Department of Economics and Finance

Course of Econometric Theory

A FACTOR AUGMENTED VAR MODEL FOR MONETARY POLICY UNDER CLIMATE RISK

Supervisor:
Prof. Paolo Santucci de Magistris

Co-Supervisor:
Prof. Francesco Lippi

Candidate:
Mariacristina Bottaro
ID: 734041

Academic year 2021/2022

Contents

Introduction

Climate change is an important topic nowadays since the effects are becoming stronger and clearer to society. If no action is taken to reduce the increase in temperature and lower Green House Gasses (GHG) emissions[1] consequences will be at best costly for the economy, at worst disastrous for human society. From the economic side what is expected from institutions and companies is that they should incorporate climate risk as a new risk in their risk management models and check whether and for how much, they are exposed to this risk. Climate risk could also be embedded in already common stated risks such as credit risk, market risk, reputational risk etc. A concrete example of the impact of climate risk for a company is for example the PG&E (Pacific Gas and Electric) case. PG&E is an american company that provides electricity and gas to households in the region of California. California since last few years has been subject to high increase in temperature that caused several fires. PG&E has been defined –by the Wall street Journal– as the first corporation subject to climate bankruptcy. Basically, PG&E declared bankruptcy after being hit with what was then estimated to be 30$ billion in liabilities tied to tragic wildfires in 2017 and 2018, which took 86 lives. Heat waves, and so wildfires, increased in number due to increasing temperature within the country. The company did not consider climatic risks -physical risk that are risk linked to one time event ie. floods, fires or permanent change in the temperature- in its models, so company's capital was not enough to repay clients when the event happened.

Climate risk is also important for financial institutions. Banks that for instance, lend to corporations that operate in "brown" sectors, meaning industries that produce high GHG emissions, are as a consequence, affected by climate risk. This is because those companies can be subject to policy measures e.g. products cannot be used in specific places due to their emission intensity –so called transition risk–, that lead to an increase in the price of their products or in a lower demand for those products (as more costly for company's customers). The decrease in demand cause company's PDs (Probability of defaults) to increase. Higher PDs, affect the bank's capital ratio directly. If the financial sector is affected so is the real economy.

In my thesis, however, I want to check to what extent climate risk, in particular physical risk, affect monetary policy. From the ECB Strategy Review published in September 2021, is highlighted that monetary policy is impacted by climate change as a result of climate impact on macro economy and financial markets that

[1]GHG emissions are compound gases that trap heat or longwave radiation in the atmosphere.Their presence in the atmosphere makes the Earth's surface warmer an example are CO2 or N20 (Nitrous dioxide)

are the transmission channels of monetary policy. Two aspects are important to consider. The first aspect concerns the possible implications of climate change and mitigation policies for the ability of central banks to fulfill their price stability mandate. The second aspect concerns the extent to which central banks themselves can play a supporting role in mitigating the risks associated with climate change, while staying within their mandate. The way through which climate change could affect monetary policy is then by impacting the transmission channel of monetary policy. As to what concerns the impact that climate has on financial markets as we have seen in the example before, one issue could be bank's capital ratio that should be increased accordingly to bank's exposures to "brown" sectors ie.sectors with high GHG emissions. A second issue could be that climate brings to a sudden re-pricing of climate-related assets or to re-evaluate – in negative- the collaterals that banks own for their loans. A change in value in banks' collaterals could lead to capital and liquidity shortage that weaken banks' ability to channel funds to the real economy. If the financial system is weakened the transmission of monetary policy is impaired. Another thing, that can happen, more due to macroeconomic shocks, is effects on market interest rates. If climate-related factors were to cause the interest rates to fall further, the policy rate could hit the lower bound more often thus limiting the monetary policy space for conventional tools. Reasons why the interest rate could fall are that climate change causing macroeconomic shocks on the demand or supply side, by reducing consumption from the demand side or production from the supply side can push interest rates downward. Moreover, another effect caused by climate related variables is that they could correct the identification of shocks relevant for the medium-term inflation outlook. This would make it more difficult to assess the monetary policy stance and potentially increase the prevalence of output and price stabilization trade-offs for central banks that focus on price stability. Uncertainty surrounding the magnitude of the effects of climate change and the horizon over which they will play out in the economy will further complicate the assessment of appropriate monetary policy actions. Uncertainty may also destabilize the expectation-formation process of economic agents, in particular with regard to inflation expectations. The credibility of the central bank may be compromised if the time horizon is extended too far into the future and inflation targets are missed too often. In this case, clear communication about the policy intention of the central bank will be essential to mitigate credibility losses. Moreover, it is worth mentioning that the effects of physical risk, meaning extreme climate event ie. flood, earthquakes, or permanent change in temperature and transition risks meaning policy measures taken to let a smooth transition from high emission world to a low-carbon emissions one, could be asymmetric and heterogeneous across countries or regions, complicating even more the behavior of a central bank.
So, to sum up, the interest rate channel, the credit channel, the asset price channel, the exchange rate channel and the expectation channel can all be affected by climate change variables. Moreover, they can be affected differently according to the physical or transition risk.

I use factor models to check these theories, more in particular FAVAR model, that in the climatic econometrics literature was never used before to assess the impact of climate change. FAVAR model is actually an extension of Dynamic Factor Models and Vector Autoregressive models Kilian and Lütkepohl (2017), Stock and Watson (2016). Factor models are useful in this case because they allow to summarize a wide concept expressed by several variables within just one or few factors of interest, moreover FAVAR overcomes the limit of VAR models ie. omitted variables bias and small-scale studies. Actually I used VAR to check, just for a few variables, as a preliminary assessment, whether climate can have a direct impact on monetary policy. I hereby used a VAR(2) model with variables EU GDP, ECB Refinancing rate and GHG emission to compute the impulse response of a shock in GHG emission. The results I have obtained perfectly match with theory meaning that the impulse response of the ECB Refinancing Rate is close to zero pointing out that no direct climate effect exist on monetary policy. Historically, in fact, the ECB has never tailored monetary policy to address a climate-related event. I provide details of this study in the last chapter.

Starting from this result, as I showed that no direct impact of climate exist on monetary policy, represented by ECB Refinancing Rate, I check then the impact of climate change in a variety of economic variables, that represent transmission channels for monetary policy. To do so I adopt a FAVAR model.
The FAVAR model I use is based on Bernanke et al. (2005) model, where the authors measured monetary policy shock on a vast economic panel. Actually as I am checking for a different shocks I re-calibrate the model to conduct my analysis. I create two different panels one for economic variables and another one for climate variables.

The economic panel is formed by GDP at EU aggregate and of major european countries, Crop Prices (Soft Wheat) of major European countries and European inflation (HCPI) at EU aggregate level and of major European countries. The climatic panel is instead formed by mainly EU GHG emissions and EU CO2 emissions in kilotonnes. All the data I use are historical with annual frequency starting from 1990 to 2019 and extracted from Eurostat and the World Bank Database.
From these two panels I extract two factor of interests one summarizing the economic panel the other one summarizing the climatic panel. Starting values for the factors have been computed by PCA (Principal Components Analysis) further updates of the factors have been estimated using Carter and Kohn (1994) algorithm which relies on Bayesian estimation Mira and Sargent (2003).
The outcome of my study is that the economic panel GDPs, HCPI and Prices so as interest rates are impacted negatively by a shock in the climatic factor, but it is difficult to understand whether the shock is demand or supply driven. Determining the overall balance of supply and demand shocks may differ between individual events. Moreover the effect of the shock seems to be statistically significant only in the short term. Similar results have been found previously by Ciccarelli and Marotta (2021) that studied the impact of physical and transition risks using a VAR(2) on GDP, HPCI and prices. As a consequence conclusions on the effect on monetary policy cannot be yet drawn. Further investigations are needed, both by changing the model and use for instance mixed-frequency VAR model or re-conduct this study when more data will be available so time series will be longer.
The thesis is dived as follows: factors models literature review (Chapter I), description of VAR and FAVAR

model according to Bernanke et al. (2005) (Chapter II), a focus on Bayesian Estimation of FAVAR model (Chapter III), empirical application (Chapter IV).

Chapter 1

Literature Review of factor models

1.1 Introduction

Due to the fact that climate change is a broad concept represented by different time series ie. variables, I decided to analyze its impact on economic variables, through factor models. By identifying just one factor that summarizes a bigger climatic dataset –so not to lose information– make it easier to model the phenomenon and conduct analyses. Generally speaking, factor models have been used widely in recent year due to the increasing volume of available financial and economic data, that pushed econometricians to develop or adapt methods that efficiently summarise the information contained in large databases within few factors of interest. More in particular, as stated in the introduction I use a FAVAR model in my application that is an extension of dynamic factor models and VAR (Vector Autoregressive) firstly applied in early 2000s. Within this chapter I provide a review of factor models, that are the basis for FAVAR, specifically Static Factor Models (SFMs) referring to the studies of Choi (2012), approximate static factor models, Dynamic Factor Models based on the review done by Stock and Watson (2016), the estimation methodology of Doz et al. (2011). Within the last paragraph, I briefly discuss how to determine the number of factors for DFMs according to the papers of Bai and Ng (2002), Bai and Ng (2007) and Amengual and Watson (2007) that provide a re-elaboration of the information criteriab (IC) proposed by Bai and Ng (2007).

