester conference series on public policy* (Vol. 49, pp. 207–244).

Faust, J., & Leeper, E. M. (1997). When do long-run identifying restrictions give reliable results? *Journal of Business & Economic Statistics*, *15*(3), 345–353.

Fengler, M., & Polivka, J. (2022). *Structural volatility impulse response analysis*. School of Economics and Political Science, Department of Economics ....

Fry, R., & Pagan, A. (2011). Sign restrictions in structural vector autoregressions: A critical review. *Journal of Economic Literature*, *49*(4), 938–960.

Fülle, M. J., Lange, A., Hafner, C. M., & Herwartz, H. (2023). Bekks: Multivariate conditional volatility modelling and forecasting [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=BEKKs` (R package version 1.4.3)

Hafner, C. M., & Herwartz, H. (2006). Volatility impulse responses for multivariate garch models: An exchange rate illustration. *Journal of International Money and Finance*, *25*(5), 719–740.

Hafner, C. M., Herwartz, H., & Maxand, S. (2022). Identification of structural multivariate garch models. *Journal of Econometrics*, *227*(1), 212–227.

Harpaz, R. (2007). Tr-2007007: Independent component analysis: An introduction.

Hau, H., & Rey, H. (2004). Can portfolio rebalancing explain the dynamics of equity returns, equity flows, and exchange rates? *American Economic Review*, *94*(2), 126–133.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE. Transactions on Neural Networks*, *10*(3), 626–634.

Hyvärinen, A., & Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural computation*, *9*(7), 1483–1492.

Izenman, A. J. (2008). *Modern multivariate statistical techniques : Regression, classification and manifold learning* (Vol. 1). Springer.

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, *99*(3), 1053–1069.

Kilian, L., & Lütkepohl, H. (2017). *Structural vector autoregressive analysis*. Cambridge University Press.

Leurgans, S. E., Moyeed, R. A., & Silverman, B. W. (1993). Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *55*(3), 725–740.

Lewinson, E. (2022). *Python for finance cookbook: Over 80 powerful recipes for effective financial data analysis, 2nd édition*. Packt Publishing Ltd.

Lütkepohl, H., Meitz, M., Netšunajev, A., & Saikkonen, P. (2021). Testing identification via heteroskedasticity in structural vector autoregressive models. *The Econometrics Journal*, *24*(1), 1–22.

Lütkepohl, H., Staszewska-Bystrova, A., & Winker, P. (2018). Estimation of structural

impulse responses: Short-run versus long-run identifying restrictions. *AStA Advances in Statistical Analysis*, *102*, 229–244.

Maxand, S. (2020). Identification of independent structural shocks in the presence of multiple gaussian components. *Econometrics and Statistics*, *16*, 55–68.

Mertens, K., & Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the united states. *American economic review*, *103*(4), 1212–1247.

Miranda-Agrippino, S., & Ricco, G. (2023). Identification with external instruments in structural vars. *Journal of Monetary Economics*, *135*, 1-19. Retrieved from https://www.sciencedirect.com/science/article/pii/S0304393223000132 doi: https://doi.org/10.1016/j.jmoneco.2023.01.006

Moneta, A., & Pallante, G. (2022). Identification of structural var models via independent component analysis: a performance evaluation study. *Journal of Economic Dynamics and Control*, *144*, 104530.

Preston, A. J. (1978). Concepts of structure and model identifiability for econometric systems. *Stability and inflation*, 275–97.

Rigobon, R. (2003, 11). Identification Through Heteroskedasticity. *The Review of Economics and Statistics*, *85*(4), 777-792. Retrieved from https://doi.org/10.1162/003465303772815727 doi: 10.1162/003465303772815727

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

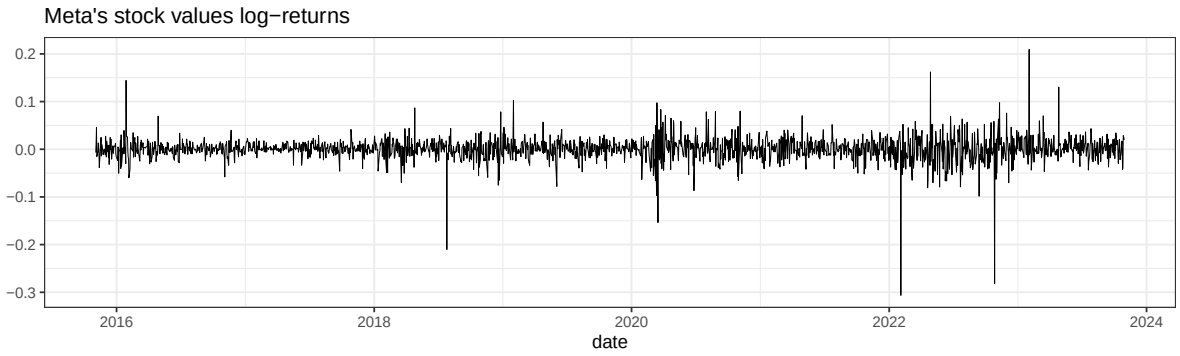
Stock, J. H., & Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, *128*(610), 917–948.

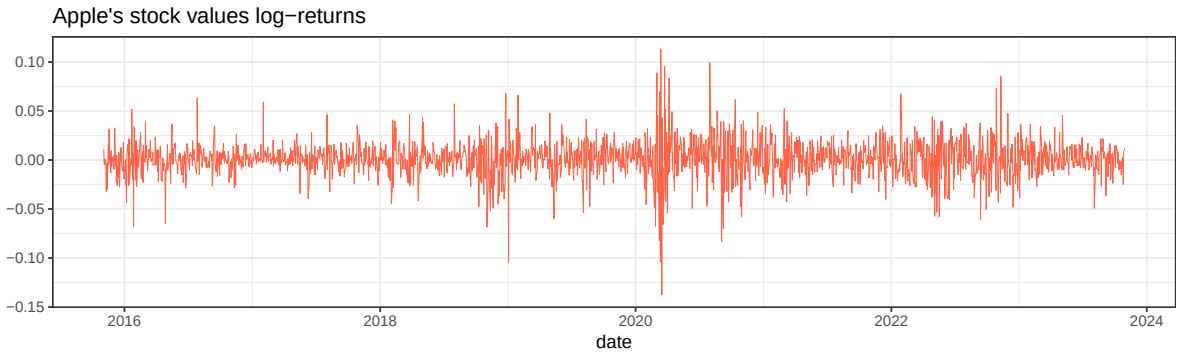
Tse, Y. K., & Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics*, *20*(3), 351–362.

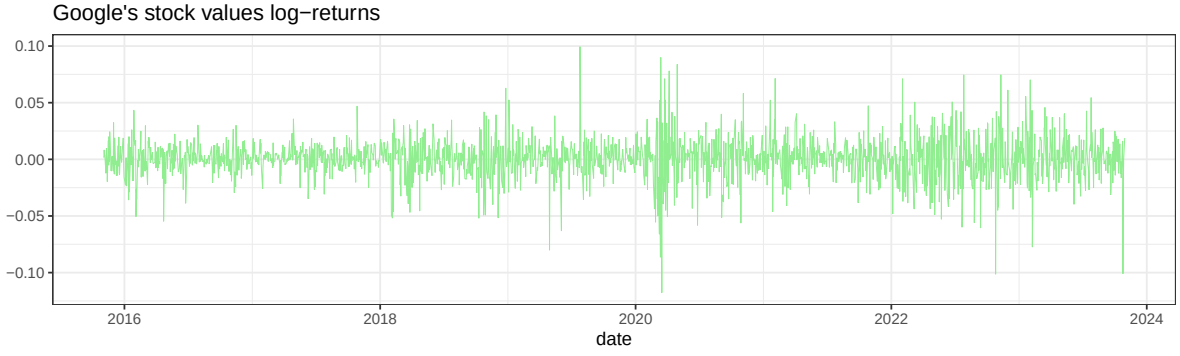
Appendix A

GAFAM stock data analysis

A.1 Indexes log-returns plots







A.1. Indexes log-returns plots

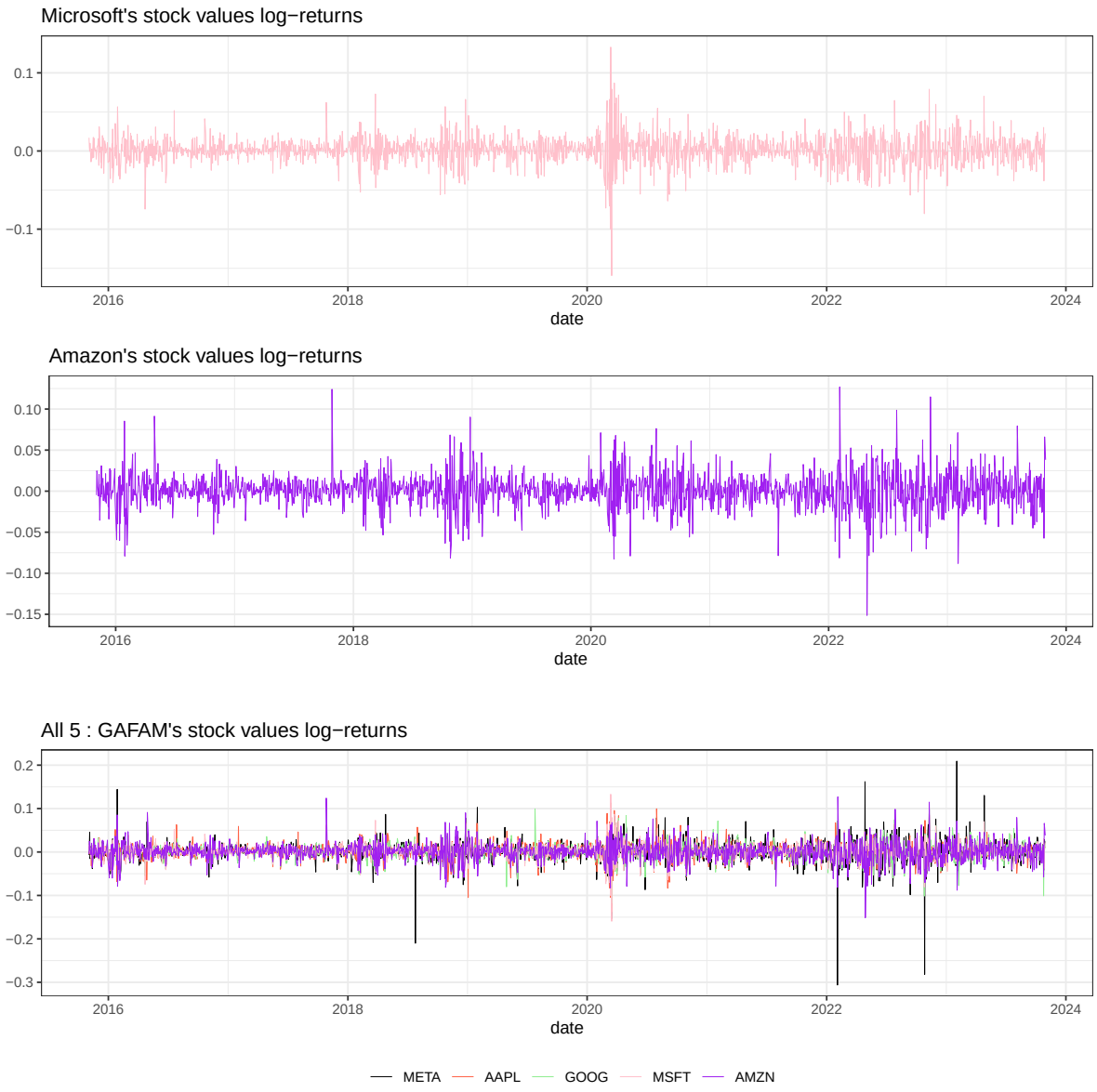


Figure A.1: log-returns of GAFAM stock values from 2015-10-31 to 2023-10-31

A.2 Data stationarity study

A.2.1 Dicker Fuller test for stationarity

	test.stat	lag.order	p-value	Decison
	Observed data			
Meta	−1.8641425	12	0.6358345	non-stationary
Apple	−2.6315216	12	0.3109658	non-stationary
Google	−2.0537982	12	0.5555441	non-stationary
Microsoft	−2.4593605	12	0.3838499	non-stationary
Amazone	−1.5950683	12	0.7497467	non-stationary
	log-returns data			
Meta	−12.8198284	12	0.01	stationary
Apple	−11.8066805	12	0.01	stationary
Google	−12.8203464	12	0.01	stationary
Microsoft	−13.4723471	12	0.01	stationary
Amazone	−12.5705211	12	0.01	stationary

Table A.1: Dicker Fuller test for stationarity results

A.2.2 Causality exploration

	Min	1st Qu.	Median	Mean	std	3rd Qu.	Max
Meta	-85.4	-2.7	0.0	0.0	6.1	3.1	35.5
Apple	-10.6	-1.1	0.0	0.0	2.4	1.2	11.9
Google	-13.5	-0.8	0.0	0.0	2.0	0.9	10.1
Microsoft	-23.6	-2.1	0.0	0.0	4.4	2.2	19.8
Amazon	-20.4	-1.4	0.1	0.0	2.9	1.4	18.8

Table A.2: De-meaned data summary statistics

response \ predictor	Meta	Apple	Google	Microsoft	Amazone
Meta	-	2.660 (0.003)	1.562 (0.111)	1.146 (0.323)	4.445 (0.000)
Apple	2.546 (0.004)	-	2.301 (0.011)	2.797 (0.001)	2.809 (0.001)
Google	2.492 (0.005)	1.940 (0.036)	-	2.118 (0.020)	3.547 (0.000)
Microsoft	3.5102 (0.000)	5.934 (0.000)	3.751 (0.000)	-	2.920 (0.001)
Amazon	0.695 (0.729)	1.693 (0.077)	1.389 (0.179)	1.276 (0.238)	-

Table A.3: Granger causality test results.