1.1.1 Static Factor Models

In this type of model, a small number of unobservable variables k provides a linear explanation of a small number of observed N variables so that the number of factors is such that a single factor can generally explain most of the variance, that is, k = 1. The classical static factor model assumes the form:

$$X_t = \Lambda F_t + \xi_t, \quad (1.1.1)$$ [2]

[2] In which the factors f_t do not possess their own dynamic and the relationship between the factors and variables is linear with constant weight over time. This model can be estimated either by assuming that the variables are IID or by assuming

where $X_t = (x_{1t}, ..., x_{Nt})'$ each $x_{it} \sim (0, \sum_x)$ i.i.d is a vector of N observed variables, $F_t = (f_{1t}, ...f_{kt})'$ each F_t is a k-dimensional vector of unobserved common factors, and k is typically much smaller than N. Accordingly Λ is a $N \times k$ matrix of factor loading. Finally, $\xi_t \sim (0, \Sigma_\xi)$ i.i.d is a N-dimensional vector of idiosyncratic components. As we can see from the equation, these type of models establish the relation between factors and observed variables X_t. The series x_{it} and f_{it} are assumed to be stationary, to have finite variance and to be standardised. Meaning that we have the following assumptions:

- The factors F_t, are centered $E(F_t) = 0$, and are mutually orthogonal for all t, that is: $\forall t$, $E(F_{jt}F'_{jt}) = 0$ for $j \neq j'$. Consequently, the variance-covariance matrix of F_t, $\Sigma_F = E(F_tF'_t)$ is a diagonal matrix.
- The idiosyncratic process ξ_t and ξ'_t are mutually orthogonal for all $i \neq i'$, with $E(\xi_t) = 0$ So, the variance-covariance matrix is a diagonal matrix
- The factors F_t and the idiosyncratic noise ξ_t are not correlated so we have that $E(F_t\xi_t) = 0$
- The variables are assumed to be independent and identically distributed over time so that $t \neq t'$, $E(F_t, F'_t) = 0$ and $E(\xi_t\xi'_t) = 0$

This model has a problem of identification. In fact, if we take any nonsingular kxk matrix Q and defining $F^*_t = QF_t$ and $\Lambda^* = \Lambda Q^{-1}$ we end up in having $\Lambda F_t = \Lambda^* F^*_t$. Exact identification of the model can be ensured by choosing Λ such that $\Lambda'\Lambda = I_k$ holds and choosing the factors such that Σ_f is a diagonal matrix with elements in decreasing order. According to that we need to impose at least k^2 restrictions (order condition) to remove the indeterminacy. Order conditions can be applied to the factors that can be estimated by PCA (Principal Components Analysis) meaning that the components explaining the majority of the variance (Principal Components) of the model will be considered as factors. Referring to Bai and Wang (2014), we can say that three commonly applied normalizations are PC1, PC2 and PC3 (where PC stands for Principal Components):

- PC1=$\frac{1}{T}F'F = I_k, \Lambda'\Lambda$ is diagonal with distinct entries
- PC2: $\frac{1}{T}F'F = I_k$ the upper k x k block is lower triangular with nonzero diagonal entries
- PC3: the upper kxk block of Λ is given by I_k.

PC1 is often imposed when the maximum likelihood estimation is used in classical factor analysis. PC2 can be applied to reduce the system to a recursive system of simultaneous equations, because by blocking the factors' loadings, then the system can be estimates as a SUR. PC3 is linked with the measurement error problem such that the first observable variable x_{1t} is equal to the first factor f_{1t} plus a measurement error e_{1t} and so on. Each of the above set of normalizations yields to a k^2 restrictions meeting the order

that there is a time dynamic withing the variables, so point 4 is abandoned.

condition for identification. In our discussion we will take as restrictions as it is common to do in classical factor analysis, PC1. This means that we will choose the factor loading matrix such that it has orthonormal columns, implying that:

$$\Lambda'\Lambda = I_k, \tag{1.1.2}$$

or to choose uncorrelated factors with variances normalized to 1,

$$F_t \sim (0, I_k), \tag{1.1.3}$$

in the latter case, as F_t and ξ_t are not correlated as in assumption four above and the factors are orthogonal we obtain that:

$$\Sigma_x = \Lambda\Lambda' + \Sigma_\xi. \tag{1.1.4}$$

such normalizations are useful, but they are not sufficient for uniquely identifying the model. Because if we normalize the factors as said $f_t \sim (0, I_k)$, Λ is still not unique without further restrictions. This can be seen by choosing an orthogonal matrix Q and defining $\Lambda^* = \Lambda Q$. Thereby we arrive at the decomposition, $\Sigma_X = \Lambda^*\Lambda^{*\prime} + \Sigma_\xi$ Exact identification of the static factor model can be ensured by choosing both restrictions on factors and factor loadings. Λ such that $\Lambda'\Lambda = I_k$ holds and choosing the factors such that Σ_F is a diagonal matrix with distinct diagonal elements in decreasing order so by imposing two condition both on the loading matrix and on the factors. By elements in decreasing order within the Σ_F it is meant that the first factor has the largest variance and, hence, explains the largest part of the variance of x_t etc. The requirement that the variances of the factors have to be distinct means that the columns of Λ cannot simply be reordered. Further identification restrictions for the model are pointed out in Bai and Ng (2013).

1.1.2 Estimating Static factor models

Starting from the assumptions and restrictions applied in the paragraph above i.e. that the factor loadings were known and normalized such that $\Lambda'\Lambda = I_k$, a natural estimator for the factors would be obtained by left multiplying (1.1.1) with Λ' and dropping the idiosyncratic term, $F_t = \Lambda' X_t$.

However, in practice the factor loadings are typically unknown. Despite this, to estimate the factors we still want to minimize the sum of the squared idiosyncratic errors. Minimizing the variance of the idiosyncratic components amounts to maximize the part of the variance of the observed variables explained by the common factors. In other words, we may estimate the factor loading and factors by minimizing the sum of squared errors,

$$\arg\min_{\Lambda, F_T} T^{-1} \sum_{t=1}^{T} (X_t - \Lambda F_t)'(X_t - \Lambda F_t). \tag{1.1.5}$$

Given the problem above of having two unknowns, we can find a solution by doing PCA so:

1. Find the r largest eigenvalues $\lambda_1 > \lambda_2$... etc. of $\Sigma_X = T^{-1} \sum_{t=1}^{T} X_t X_t'$ and the corresponding orthonormal eigenvectors

2. Choose $\hat{\Lambda} = [\lambda_1, \lambda_2 ...]$

3. Express the factor estimate as $\hat{F}_t = \hat{\Lambda}' X_t$.

The estimator $\hat{\Lambda}$ is the so-called principal components estimator of Λ. Given ortogonality of the eigenvectors, it satisfies $\Lambda'\Lambda = I_k$. Then the factors are the principal components and $\hat{\Sigma}_F = T^{-1} \sum_{t=1}^{T} F_t F_t' = diag(\lambda_1, \lambda_2, ...)$. The eigenvalues are the empirical variances of the factors with λ_1 representing the variance of the principal component with the largest distribution to the variance of the data etc. Proof is showed in Appendix I. The PC estimator is the Maximum Likelihood estimator if the observations x_{it} come from a normal distribution and the idiosyncratic components have equal variances such that $\Sigma_\xi = \sigma^2 I_N$. In that case, the factors and idiosyncratic components are normally distributed according to $F_t \sim N(0, \Sigma_F)$ and $\xi_t \sim N(0, \sigma^2 I_N)$. If the variances of the idiosyncratic components are heterogeneous,instead, the log-likelihood is:

$$logl(\Lambda, F_T, \Sigma_\xi) = c - \frac{T}{2} \log(\det(\Sigma_\xi)) - \frac{1}{2} tr(\sum_{t=1}^{T} (X_t - \Lambda F_t)(X_t - \Lambda F_t)' \Sigma_\xi^{-1}). \tag{1.1.6}$$

Where c is a constant. Since Λ has $N \times k$ parameters and F has $T \times k$ parameters, the number of parameters to be estimated is very large. This leads to a loss in efficiency. The thing is that if N is small, but T is large, then it can be shown that the likelihood function diverges to infinity by a certain choice of parameters, so a global maximum does not exist.[3]

This log likelihood function is therefore unbounded in general, hence, does not have a global maximum. Thus, standard ML estimation cannot be used. The likelihood function has local maxima, however, allowing us to consider local maxima in the neighborhood of the true parameter vector. If an estimator of Σ_ξ is available, the factor loadings and factors may be estimated by a feasible GLS method based on the minimization problem:

$$\arg\min_{\Lambda, F_T} T^{-1} \sum_{t=1}^{T} (X_t - \Lambda F_t)' \tilde{\Sigma_\xi}^{-1} (X_t - \Lambda F_t). \tag{1.1.7}$$

1.1.3 Approximate Static Factor Models

So far, we have considered what can be called an exact static factor model in which the idiosyncratic components are clearly separated from each other and from the factors. For economic data such an assumption may be too strict. An alternative approach is to work with a approximate factor models that allows for the

[3] According to Anderson T.W (2003). If we express the log-likelihood function like

$$L = \frac{1}{[2\pi)^p \prod_{i=1}^{p} \Phi_{ij}]^{\frac{N}{2}}} \prod_{i=1}^{p} exp - \left(\frac{1}{2} \sum_\alpha = 1^N \frac{(x_{i\alpha} - \mu_i - \sum_{j=1}^{m} \lambda_{ij} f_{ja}^2)}{\Phi_{ii}} \right)$$