A.3 Standardized residuals analysis

A.3.1 Mean and Covariance matrix

	Meta	Apple	Google	Microsoft	Amazon
Mean	0.021	0.052	0.013	0.049	-0.005

Table A.4: BEKK model's residual mean

	Meta	Apple	Google	Microsoft	Amazon
Meta	1.093	-0.026	-0.046	0.002	-0.018
Apple	-0.026	1.040	0.052	0.029	0.024
Google	-0.046	0.052	1.046	-0.002	0.008
Microsoft	0.002	0.029	-0.002	0.989	-0.016
Amazon	-0.018	0.024	0.008	-0.016	1.033

Table A.5: Covariance matrix of BEKK's model residuals

A.3.2 Standardized residuals distribution

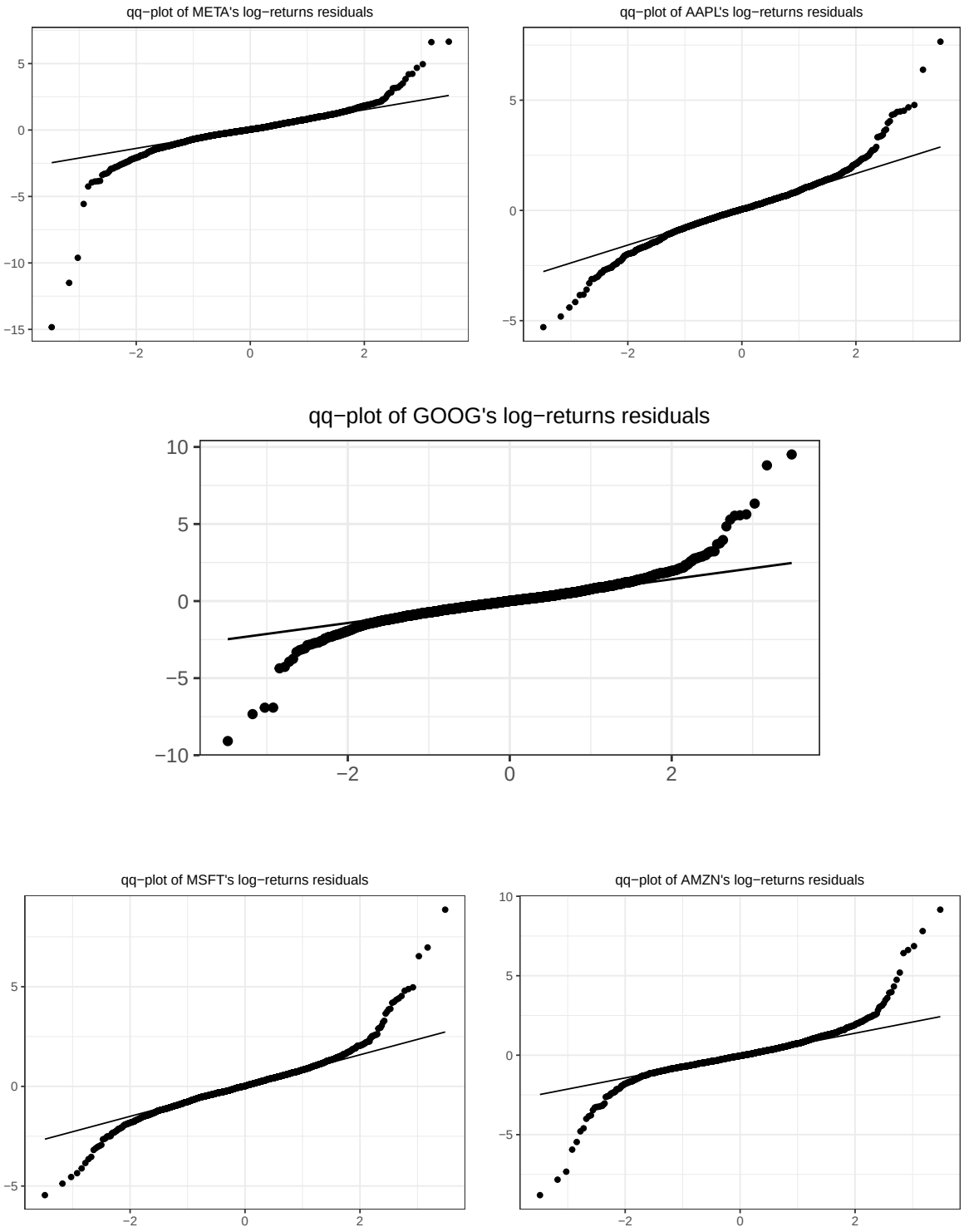
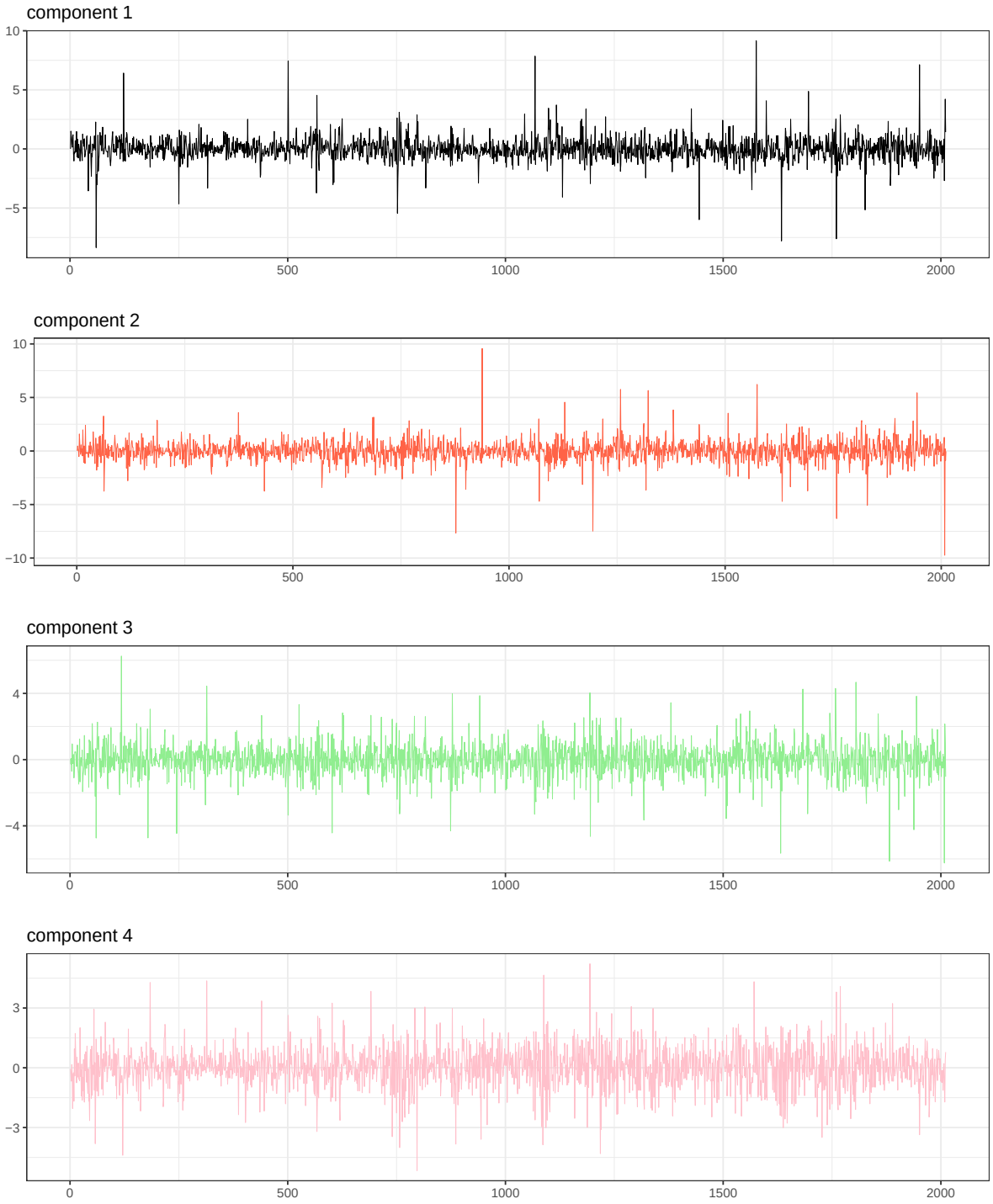


Figure A.2: qq-plot of standardized residuals

A.4 Retrieved components

A.4.1 Retrieved components plots



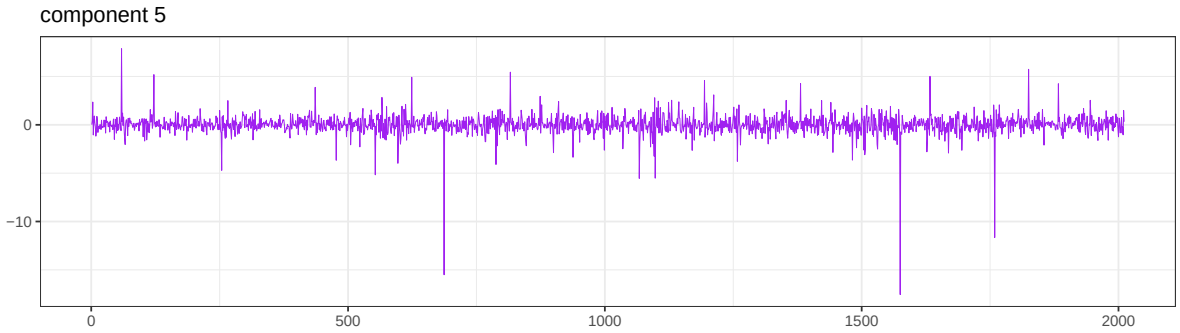
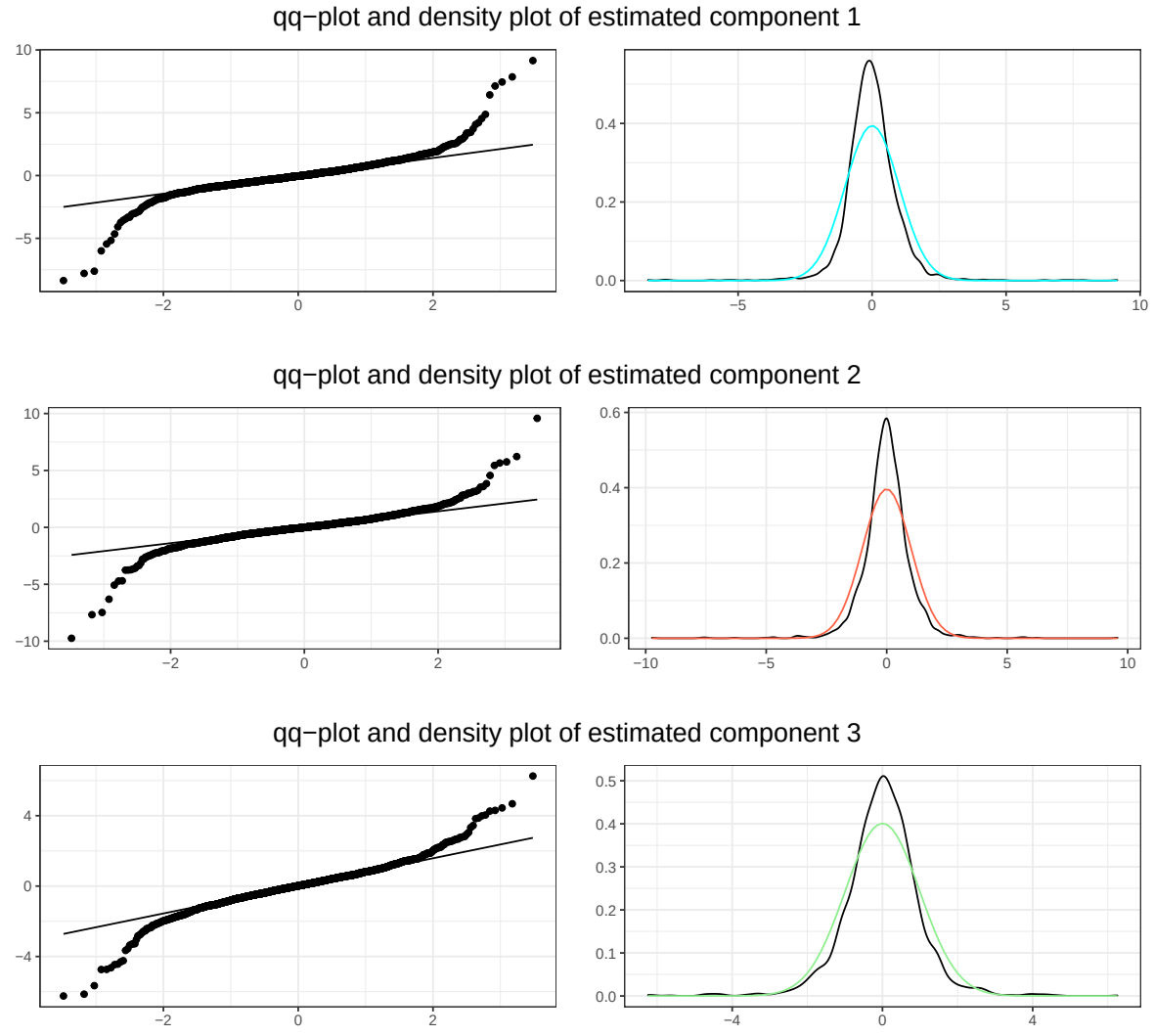