Where $x_\alpha = (x_{1\alpha}, \ldots x_{p\alpha})'$ is an observation on X_α given by $X_\alpha = f_\alpha + \mu + U_\alpha$ with f_α being a nonstochastic vector satisfying $\sum_{\alpha=1}^{N} f_\alpha = 0$. It can be shown that if $\mu_1 = 0$, $\lambda_1, 1 = 1$, $\lambda_{1,j} = 0$ with $j \neq 1$, $f_{1,a} = x_{1,a}$. Then $(x_{i\alpha} - \mu_i - \sum_{j=1}^{m} \lambda_{ij} f_{ja})^2 = 0$ and $\Phi_i i$ does not appear in the exponent but appears only in the constant. This means that $\Phi_{ii} \to 0, L \to \infty$. So, it has no global maxima.

possibility that the idiosyncratic component has some correlation across series, so maybe data derived from the same survey might have correlated measurement error. In this case we talk about approximate factor model. In the approximate factor model

$$\Sigma_x = \Lambda \Sigma_F \Lambda' + \Sigma_\xi, \tag{1.1.8}$$

where Σ_ξ is not necessarily a diagonal matrix. Assuming that the common factors are normalized to have variance one, Chamberlain (1983) define approximate factor model by the conditional that Σ_x has only k unbounded eigenvalues when $N \to \infty$, where N is the number of variables considered by the researcher. The common factors are defined by the requirement that there exists a sequence of $N \times k$ matrices Λ and positive definite covariances Σ_v such that: $\Sigma_x = \Lambda\Lambda' + \Sigma_\xi$ and the maximum eigenvalue of Σ_ξ is bounded when $N \to \infty$. Thus, the relative variance share of each idiosyncratic component is small, when the number of variables is large. Chamberlain (1983) apply this model to a financial market with many assets. Obviously, in that case identification of the model becomes more difficult and conditions different from those stated earlier are required. In fact, is is even possible that Σ_ξ, has a factor decomposition that needs to be clearly separated from the common factor part captured by $\Lambda\Lambda'$, at least asymptotically.

1.2 The Dynamic Factor Models (DFMs)

The DFMs represent the evolution of a vector of N observed time series, X_t, in terms of a reduced number of unobserved common factors which evolve over time, plus uncorrelated disturbances which represent measurement error and/or idiosyncratic dynamics of the individual series. There are two ways to write the model. The dynamic form of DFM that represents the dependence of X_t on lags of the factors explicitly, while the static form represents those dynamics implicitly. The two forms lead to different estimation methods. What makes the DFM stand out for macroeconometric application is that the complex comovements of a potentially large number of observable series are summarized by a small number of common factors, which drive the common fluctuations of all the series.

1.2.1 Dynamic form of DFM

The dynamic form of DFM is obtained if the factors, f_t explaining the variables X_t are also allowed to enter in lagged form. The DFM, in fact, expresses a Nx1 vector X_t of observed time series as depending on a reduced number of unobserved or latent factors f_t and a mean-zero idiosyncratic component e_t, where both the latent factors and idiosyncratic terms are in general serially correlated. The DFM is:

$$X_t = \Lambda_0^f f_t + \Lambda_1^f f_{(t-1)} + \dots \Lambda_{q^*}^f f_{t-q^*} + v_t. \tag{1.2.1}$$

Assuming the same DGP (data generating process) for f_t and v_t this model can be written in lag operator notation as:

$$X_t = \Lambda(L) f_t + e_t. \tag{1.2.2}$$

$$\Phi(L)f_t = \eta_t, A(L)v_t = u_t. \tag{1.2.3}$$

With:

$$\Lambda(L) = \Lambda_0^f + \Lambda_1^f L + ... + \Lambda_{q*}^f L^{q*},$$

$$\Phi(L) = I_r - \Phi_1 L - ... - \Phi_s L^s,$$

$$A(L) = diag\left[a_1(L), ..., a_N(L)\right].$$

Where the lag polynomial matrices $\Lambda(L)$ and $\Phi(L)$ are N x q[4] and $q \times q$, respectively, and η_t is the $q \times 1$ vector of (serially uncorrelated) mean-zero innovation to the factors. The idiosyncratic disturbances are assumed to be uncorrelated with the factor innovations at all leads and lags, that is, $E[e_t \eta'_{t-k}] = 0 \ \forall$ k. In general, though, e_t can be serially correlated. The ith row of $\Lambda(L)$, the lag polynomial $\Lambda_i(L)$, is called the dynamic factor loading for the ith series, X_{it}. The term $\Lambda_i(L)f_t$ is the common component of the ith series. The idiosyncratic component in (1.2.1) can be serially correlated. If so, equation (1.2.1) and (1.2.3) are misspecified because they cannot be clearly estimated. In this case another method of estimation must be used, like FGLS giving a process to the idiosyncratic component. One possibility could be to assume that the idiosyncratic component, follows the univariate autoregression, $e_t = \delta_i(L)e_{it-1} + v_{it}$ where v_{it} is serially uncorrelated. To estimate DFMs there are two possibilities. The first one is based on estimate directly the dynamic form of the model which follow a Dynamic PCA approach described by Forni et al. (2000). The other method is to rewrite the dynamic model in static form and estimate it with standard PCA. As so, I will present how to rewrite the DFM in a static form that as showed in the Appendix the two forms are equivalent.

1.2.2 Static Form of DFMs

The static form of the DFM rewrites the dynamic form of a DFM model to depend on r static factors F_t. Rewriting the model in this way makes it likely to be estimated by standard principal components analysis and least squares estimators. Let $F_t = (f'_t, f'_{t-1}, \dots f'_{t-p})'$ denote a k x 1 vector of so-called "static" factors in contrast to the f_t that will be the dynamic factors called also primitive factors. Let $\Lambda = (\Lambda_0^f, \Lambda_1^f, \dots, \Lambda_p^f)$, where Λ_p^f is the N x q matrix of coefficients on the hth lag in $\Lambda(L)$. Similarly, let $\Gamma(L)$ be the matrix consisting of 1s, 0s and elements of $\Phi(L)$ in equation (1.2.3) such that the vector autoregression is rewritten in terms of F_t. So, the DFM can be rewritten like:

$$X_t = \Lambda F_t + e_t, \tag{1.2.4}$$

$$F_t = \Gamma(L)F_{t-1} + G\eta_t, \tag{1.2.5}$$

where equation (1.2.4) is the evolution of the dynamic factor in equation (1.2.3). G= $[I_q \quad 0_{qx(k-q)}]'$ because it is the identity used to express Φ in (1.2.3) in first-order form.

[4] q is the number of unobserved factors

As an example, consider there is a single dynamic factor f_t (so q=1), that all X_{it} depend only on the current and first lagged values of f_t, and that the equation for f_t in (1.2.4) has two lags, so $f_t = \Gamma_1 f_{t-1} + \Gamma_2 f_{t-2} + \eta_t$. Then the correspondence between the dynamic and static forms for X_{it} is:

$$X_{it} = \Lambda_0 f_t + \Lambda_1 f_{t-1} + e_{it} = \begin{bmatrix} \Lambda_0 & \Lambda_1 \end{bmatrix} \begin{bmatrix} f_t \\ f_{t-1} \end{bmatrix} + e_{it}.$$

$$F_t = \begin{bmatrix} f_t \\ f_{t-1} \end{bmatrix} = \begin{bmatrix} \Gamma_1 & \Gamma_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} f_{t-1} \\ f_{t-2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \eta_t = \Gamma F_{t-1} + G\eta_t. \tag{1.2.6}$$

The first expression writes out the equation for X_{it} in the dynamic form $\Lambda_i = [\Lambda_0 \Lambda_1]$ is the ith row of Λ, and the second expression is the equation for X_{it} in the static form.

Estimation of DFMs. In order to estimate the model, it is useful to left-multiply equation (1.2.4) by A(L) which is a lag operator (Kilian and Lütkepohl (2017)). So the model become:

$$A(L)X_t = \Lambda F_t + u_t, \tag{1.2.7}$$

$$F_t = \Gamma F_{t-1} + G\eta_t, \tag{1.2.8}$$

Using similar notation as before, $F_t = (f_t', \dots, f_{t-q}')'$, $\Lambda = [\Lambda_0, \Lambda_1, \dots \Lambda_q^f]$, $u_t = A(L)e_t$

and

$$\Gamma = \begin{bmatrix} \Gamma_1 & \Gamma_2 & \dots & \Gamma_q & \Gamma_{q-1} \\ I_r & 0 & ... & 0 & 0 \\ 0 & I_r & ... & 0 & 0 \\ ... & ... & ... & 0 & 0 \\ 0 & 0 & ... & I_r & 0 \end{bmatrix} \text{and G=} \begin{bmatrix} I_q \\ 0 \\ 0 \\ ... \\ 0 \end{bmatrix}.$$

Taking as given the number of lags and the number of factors and assuming $A(L)$ and Σ_u to be diagonal, we can proceed to the estimation of this model (1.2.7). Firstly, it is recommended that the variables are first standardized such that they have zero mean and variance one. Following Stock and Watson (2005), the dynamic form of the DFM can be estimated as follows:

1. Construct an initial estimate $\tilde{A}(L)$ of $A(L) = \text{diag}[a_1(L), \dots a_N(L)]$ for example by regressing the individual variables on their lags. (AR)

2. Compute the PC estimator $\hat{\Lambda}$, of Λ, from the model $\tilde{A}(L)\ X_t = \Lambda F_t + \tilde{u}_t$ where $\tilde{A}(L)X_t$ assumes the role of X_t in the equation of the static factor model and estimate the factors as $\hat{F}_t = \hat{\Lambda}'\tilde{A}(L)x_t$.

3. Estimate $A(L)X_t = \Lambda\hat{F}_t + \hat{u}_t$ by single-equation LS for each equation to update the estimate of A(L).

4. Iterate Steps 2 e 3.

Using a single-equation LS in step 3 is justified by the assumption that Σ_u is diagonal. If the assumption is correct, estimation efficiency can be improved by using a FGLS because the regressor in the different equations of the system are not identical.