Figure A.3: Obtained components from kernelICA-KGV

A.4.2 Retrieved components distribution



A.4. Retrieved components

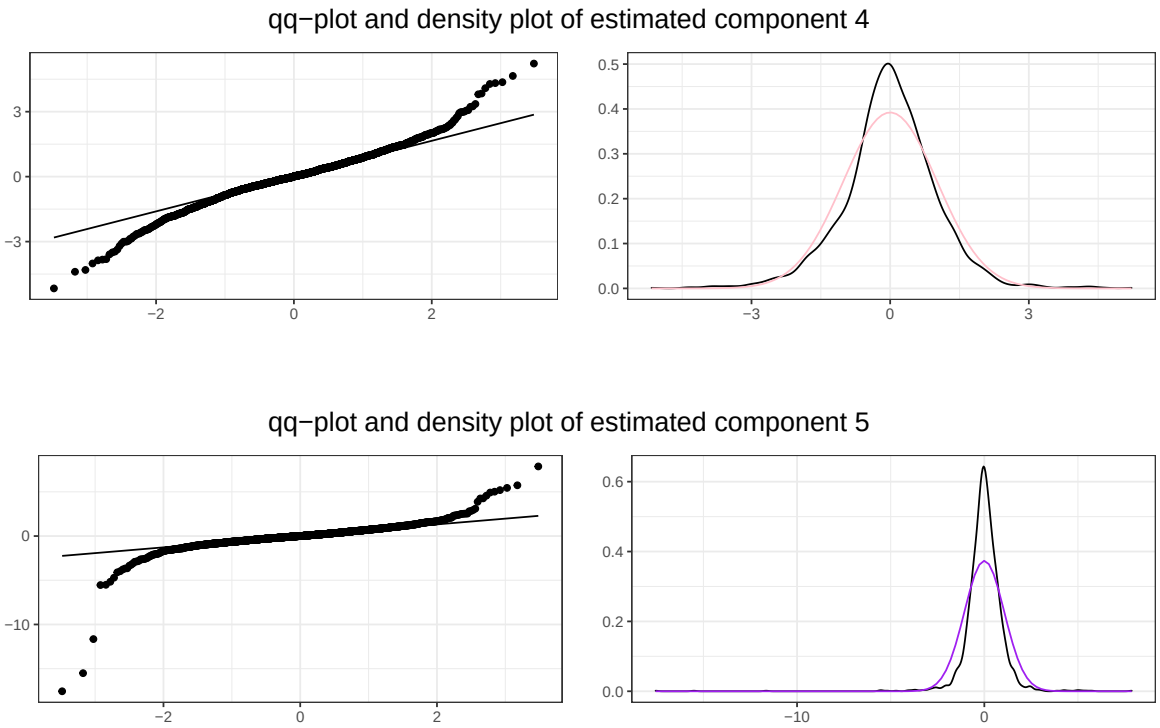


Figure A.4: Obtained components from kernelICA-KGV quantile-quantile and density distribution plots.

Appendix B

Volatility impulse responses

B.1 BEKK models parameters significance

matrix	estimated	t-value
C	$\begin{pmatrix} 0.0017 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0018 & 0.0022 & 0.0 & 0.0 & 0.0 \\ 0.002 & 0.0001 & 0.0016 & 0.0 & 0.0 \\ 0.0024 & 0.0003 & 0.0003 & 0.0017 & 0.0 \\ 0.0012 & 0.0005 & 0.0006 & 0.001 & 0.0011 \end{pmatrix}$	$\begin{pmatrix} 27.7 & 0.0 & 0.0 & 0.0 & 0.0 \\ 8.6 & 20.9 & 0.0 & 0.0 & 0.0 \\ 12.1 & 1.3 & 26.2 & 0.0 & 0.0 \\ 10.2 & 1.7 & 1.9 & 17.4 & 0.0 \\ 13.2 & 6.9 & 7.8 & 6.4 & 6.0 \end{pmatrix}$
A	$\begin{pmatrix} 0.1166 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.1538 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.149 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.1762 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.1106 \end{pmatrix}$	$\begin{pmatrix} 39.55 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 22.6 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 27.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 27.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 24.6 \end{pmatrix}$
B	$\begin{pmatrix} 0.989 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.971 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.975 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.963 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.987 \end{pmatrix}$	$\begin{pmatrix} 2522.4 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 399.3 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 627.2 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 354.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1179.2 \end{pmatrix}$

Table B.1: Estimated parameter matrices of BEKK model

B.2 VIRFs of negative shocks

B.2. VIRFs of negative shocks

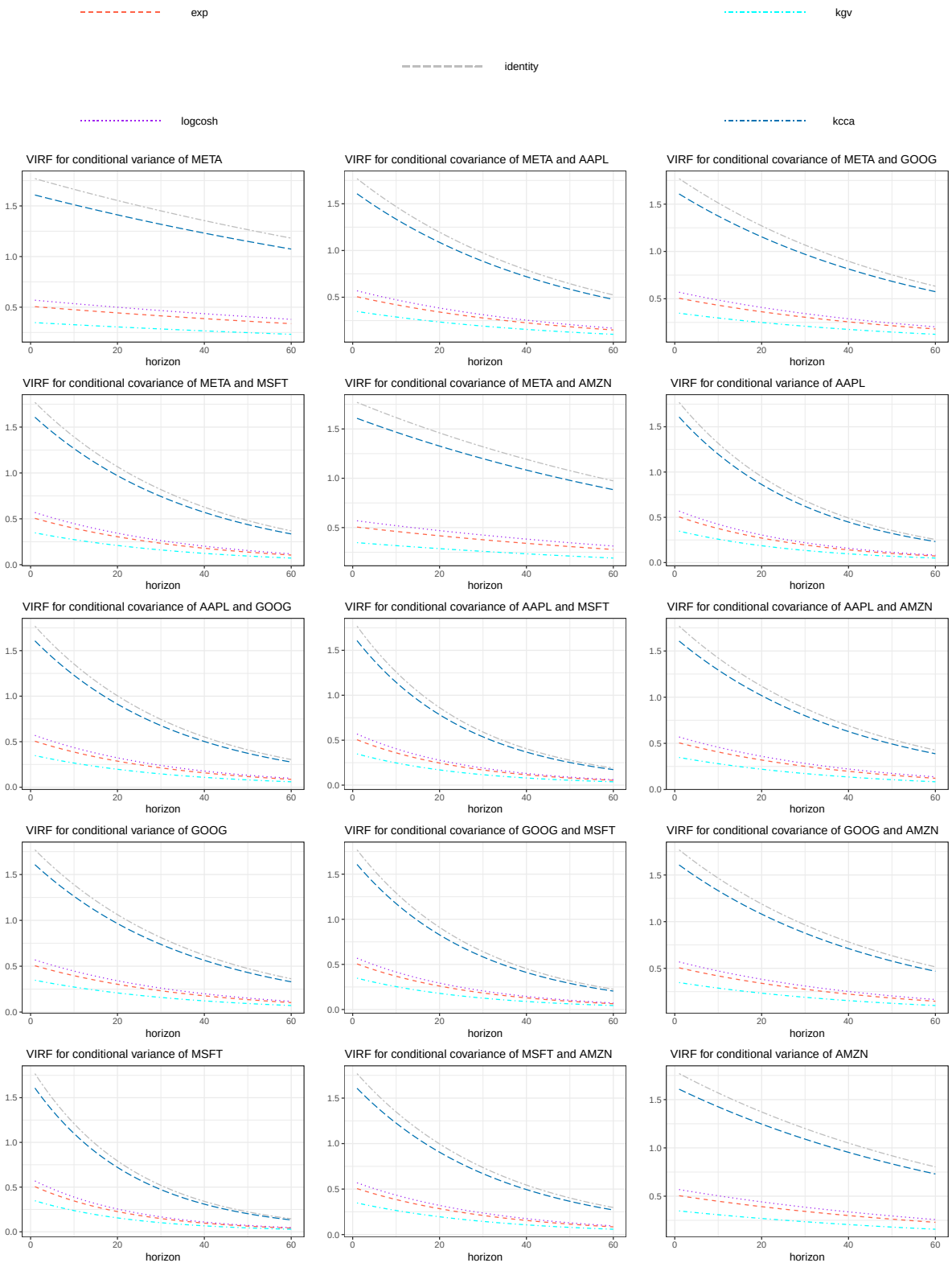


Figure B.1: Plots of the estimated VIRF of -6 points shock on 2023-10-13 on META

B.2. VIRFs of negative shocks

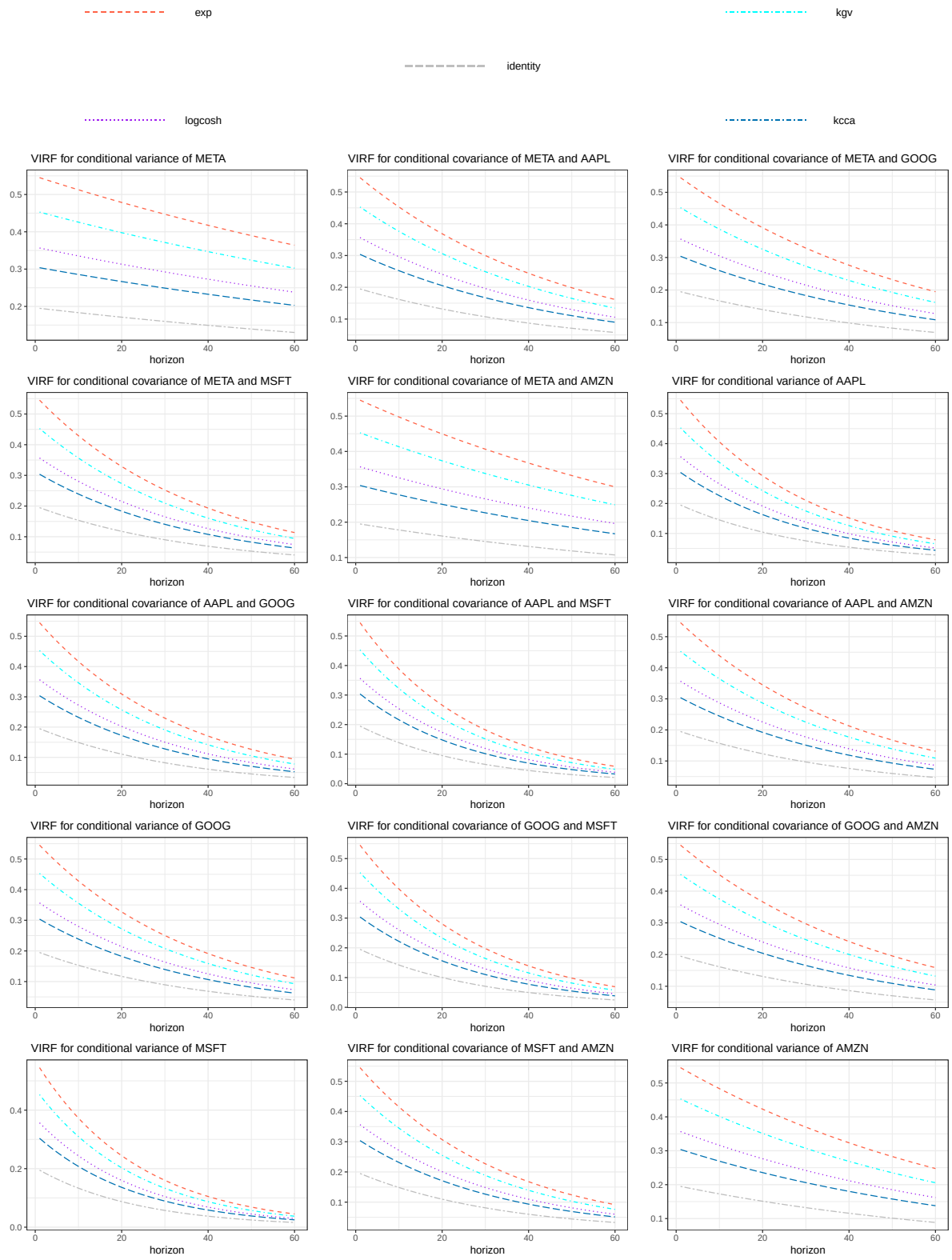


Figure B.2: Plots of the estimated VIRF of -6 points shock on 2023-10-13 on APPL

B.2. VIRFs of negative shocks