Once the estimated factors $\hat{F}_t$ are available, the coefficient Γ matrix in the transistion equation and the underlying dynamic factors f_t in model (1.2.8) can be estimated as well. The estimation of Γ is complex because in practice we can only estimate some linear transformation of the static factors F_t. The four-step estimator of the DFM discussed above uses a statistical normalization that may not result in the primitive dynamic factors f_t. Thus, estimating Γ by regressing $\hat{F}_t$ on (F_{t-1}) produces an estimator Γ^* of some linear transformation of Γ. This issue may be addressed by implementing a second-stage estimator which determines the r linearly independent primitive factors f_t underlying F_t . Let $\hat{W}$ be the matrix of eigenvectors corresponding to the r largest eigenvalue of the residual covariance matrix $\hat{\Sigma}_\epsilon = T^{-1}\Sigma_t\hat{\epsilon_t}\hat{\epsilon_t}'$. Where, $\hat{\epsilon_t} = \hat{F_t} - \hat{\Gamma^*}\hat{F_{t-1}}$, and Γ^* is the estimator obtained by regressing $\hat{F}_t$ on $\hat{F}_{t-1}$. Then $\hat{\eta_t} = \hat{W}'\hat{\epsilon_t}$ and the primitive factors f_t can be estimated as $\hat{f_t} = \hat{W}'\hat{F_t}$. If estimates of $\Gamma_1, \ldots \Gamma_{q+1}$ are required, they may be obtained by regressing $\hat{f}_t$ on $\hat{f}_{t-1}, \ldots \hat{f}_{t-q-1}$. Finally, the covariance matrix of η_t can be estimated in the usual way using the residual covariance estimator based on the latter regression. It is also possible to estimate all DFM parameters simultaneously by ML estimation under the assumption that X_t is Gaussian. The Gaussian likelihood may be evaluated with the Kalman filter because the model is in a state space form. If we drop the assumption of Σ_u being diagonal, we end up in the GDFM (Generalized Dynamic Factor Models) for the estimation of this model look at Kilian and Lütkepohl (2017).
Anyway, this model is called the static form of the DFM because the relation between the observed X_t and the dynamic factors is contemporaneous. No lagged f_t appear in equation 1.2.4 it is implicit, but the factors are dynamic.

1.3 Determining the number of Factors in DFM

1.3.1 Determining the number of factors in Static form of DFM

Another problem encountered within factor models –that until now we took as given– is deciding how many factors are needed to explain our observable variables. According to the notation I used for static form of DFM, this means deciding k (number of factors) that explain our matrix X_t. In classical static factor models (the one a paragraph 2.1.1), criteria such as choosing as many factors as are necessary to explain a prespecified fraction of the overall variance are commonly used. More precisely, this means including additional common factors until the sum of the variances of the common factors exceeds a prespecified fraction of the sum of the eigenvalues of the sample covariance matrix.

Sometimes also a scree plot can be used. A scree plot displays the marginal contribution of the *kth* principal component to the average R^2 of the N regression of X_t against k principal components. This marginal contribution is the average additional explanatory value of the kth factor. When there are no missing data, the scree plot is a plot of the ordered eigenvalues of $\hat{\Sigma}_X$, normalized by the sum of the eigenvalues. Instead, Bai and Ng (2002), still for static form of DFMs, proposed using information criteria to select the optimal number of static factors k when N and T tend to infinity. Bai and Ng (2002) propose information criteria (IC) based on the quality of adjustment of the model to the data measures by the variance $V(j, F)$ such that:

$$V(j, F) = (NT)^{-1} \sum_{t=1}^{T} (X_t - \hat{\Lambda} F_t)'(X_t - \hat{\Lambda} F_t). \tag{1.3.1}$$

Where j is a given number of factors. Thus, if the number of factors j increases, the variance of the factors increases mechanically and the sum of the squares of the residuals decreases in turn. Bai and Ng (2002) suggest introducing a penalty function in the criterion to be optimized and propose three different Information Criteria (IC), corresponding to different penalty functions all based on the generic:

$$IC(r) = \log V_j(\hat{\Lambda}_t, \hat{F}) + rg(N, T). \tag{1.3.2}$$

Where $V_j(\hat{\Lambda}_t, \hat{F}_t)$ is the least-squares objective function evaluated at the PCs (Principal Components) $(\hat{\Lambda}_t, \hat{F}_t)$, and where g(N,T) is a penalty factor:

$$g(N, T) = \left[\frac{N + T}{NT} \right] \log[\min(N, T)].$$

such that as $N, T \to \infty$, $g(N, T) \to 0$ and $\min(N, T) g(N, T) \to \infty$. Bai and Ng (2002) provide conditions under which the value of r that minimizes an information criterion, with g(N,T) satisfying these conditions, is consistent for the true value of r. A commonly used penalty function among the three presented is the Bai and Ng (2002) penalty is:

$$IC_{p^2} = (R) = \log V(R) + R \left(\frac{N + T}{NT} \right) \log(\min(N, T)).$$

Using this criterion we can estimate the number of static factors, which determines the dimension of F_t. When N=T, this penalty simplifies to two times the BIC penalty, $T^{-1} \log(T)$.

1.3.2 Determine the number of factors in Dynamic for of DFMs

In the context of dynamic factor models, the number of dynamic shocks q can be determined using the Bai and Ng (2007) information criterion (IC). This criterion is obtained by considering the k estimated static factors as given and then estimating a VAR model of order p on these factors, where the order p is selected using the BIC criterion. Next, a spectral decomposition of the variance-covariance matrix of the estimated residuals of the VAR model, denoted Σ_ϵ of dimension $(k \times k)$, is calculated.[5]

Let $\rho_1 \geq \rho_2 \geq \rho_3 \ldots$ be the eigenvalues obtained from a PC analysis of the estimated residual covariance matrix of $\epsilon_t$and define

$$\hat{D}_1(k) = \left(\frac{\rho_{r+1}^2}{\sum_{i=1}^{R} \rho_i^2} \right)^{1/2}, \tag{1.3.3}$$

and

$$\hat{D}_2(k) = \left(\frac{\sum_{i=r+1}^{R} \rho_{i^2}}{\sum_{i=1}^{R} \rho_{i^2}} \right)^{1/2}. \tag{1.3.4}$$

Based on these quantities Bai and Ng (2007) propose to estimate the number of primitive dynamic factors as

$$\hat{k} = min \in r \Big| \hat{D}_1(r) < \frac{1}{min(N^{\frac{1}{2}-\delta}, T^{\frac{1}{2}-\delta})}, \tag{1.3.5}$$

Or:

$$\hat{k} = min \in r \Big| \hat{D}_2(r) < \frac{1}{min(N^{\frac{1}{2}-\delta}, T^{\frac{1}{2}-\delta})}. \tag{1.3.6}$$

Where δ is a small number between 0 and 1/2 . In a simulation study they used $\delta = \frac{1}{4}$. In practice, these different criteria are used at three stages:

(a) One of the Bai and Ng (2002) criteria is used to determine the optimal number of factors $k \in [1, r_{max}]$ in a static context.

(b) VAR(p) is estimated on these k estimated factors and the order p of the VAR is selected to minimize the BIC criterion.

(c) The Bai and Ng (2007) criteria are applied to the variance-covariance matrix or correlation matrix of the residuals ϵ_t of the VAR(p) to obtain the optimal number of dynamic factors q.

Until now I have briefly discussed dynamic factor models with their representation and estimation. Extentions of DFMs are VAR and FAVAR models that I use to address my research question. According to that I present the features of these two types of models in the next chapter.

[5]They utilize the fact that the error term in the transition equation in the static form (1.2.4) $G\eta_t$, has covariance matrix $G\Sigma_\eta G'$ of rank r and devise a procedure for determining that rank. Starting from estimates $\hat{F}_t$ of the static factors, they propose to fit a VAR model to the $\hat{F}_t$. In our current framework that VAR model is of order one because we have assumed qs. Thus, fitting $\hat{F}_t = \Gamma \hat{F_{t-1}} + \epsilon_t$.

Chapter 2

VAR and FAVAR Models

2.1 VAR Model

VAR are models used to capture the relationship between multiple quantities as they change over time. VAR is a type of stochastic process model. They generalize the single-variable autoregressive model (AR) by allowing for multivariate time series. More in particular, VAR models are characterized by their order (as AR), which refers to the number of earlier time periods the model will use. Usually, it is indicated as VAR of p order or VAR(p).VAR models are linked to DFM because as we can see from equation (1.2.8) or (1.2.4) those are VAR equations for factors, factors F_t are regressed on their lags. VAR stands for Vector Autoregressive. The generic equation of the model is:

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + e_t. \tag{2.1.1}$$

[6]

Where c is the intercept of the model, A_i is the time-invariant (kxk) matrix of coefficient and e_t is a k-vector of error terms. X_t is a vector of observed time-series. The error terms in the basic model satisfy three conditions:

- $E(e_t) = 0$
- $E(e_t e_t') = \Omega$ meaning that the variance-covariance matrix is a kxk positive-semidefinite matrix
- $E(e_t e_{t-k}') = 0$ for any non-zero k. There is no correlation across time. In particular, there is no serial correlation in individual error terms.

The process of choosing the maximum lag p in the VAR model requires special attention because inference is dependent on correctness of the selected lag order. VAR models were used to value, for

[6] If we adapt this equation to factor models we should have $F_t = c + A_1 F_{t-1} + A_2 F_{t-2} + ... + A_p F_{t-p} + e_t$ where $F_t = [f_{1t}, ..f_{kt}]$ and k is the number of factors

example, shocks on inflation Sims (1993), on prices Sims (1992) etc. Actually Sims (1992) showed the limits of VAR model because in his study he found out that after a restrictive monetary policy taken by the central bank, the impulse response for prices was initially positive that is counter intuitive if analyzed based on Keynesian macroeconomic approach. This result is the so called "price puzzles". Mainly this problem is due to the fact that VAR model has a huge, omitted variables bias. VAR is good if there are few factors to consider otherwise there could be problems of missing some of the important factors that instead could explain a bigger slice of variance-covariance matrix of the model. Another problem in VAR were restrictions needed to identify the parameters. To solve these issues, based on Geweke (1977) thesis, dynamic factor analysis seemed to help VAR models. Even if the number of common shocks is small in DFMs, as we have seen above, it still gives better results compared to a conventional VAR analysis with a small or moderate number of variables – VAR would fail to span the space of the structural shocks to the dynamic factors –. The new methods for estimating and analyzing dynamic factor models, combined with the empirical evidence that perhaps only a few dynamic factors are needed to explain the comovement of macroeconomic variables, has motivated research on how to best integrate factor methods into VAR and SVAR analysis (as we will see in the following paragraph FAVAR model of BBE is an example).

2.2 FAVAR Model

Bernanke et al. (2005) to overcome the limits of VAR model thought about a factor augmented vector autoregressive model (FAVAR). They have been the first to present and apply this model in their paper in which they value the impact of monetary policy shocks on three observed variables (Y_s) ie. inflation, interest rates and GDP so as on a background panel of 120 economic variables, from which they extract the factor of interests. Their aim of their paper is to augment the VAR model with these extracted factors, in order to summarize a wider economic concept so to overcome the omitted variable bias of VAR. In the following paragraph I explain in more detail how FAVAR work and I also provide reference to FAVAR estimation's method by Bai et al. (2016), Bai (2003).

Generally speaking, one way to obtain FAVAR model is to augment the set of observed variables Y_t in a given VAR model by unobserved or latent factors, F_t extracted from a panel of observed variables X_t that does not include Y_t. According to the representation of Bernanke et al. (2005), consider an observed variable Y_t of dimensions N x 1, and still F_t be a Kx1 vector of unobserved factors, where K is "small". Then the FAVAR model, assuming now the joint dynamics of (F_t, Y_t) is given by:

$$\begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \Phi(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + v_t. \tag{2.2.1}$$

Where $\Phi(L)$ is a conformable lag polynomial of finite order d, which may contain a priori restrictions as in the structural VAR literature Amisano and Giannini (2012). The error term v_t is mean zero with covariance matrix Q. This equation cannot be estimated directly because the factors F_t are unobservable. However, if we interpret the factors as we said before ie. described by many economic or climatic variables, we can infer something about the factors from the observation of a variety of economic and climatic time series contained in another panel. For concreteness, suppose that we have a number of background panel time series denoted by M x 1 vector X_t that represent this variety of economic and climatic time series. More precisely in my analysis, X_t is further divided in two parts: X_1 and X_2. X_1 is the economic panel composed by 18 variables, and X_2 is the climatic one composed by 11 variables. M is the number of time series in the panel X_t and M is "large" in particular, it is assumed to be much greater than the number of factors $(K + N << M)$[7]. We assume that the informational time series in X_t are related to the unobservable factors F_t and the observable variables Y_t[8] by:

$$X_t' = \Lambda^f F_t' + \Lambda^y Y_t' + e_t'. \tag{2.2.2}$$

This is a static form of a Dynamic Factor Model because X_t depends on the current and not lagged values of the factors. Where Λ^f is an N x K matrix of factor loadings and Λ^y is a NxM and the Nx1 e_t error terms are mean zero and will be assumed either weakly correlated or uncorrelated, depending on whether estimation is by PC or likelihood methods. This equation captures the idea that both Y_t and F_t, which in general can be correlated, represent pervasive forces that drive the common dynamics of X_t. Conditional on the Y_t, the X_t are thus noisy measures of the underlying unobserved factors F_t. The implication of equation (2.2.2) that X_t depends only on the current and not lagged values of the factors is not restrictive in practice, as F_t, can be interprets as including arbitrary lags of the fundamental factors, thus as a dynamic factor model.

Here we can see that the part concerning the factor specification is closely linked to DFMs as factors are estimated in the same way. Nonetheless the extracted factors are combined in a VAR. FAVAR, then is an extention of both VAR and DFMs. This approach is often intended to address the informational deficiencies of conventional small-scale VAR models because FAVAR allows to approximate the responses of many more variables to a given structural shock that is not possible in conventional VAR. Specifically with FAVAR it is possible to compute the impulse response for the panel X_t and for the Y_t not only for the Y_t as in VAR. In my research question, as stated in the introduction, I want to address the impact that climate has on a variety of economic variables that cannot be included all in a VAR model. Moreover, climate is represented by different variables so using just one factor to summarize this concept makes it easier to conduct the analysis so FAVAR could be a good model for the analysis.

[7] Just a reminder N is the number of observed time series in Y_t

[8] I use this notation to unify notation to the one used in the previous paragraph and in Chapter I for DFMs

2.2.1 Estimation of FAVAR model PCA approach

To estimate the FAVAR model I will present, as Bernanke et al. (2005) did in their paper, the two-step approach (Stock and Watson approach) and the Bayesian approach applied through Gibbs sampling used in the MATLAB code. Starting with the first approach, it is based on a two-step principal components methodology, which provides a non-parametric way of uncovering the space spanned by the common components $C_t = (F_t, Y_t)'$. Common components indicated as C_t, are called like this because are the regressors for each equation $x_i = \lambda^f f_{it} + \lambda^y y_{it} + e_t$ in system (1.2.1) as $X_t = [x_{1t}...x_{it}]$, $F_t = [f_{1t}...f_{it}]$ and $Y_t = [y_{1t}...y_{it}]$. In the first step, the common components, C_t, are estimated using the first $K + N$ (where K is the number of factors of F_t, instead N is the number of observed variables Y_t) principal components of X_t as to obtain $\hat{F}_t$. Notice that even if Y_t is observed, so we do not have to estimate it, in the first step the PCA is conducted by using the true number of factors ie. the number of observed plus the number of unobserved factors. $\hat{F}_t$ is then obtained as the part of the space covered by $\hat{C}_t$ that is not covered by Y_t. In the second step, the FAVAR equation (2.2.1) is estimated by standard methods ie. OLS, with of course F_t replaced by $\hat{F}_t$. This procedure has the advantages of being computationally simple to implement and as discussed by Stock and Watson (2002), it also imposes few distributional assumptions and allows for some degree of cross correlation in the idiosyncratic error term e_t. However, this two-step approach implies the presence of "generated regressors" in the second step, because we do use factors estimated by PCA to estimate the coefficients. To end the estimation and obtain accurate confidence intervals on the impulse response function a bootstrap method implemented by Kilian (1998) or Berkowitz and Kilian (2000) is used.

2.2.2 Identification of the model

As we have said for DFMs, it is also necessary for FAVAR model, to decide the restrictions to identify the factors and the associated loadings. We can use the same restriction pointed out in Chapter I, proposed by Bai and Wang (2014) for DFMs as the factor equation is in DFM form. In two-step estimation by principal components, the factors are obtained entirely from the observation equation (2.2.1). In this case we can choose either to restrict loadings by $\Lambda^{f\prime}\Lambda^f/N = I$ or restrict the factors by $F'F/T = I$. Either approach delivers the same common component $F\Lambda^{f\prime}$ and the same factor space. Here we impose the factor restriction, obtaining $\hat{F}_t = \sqrt{T}\hat{Z}$, where the $\hat{Z}$ are the eigenvectors corresponding to the K largest eigenvalues of XX' sorted in descending order. This approach identifies the factors against any rotations. The estimated system ((2.2.1)) -((2.2.2)) can be used to draw out the dynamic responses of not only the "main" variables Y_t but of any series contained in X_t. Hence the "reasonableness" of a particular identification can be checked against the behavior of many variables, not just 3 or 4. Second, one might also consider constructing the impulse response functions of factors (or linear combinations of the factors) that can be shown to stand in for a broad concept.