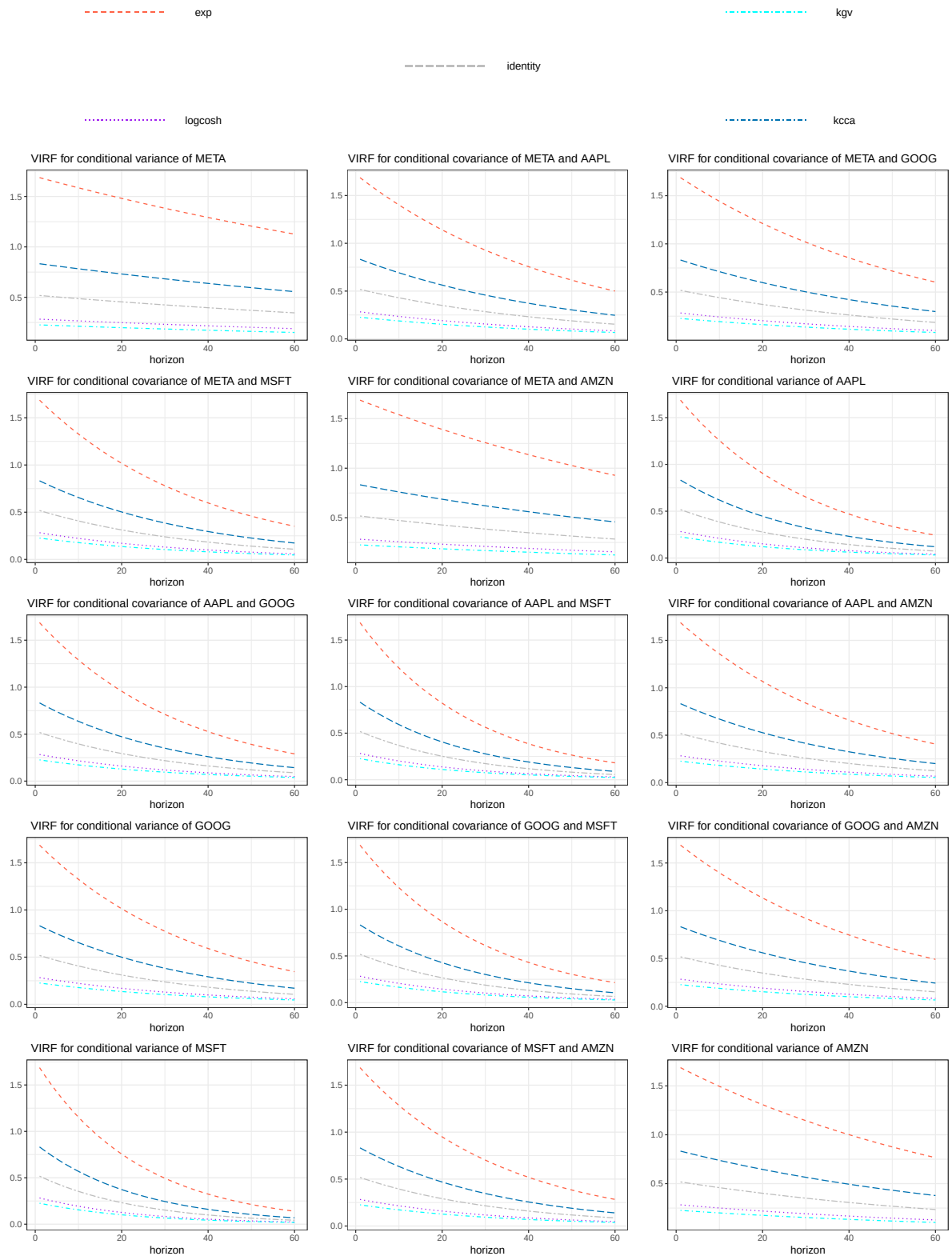


Figure B.3: Plots of the estimated VIRF of -6 points shock on 2023-10-13 on GOOG

B.2. VIRFs of negative shocks

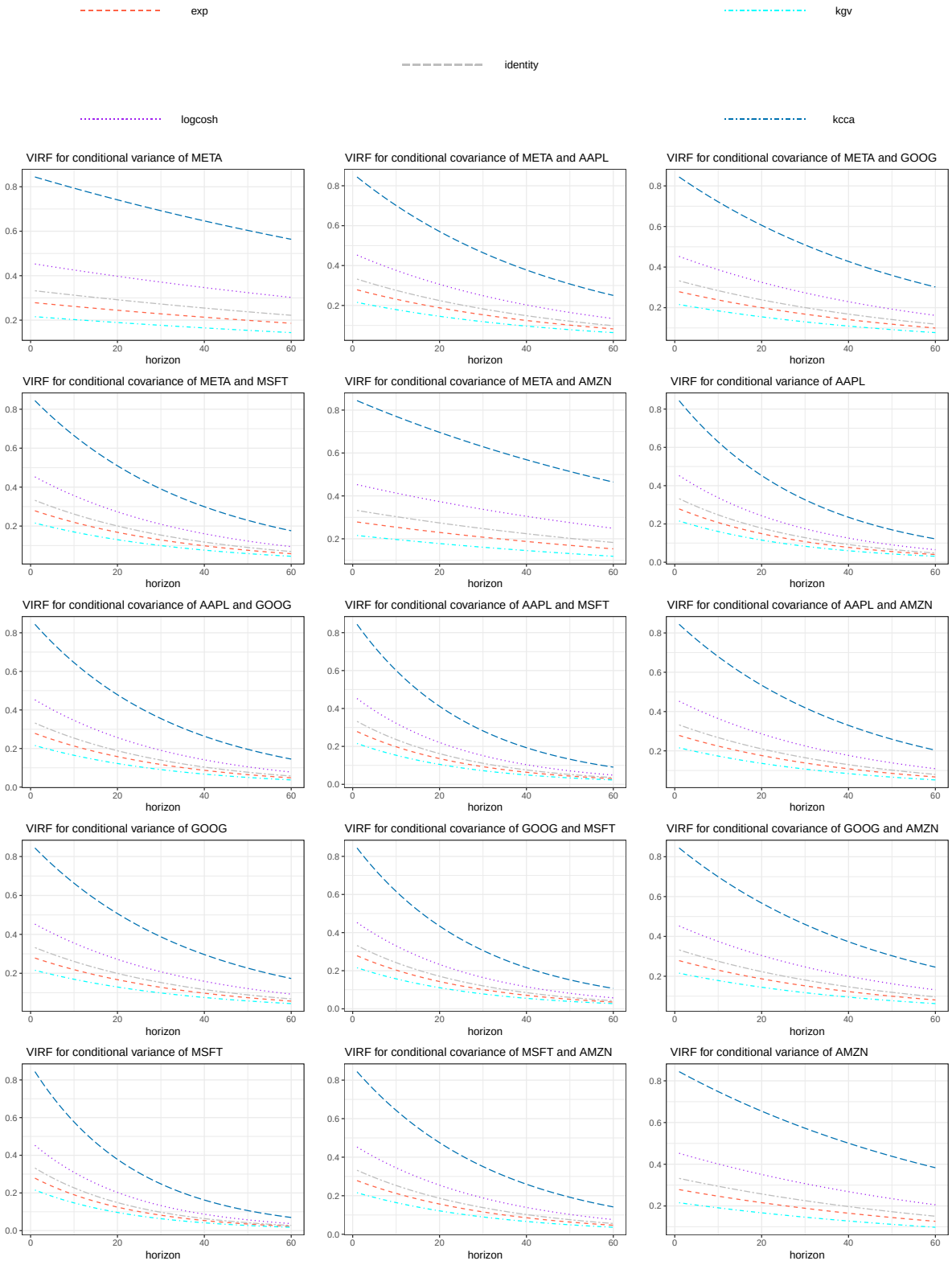


Figure B.4: Plots of the estimated VIRF of -6 points shock on 2023-10-13 on MSFT

B.2. VIRFs of negative shocks

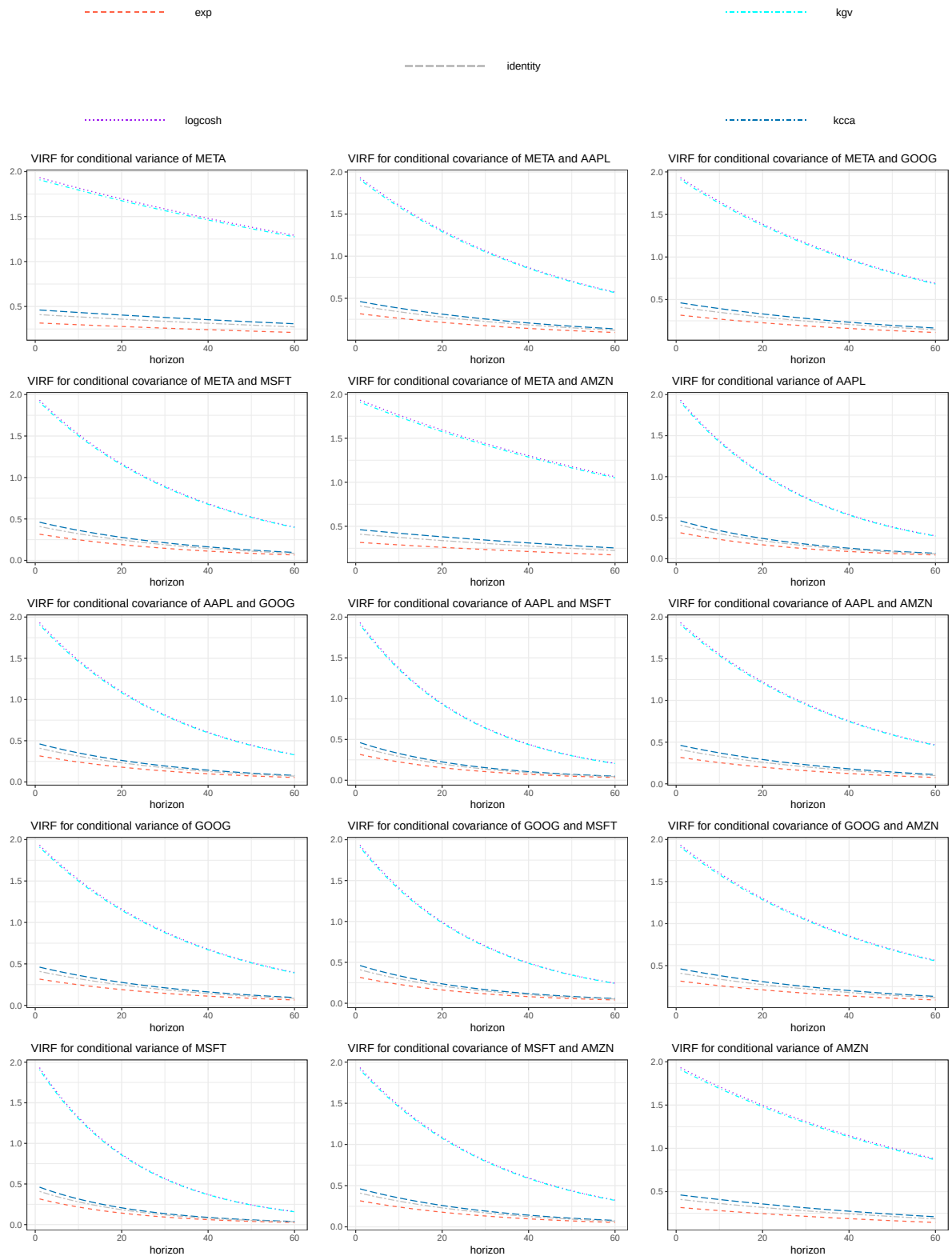


Figure B.5: Plots of the estimated VIRF of -6 points shock on 2023-10-13 on AMZN

Appendix C

R & python code

The R and python codes to replicate the results reported in this document are available at the following GitHub repository:

`https://github.com/mdahfienon/master-data-science-thesis.git`

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté des sciences
Place des Sciences, 2 bte L6.06.01, 1348 Louvain-la-Neuve, Belgique | www.uclouvain.be/sc