Chapter 3

Bayesian Estimation of FAVAR model

3.1 Bayesian notation

Another way to estimate FAVAR model, is to use Bayesian techniques. The upgrade in using these techniques is that a more precise estimate of the factors can be obtained. In a nutshell, using Bayesian techniques allow the state vector of factors, $F_t = [F_{1t}, ..F_{kt}]$, to be updated as soon as new information is available. I now explain in more detail this statement. First of all, Bayesian statistics is based on a few concepts that can be expressed analytically. Differently to frequentist analysis, where the objective is to try to infer the parameters θ (mean, variance etc.) of an observed distribution of data by actually making assumption on the distribution of the parameters, bayesian statistics focuses on finding the distribution of the vector of parameters. According to that, Bayesian statistics is based on two concepts: prior and posterior distribution. Bayesian analysis allows the researcher to incorporate his prior beliefs about the parameters vector θ, these will constitute the prior distribution. These beliefs and so priors, are based on past experience. For past experience it is meant information the researcher has about the parameters or experience he has developed by studying other similar datasets etc. The key point is that these prior beliefs can be expressed with a probability distribution. Suppose we have a set of data $y = (y_1', \dots y_T')$, Bayesian treats the data as given and the parameter of interest θ, as unknown. Firstly, without looking at the distribution of the data, they lay down a probability distribution (priors). After stating that, the researcher collects data on y and writes down the likelihood function of the data. This phase is identical to the one done by classical econometrician. The last phase is that the Bayesian researcher, conditioned to the information on the data he has now observed, develops the posterior distribution. This distribution is proportional to the prior distribution times the likelihood function.

$$g(\theta|y) \propto L(y|\theta)P(\theta), \tag{3.1.1}$$

Where, $P(\theta)$ is the pdf of the prior beliefs of the researcher, L$(y|\theta)$ is the Loglikelihood and $g(\theta|y)$ is the posterior distribution. According to Bayesian approach after we reach the posterior distribution we might be interested in the moments of this distribution. This is the part related to point estimation. Often moments of the posterior distribution are of interest such as $E(\theta|y)$. The posterior mean is often used as a point estimator for θ if the posterior distribution is Gaussian or symmetric. Otherwise we can use the mode or median of the distribution. Broadly speaking the researcher is often interested in expected value of functions of θ. If we are interested in figuring out a point estimate of some function $h(\theta)$ that may be vector-valued or a scalar, we minimize the expected loss of $h(\theta)$ based on some loss function. Denoting this loss function by $L(h^*, h(\theta))$, the point estimate, say $\hat{h}$ is chosen such that:

$$\hat{h}^* = argmin_{h^*} \int L(h^*, h(\theta))g(\theta|y)d\theta. \tag{3.1.2}$$

Where h^* denotes an element of the range of $h(\theta)$. For example, if the loss function is quadratic and the first two moments of the posterior distribution exist, the point estimate corresponds to the posterior mean. Changing the loss function if it is no more quadratic for example, then we will get a different estimator that could be the mode or median.

3.1.1 Simulating the posterior distribution

Bayesian inference as we have said till now, is based on finding the posterior distribution. In non-trivial cases, the closed form for the posterior distribution is not available. This means that we do not know the shape of our ending distribution. To get it, it is necessary to use numerical methods to simulate. Usually, the objective is to generate random draws for the parameter vector $\theta^1, \dots \theta^n$ from the posterior distribution of the parameters or some function of these parameters such as the impulse response associated with a VAR model. Suppose we are interested in the expectation of $h(\theta)$ which may be a vector or a scalar function. Laws of large numbers suggest that we can obtain a good approximation to this quantity by drawing at random $\theta^1, \dots \theta^n$ from the posterior distribution and noting that

$$\frac{1}{n}\sum_{i=1}^{n} h(\theta^i|y) \rightarrow^{a.s} E(h(\theta|y)) = \int h(\theta|y)g(\theta|y)d\theta. \tag{3.1.3}$$

Where a.s stands for almost surely convergence. A sequence of random variables is said to converge almost surely when, $X_n\ (w) \rightarrow X(w)$ not on ω as for pointwise convergence, but on an event B with probability P(B)=1. Where omega is the support. Both the bounded convergence theorem and the monotone convergence holds in almost surely convergence. The approximation precision increases with the number n of random draws. Of course, simulating the posterior distribution when it is not known or nonstandard is not easy.

3.1.2 Markow Chain Monte Carlo (MCMC)

The usual method to start sampling is Monte Carlo simulation, but Monte Carlo simulation can only be used if we can sample from a distribution in a i.i.d fashion. Meaning we are dealing with a not so difficult distribution or a distribution in closed form. If we cannot do this, meaning we do not have a well-defined distribution, for the posterior one in this case, we have to change our way to estimate and use something more complex that is the MCMC. MCMC stands for Markov Chain Monte Carlo, and it constructs a sequence of correlated samples let's say $Y_1, Y_2, \ldots$ that explore the region of high probability by making a sequence of incremental movements. Even though the samples are not independent, it turns out that under very general conditions, sample averages $\frac{1}{N}\sum_{i=1}^{N} h(X_i)$ can be used to approximate expectations $E[h(X)]$ just as in the case of simple Monte Carlo approximation and by ergodic theorem, these approximations are guaranteed to converge to the true value. Suppose we are interest in knowing the expected value of a certain function $f(x)$ with respect to a certain distribution π.

$$\mu = E_\pi[f(x)]. \tag{3.1.4}$$

This is usually expressed as:

$$\int f(x)\pi(x)dx. \tag{3.1.5}$$

Supposing that we cannot derive this integral analytically nor sampling in a i.i.d fashion from π, then we have to use another method to estimate the integral that is MCMC. To use MCMC we need to ensure that π is a unique, limiting and stationary distribution and also the Markov Chain must be irreducible such that the ergodic system holds.[7]

For stationary distribution it is meant that taken any non-singular matrix T if multiplied by the probability distribution function it gives back the same pdf. For irreducibility it is meant that for any pair a,b there is some π such that $P(X_t = b \big| X_0 = a) > 0$ meaning that starting from any kind of value at time 0 we arrive at point b with a probability greater than 0. After building as such a distribution we can now start sampling with Monte Carlo. We will then obtain a parameter estimate $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$ To check whether this estimate is good for μ, we have to use an indicator

$$V(f,P) = \lim_{n\to\infty}[nVar_\pi\hat{\mu}_n] = \sigma^2 \sum_{k=-\infty}^{\infty} \rho_k. \tag{3.1.6}$$

That is the variance basically. $\sigma^2 = Var_\pi f(x)$ and $\rho_k = Cov_\pi[\frac{F(X_0, f(X_k)}{\sigma^2}]$. Basically the "pity" of using the MCMC methods instead of the classic Monte Carlo is that we have a higher variance

[7] A Markov chain is said to be ergodic if there exists a positive integer T_0 such that for all pairs of states i,j in the Markov chain, if it is started at time 0 in state I then for all $t > T_0$, the probability of being in state j at time t is greater than 0.

in our estimate because of the term ρ_k. This is within the functioning of the Markov Chain, that estimates correlated samples. The draws of the parameter vector are not independent, but serially dependent -we pay more variance-. Laws of large numbers and central limit theorems for dependent samples can be used to justify the use of such samples. Thus, longer samples may be necessary to have a more precise inference. This is the price we pay to have the impossibility of sampling in a i.i.d fashion from the distribution. If the chain of MC runs long enough, the distribution converges to the posterior distribution of the parameters. In fact, for the construction of approximately independent samples from the joint posteriors, econometricians may want to work only with every m^{th} sampled vector, where m is a sufficiently large number (to reduce correlation among series). Moreover, a large number of initial sample values are usually discarded (burns-in) to ensure a close approximation of the posterior because we go faster to the support of the posterior distribution. There are basically two ways of simulating MCMC method. One is the Metropolis-Hastings algorithm. Practically, a draw is accepted with probability one if it increases the posterior. Otherwise, it is accepted with a probability less than 1, the precise value of which depends on how much lower the current posterior value is compared with the previous draw. The other way is the Gibbs sampler which will be used in the MATLAB application.

3.1.3 Gibbs Sampler

The Gibbs sampler is within the MCMC methods more in particular it can be adopted if the parameter vector can be partitioned. This can be done to simplify the estimation of the posterior distribution because sometimes computing the integral for the posterior distribution can be a difficult process. If the conditional posterior of one of the sub-vectors, given the remaining elements, has a known, conventional distribution from which it can be possible to start sampling, it can lead to a simpler drawn procedure. To illustrate the Gibbs sampling, consider the simplest case where θ our vector of parameters of interest, can be partitioned as

$$\theta = (\theta_1', \theta_2')' \tag{3.1.7}$$

such that $g(\theta_1 \big| \theta_2, y)$ and $g(\theta_2 \big| \theta_1, y)$ correspond to known distributions that can be simulated easily. The Gibbs sampler happens when the random variable we observe, let's say $y = (y_1, \dots y_d)$, lives in a d-dimentional space and rather than update all coordinates together we update one coordinate at a time and then cycle those coordinates i.e. first we update y_1 then y_2 then y_3 etc. given the other information we estimate $\theta_{1,1} \big| \theta_{2,0}$ then $\theta_{2,1} \big| \theta_{1,1}$ etc. and then we go back to the first one till convergence is reached. This is called the fixed scan Gibbs sampler. There is also the random scan Gibbs sampler in which the cycle among the coordinates does not follow an ordinate method but rather is random meaning that we can associate some probability to the coordinates we want to update i.e let's say we want to update coordinate y_1 more often than coordinate y_2 because maybe it is more important, then we can attribute a greater probability to that coordinate to be updated, so that it will be sampled more often than the others. If we opt for random Gibbs sampler, we can find ourselves in the situation in which we do not know which of the coordinates is better to re-sample more, so maybe at the beginning can be convenient to do a fixed scan Gibbs sampler, learn what are the most important coordinates and then try and do a random Gibbs sampler. The functioning of the Gibbs sampler is always the same, though. We update the ith coordinate by taking its proposal distribution and we assume that it is the full conditional distribution for that coordinate. The full conditional distribution is equal to the marginal target distribution conditioned on all the other coordinates:

$$q(y_i,) = \pi(y_i \big| y_{-i}) = \text{full conditionals.} \tag{3.1.8}$$

Where -i indicates $j : j \neq i$. A Gibbs sampler has the advantage that these full conditionals always lead to probability of one (because divided by the "normalized constant"), so there is no need to calibrate the proposal, but on the other hand those full conditionals are needed to be computed and they need to be sampled on a i.i.d fashion. If we cannot derive the full conditional distribution we might use a Metropolis-Hastings [9] function within Gibbs for those distributions that we cannot derive a prior. Meaning we will derive the full conditionals of the distributions we can derive and for the other we could use Metropolis-Hastings. Anyway, to explain better the Gibbs concept, it can be useful to present an easy example. Suppose we have a single observation $y = (y_1, y_2)$ from a bivariate normal distribution of mean unknown $\theta = (\theta_1, \theta_2)$ and variance unknown

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \tag{3.1.9}$$

Suppose we do stabilize a prior distribution, the uniform distribution on θ, so the posterior distribution obtained as we said in the paragraph above, proportional to the prior distribution times the likelihood function, is a normal distribution:

$$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} | y \sim N\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \tag{3.1.10}$$

In this example we have a closed form for our target distribution, meaning that we already know where we will arrive (posterior distribution), and we do know everything about this distribution. To check if the algorithm works, though we can simulate a Gibbs sampler and check, by starting from different

[9] The Metropolis-Hastings Algorithm can be used as an alternative when none of the other way to simulate could work. Basically, it works that given a probability acceptance level given by: $\alpha = min[1, \frac{\Pi(y)q(x,y)}{\Pi(x)q(x,y)}]$ and a proposal distribution (q in our case), that is a distribution for the posterior given my prior, then, by starting with different x_0 we will accept the distribution generated by our xith value, if the probability is equal to α otherwise $x_i = x_{i-1}$. More formally, Metropolis-Hastings is based on a conditional density, $\eta(\theta|\theta^{i-1})$. For a given θ^{i-1}, a candidate θ^+ is drawn from the distribution corresponding to $\eta(\theta|\theta^{i-1})$, and $\theta^i = \theta^+$ is chosen with probability: $min[\frac{\frac{(g(\theta^+|y)}{\eta(\theta^+|\theta^{i-1})}}{\frac{(g(\theta^{i-1}|y)}{\eta(\theta^{i-1}|\theta^+)}}, 1]$ otherwise $\theta^i = \theta^{i-1}$.

values of $\theta_0 = initialvalues$ if, by simulating the posterior distribution will be stable at the support and reach expected value of θ which is the mean of the true distribution. To set up the Gibbs sampling we have to derive the full conditionals distribution. In this case as we are dealing with a bivariate Gaussian, we do know that also the marginal distribution will be Gaussian and so as the conditionals. We obtain:

$$\theta_1 \big| \theta_2, y \sim N(y_1 + \rho\left(\theta_2 - y_2), 1 - \rho^2\right), \tag{3.1.11}$$

$$\theta_2 \big| \theta_1, y \sim N\left(y_2 + \rho(\theta_1 - y_1), 1 - \rho^2\right). \tag{3.1.12}$$

As we said before, Gibbs sampler works that we do estimate one parameter at a time keeping the others as given. We will first estimate θ_1 given θ_2 and the data (which will be the last value of θ_2) then we will estimate θ_2, given the data and the value just estimated for θ_1 and start over for N times which is the number of simulations we want to run. To decide the number of times we do have to iterate our cycle, there is a convergence criterion developed by Rosenthal et al. (2001) that computes the "magic number" 0.234 which is a value that the convergence ratio should reach to have a pretty close distribution to the true one. For convergence ratio it is meant $\frac{\pi(y)}{\pi(x)}$ which is the distribution at which we are trying to go ($\pi(y)$) divided by the distribution at which "we are" with the current number of simulation $\pi(x)$. If this ratio is close to 0.234 it means that our posterior is good. Going back to our example, if we suppose that the observation is $y = (0,0)$ and $\rho = 0,8$ for frequentist statistics the maximum likelihood estimator for θ would have been the mean of the distributions and as in this example we do have only one observation so it would end up 0 for θ_1 and 0 for θ_2. For Bayesian approach the estimator of theta would have been the expected value of the posterior distribution and in this special case of Gaussian, it is also 0 for θ_1 and 0 for θ_2. So, in this case the two statistics branches (Frequentist and Bayesian) coincide.

This easy example summarizes the logic behind the Gibbs sample, but to apply the Gibbs sample on a more general situation we need to generalize the formula. Suppose we do not have only one observation, but we do have a sample of observed variable $Y_1, \ldots Y_d$ (as we will have in the case study I will present later) still distributed as a Gaussian distribution, in this case suppose we do not know nor the mean nor the variance matrix. Then we do have to establish some priors both on the mean and on the variance coefficient. Basically, the mean and the variance-covariance matrix will represent our parameters of interest. We can say that the prior distribution for the mean parameter is Gaussian instead the prior distribution for the variance is a Gamma (because the variance-covariance matrix is Positive Definite). So, we have our observed sample:

$Y_1, \ldots Y_d$ iid $N(\mu, \sigma^2 = \gamma^{-1})$ Priors:

$$\mu \sim N(\mu_*, \sigma_*^2)\tau = \frac{1}{\sigma^2} \sim G(\alpha_*, \beta_*). \tag{3.1.13}$$

Posteriors:

$$p(\mu, \tau \big| y_1, \ldots, y_d) L(\mu, \tau; y) p(\mu, \tau) \propto \tau^{(\frac{d}{2} + \alpha_* - 1)} e^{\beta_* \tau} e^{\frac{-(\tau S_d)}{2} - \frac{(\mu - \mu_*)^2}{(2\sigma_*^2)}}. \tag{3.1.14}$$

here we have a layered Gibbs sampling because we do want to find the posterior distribution of y but the parameters of y have their own distributions, so α and β, μ_* and σ^2 are defined "hyperparameters" $S_d = \Sigma_{i=1}^{d}(y_i - \mu)^2$ the posteriors is proportional to the likelihood times the priors.

To start sample we have to define the full conditionals:

$$(\mu|\gamma, y_1, \ldots, y_d) \sim N\left(\frac{(d\bar{y}\tau + \mu_*\tau_*)}{(d\tau + \tau_*)}, (d\tau + \tau_*)^{-1}\right), \tag{3.1.15}$$

$$(\tau|\mu, y_1, \ldots, y_d) \sim G\left(\alpha_* + \frac{d}{2}, \beta_* + \frac{S_d}{2}\right). \tag{3.1.16}$$

These two conditionals have been computed from the inverse formula of Bayes by dividing the joint distribution by the marginal distribution of each of the parameters respectively that is quite a difficult iter to follow. Actually, in this case is still possible to compute, but there are other cases in which this procedure can lead to unfeasible integrals and cannot be computed. Defined the full conditionals we have to iterate the algorithm and the cycle is always the same. We start at certain initial values μ_0 and τ_0 then we find, given τ_0, μ_1, then given μ_1 we find τ_1 etc. etc. for N times. To thin the sampling scheme and not use all the storage when running the algorithm, it could be useful to retain only fewer results from the MCMC simulation. To decide the "step" to use to thin the results (and save every K times) we have to refer to the ACF graph. Basically, we will plot the ACF of the distribution of the parameters and take the worst ACF (meaning the highest lag after which the distribution goes down to zero) and use it as a lag. For example, if in the ACF the lag after which the series stabilize is 15, we will take a step equal to 15, meaning that we will save results as: the first one, then the 15th, then the 30th and so on.

3.1.4 Gibbs sampling for state-space model

To actually apply Bayesian methods to my FAVAR model later on, it is needed to understand how Gibbs sampling changes and works within state-space models. For illustration purpose suppose we are dealing with this model state-space model:

$$Y_t = H\beta_t + Az_t + e_t, \tag{3.1.17}$$

$$Var(e_t) = R, \tag{3.1.18}$$

$$\beta_t = \mu + F\beta_{t-1} + v_t, \tag{3.1.19}$$

$$Var(v_t) = Q. \tag{3.1.20}$$

The first equation is the observation equation, the second equation is instead the transition equation. In the observation equation the unknown parameters are: the elements of H that are non-zero or not given in the data, A and the non-zero elements of R. In the transition equation instead, the unknown parameters are: μ, F and Q. Moreover also β_t is an unknown state variable. The fact that β_t is also unknown, would have make the system more complex, but if we treat β_t as observed, the observation and the transition equation go down to two linear regressions. Having said that the general Gibbs procedure for this model is:

(a) Conditional on β_t, we estimate H, R from the posterior distribution (for the observation equation)

(b) Conditional on β_t, we estimate μ, F and Q for the posterior distribution

(c) Conditional on the parameters of the state-space: H, R, μ, F and Q sample the state variable β_t from its conditional posterior distribution.

(d) Iterate step 1 to 3 until convergence.

The new step for this kind of Gibbs sampling, is Step 3 because we need to sample from a state variable posterior distribution. First of all, it is necessary to estimate $beta_t$ and then infer its distribution.

3.1.5 β_t Posterior Distribution

Defining the time series of β:

$$B_t = [\beta_1 \beta_2 \dots \beta_t]', \tag{3.1.21}$$

So that

$$B_{T-1} = [\beta_1 \beta_2 \dots \beta_{T-1}]'. \tag{3.1.22}$$

We use multi-move Gibbs sampling that basically allow us to simulate the whole vector of betas all together. Multi-move Gibbs sampling comes from the theory behind Gibbs sampling in Markow switching models[10] in which we need to estimate the whole vector of latent variable. The main difference between Single-move Gibbs sampling (that is the classic one) and multi-move is that:

- Single-move Gibbs sampling: generate each β_t
  from: $f(\beta_t \big| (\tilde{\beta}_{\neq t}), H, Q, R, F, \mu, Y_t)$
- Multi-move Gibbs Sampling: generate the whole block of B_t
  from: $f(B_t \big| H, Q, R, F, \mu, Y_t)$

We are interested in knowing the posterior distribution $p(B_t \big| H, Q, R, F, \mu, Y_t)$ that we can shortly indicate by $p(B_T|\Phi_T)$ so the joint distribution of all the betas. Φ_T represent all the available information at time T. The joint probability distribution $p(B_T|\Phi_T)$, can be split into a sequence of conditional distributions:

[10] Di Persio and Frigo (2016)

$$p(B_T \big| \Phi_T) = p(B_T|\Phi_T)p(B_{T-1}\big|B_T, \Phi_T) =$$
$$p(\mathrm{B}_T\big|\Phi_T)p(\beta_{T-1}\big|\beta_T, \Phi_T)p(B_{T-2}\big|\beta_{T-1}, \Phi_T).$$

As shown in Kim and Nelson (1999), this expression can be simplified since β_T follow a first order AR or Markov process. Thanks to that, given Φ_T and β_{T-1}, in the term $p(B_{T-2}|\beta_T, \beta_{T-1}, \Phi_T), \beta_T$ contains no additional information not contained in β_{T-1} so, this term can be re-written as $p(B_{T-2}|_{T-1}, \Phi_T)$ so in general the posterior distribution can be summed up as:

$$p(B_T|\Phi_T) = p(B_T|\Phi_T)\Pi_{t=1}^{T-1}p(\beta_t|\beta_{t+1}, \Phi_t).$$

So, we do not need to know information far more than one period ahead because we are dealing with a Markov Chain. This is typical of state space, because as we are in that state space it will give us all the information we need to conduct our estimates. Practically it means that if we go ahead in time like taking b_{100} to estimate b_{99}, b_{100} is a number extracted from the distribution found with Gibbs sampling. So, all we do is, condition on that number we work out the probability distribution and the available information at one time ahead. This methodology is the Carter and Kohn (1994). Assuming that the disturbance errors of the state-space model are normally distributed we have:

$$B_{T|\Phi_T} \sim N(\beta_{T|T}, P_{T|T}),$$
$$\mathrm{B}_{t|\Phi_t,\beta_{t+1}} \sim N(\beta_{t|t,\beta_{t+1}}, P_{t|t,\beta_{t+1}}).$$

Where:

$$\beta_{t|t,\beta_{t+1}} = E[\beta_t|\Phi_t, \beta_{t+1}] = E[\beta_t|\beta_{t|t}, \beta_{t+1}].$$
$$P_{t|t,\beta_{t+1}} = cov[\beta_t\Phi_t, \beta_{t+1}] = cov[\beta_t\big|\beta_{t|t}, \beta_{t+1}].$$

Now this representation of the problem that we have said till now, is based on the Carter and Kohn (1994) algorithm that derive an appropriate recursion that, updates the estimates conditioned on some known value of β_{t+1} that is the next value of β_t previously estiamted.

$$\beta_{t|t,\beta_{t+1}} = \beta_{t|t} - K_{t|t+1}(\beta_{t+1} - F\beta_{t|t} - \mu). \tag{3.1.23}$$

Where K is the "Kalman gain". The Carter and Kohn algorithm relies heavily on the Kalman smoother definition, and the values computed with the Kalman Filter. Before moving forward with the explanation of the Carter Kohn algorithm, then it is necessary to point out what it is meant with Kalman filter and Kalman smoother. The Kalman Filter is a recursive algorithm, that provides an estimate of the state variable at each point in time, given the information up to that time period. Basically, the Kalman Filter allow us to estimate the entire state space starting by few initial values. To use the

Kalman filter though, we need estimates on the other parameters of the model, that in our case are H, R, μ, F and Q. These estimates are available from the Gibbs sampling from previous draw of the Gibbs sampler. The Kalman filter can be constructed with the following equations:

$$\begin{aligned}
\beta_{t|t-1} &= \mu + F\beta_{t-1|t-1}, \\
\mathrm{P}_{t|t-1} &= FP_{t-1|t-1}F' + Q, \\
\eta_{t|t-1} &= Y_t - H\beta_{t\beta|t-1} - Az_t, \\
\mathrm{F}_{t|t-1} &= HP_{t|t-1}H' + R, \\
\beta_{t|t} &= \beta_{t|t-1} + K\eta_{t|t-1}, \\
\mathrm{P}_{t|t} &= P_{t|t-1} - KHP_{t|t-1}.
\end{aligned}$$

Where $K = P_{t|t-1}H'F_{t|t-1}^{-1}$. The first and the second equation are the "prediction" equation meaning that they are a first preliminary estimate of those two quantities. The first equation finds $\beta_{t|t-1}$ meaning the beta one period ahead, having information up to time t-1 using the transition equation. The equation in fact, can be found by taking the expected value of the transition equation. The second equation of the filter instead finds the variance $P_{t|t-1}$ given the information at time $t-1$ and can be found by taking the variance of β_t. The third and the fourth equations instead consist of the prediction error and the variance of the prediction error, these two quantities are needed to update the estimate of the betas and the variance matrix in the last two equations that are the updating equations. Basically, the prediction error is the error that comes from the estimates of beta one period ahead using the information up to one period before. The parameter K is the so called "Kalman gain" and it is the weight associated to the prediction error. That means that, if the information contained in the prediction error is not so impactful for the betas already estimated, then the weight will be low. So, the prediction error does not bring too much new information for the already estimated beta. Now, the problem of the Kalman Filter is that the estimate of the parameter only happens with the information up to time $t-1$. So, when we are estimating we are not considering the information generated after. In practice, this means that, if we want to estimate β_{t+1} we are using information up to time t+1 so in the estimate of β_{t+1} we use all the available information, but the estimate of β_t we do not because, β_t has been estimated by only using information up to time t. To solve this "issue" the estimates could be updated by "smoothing them". Smoothing means use all the information happened before and after time t to get a more precise estimate of β_t. This means that there will be future values $\eta_{i\big|t-1}$ that will bring some news, that can help estimate more correctly β_t. So, it becomes: $\beta_{t|T} = E(\beta_t \big| \Phi_{t-1}, \ldots \eta_{t|t-1}, \eta_{T\big|t-1})$ Where $T > t$ and Φ_{t-1} represent all the available information ad time t-1. To get a quicker way of updating the value of beta using and assuming that the future innovations are uncorrelated, we can say that:

$$\beta_{t|T} = \beta_{t|t} + \sum_{j=t}^{T} \Sigma_{\beta_t \eta_j} \Sigma_{jj}^{-1} \eta_{j|t-1}. \tag{3.1.24}$$

That means that the estimate of $\beta_{t|T}$ (so given the future information T) is equal to the value of beta estimated with information up to time t plus a prediction error given by the "news" brought by the future innovations. Recalling that:

$$\beta_{t|t} = E(\beta_{t|t}|\Phi_{t-1}, \eta_{t|t-1}). \tag{3.1.25}$$

This is called fixed interval smoother, because we are using always the same set of information T and it is used to estimate all β_t. Now, we do have another set of information, though, so we can say, we do have an estimate of $\beta_{T\big|T}$ using all the information, how can I let the smoother work better? We do have a new prediction error that answer to the question: what would have been the value of $\beta_{T|T-1}$ meaning the value at the next time given the information up to a period before? The prediction error of this estimate is then equal to:

$$\zeta_{T|T-1} = \beta_{T|T} - F\beta_{T-1|T-1} - \mu. \tag{3.1.26}$$

That is the difference between $\beta_{T|T}$ estimated with all the information minus F$\beta_{T-1|T-1}$ that is the parameter estimated at time $T-1$ with the information up to time $T-1$. Finding these residuals, allow us to use them to update the estimate of $\beta_{T-1|T}$ so we do have that:

$$\beta_{T-1|T} = \beta_{T-1|T-1} + \sigma_{\beta\zeta}\Sigma_{\zeta\zeta}^{-1}\zeta_{T|T-1}. \tag{3.1.27}$$

Where:

$$\Sigma_{\zeta\zeta} = var[\zeta_T \big| \beta_{T-1}],$$

$$\Sigma_{\beta\zeta} = cov[\beta_{T-1}, \zeta_T|\Phi_{T-1}],$$

$$\Sigma_{\zeta\zeta} = var(\beta_T - F\beta_{T-1|T-1} - \mu) =$$
$$\text{var}(\text{F}(\beta_{T-1} - \beta_{T-1|T-1}) + e_t) = FP_{T-1|T-1}F' + Q.$$

$$\Sigma_{\beta\zeta} = E[(\beta_{T-1} - \beta_{T-1|T-1})(\beta_T - F\beta_{T-1|T-1} - \mu)'] = E[(\beta_{T-1} - \beta_{T-1|T-1})(\beta_T - \beta_{T-1|T-1})'F'] =$$
$$P_{T-1|T-1}F'.$$

So, plugging this information in the update equation of beta we do obtain that the update is the result of the Kalman Filter plus the state variances

$$\beta_{T-1|T} = \beta_{T-1|T-1} + P_{T-1|T-1}F'P_{T|T-1}^{-1}(\beta_{T|T} - F\beta_{T-1|T-1} - \mu). \tag{3.1.28}$$

Applying it recursively gives:

$$\beta_{t|T} = \beta_{t|t} + P_{t|t}F'P^{-1}_{t+1|t}(\beta_{t+1|T} - F\beta_{t|t} - \mu) = \beta_{t|t} - K_{t|T}(\beta_{t+1|T} - F\beta_{t|t} - \mu). \tag{3.1.29}$$

Where K is the “Kalman gain”. Having said that, going back to our equation 3.3.24, we can now explain the alternative update system thought by Carter and Kohn. In equation 3.3.24 $\beta_{t|t}$ and $P_{t|t}$ are computed from the Kalman filter as explained above. Then, the update is done as the Kalman smoother, but what changes is the prediction error, because this time we are conditioning not on Φ_t but on β_{t+1}. In this case the prediction error is:

$$\zeta_{t+1|t} = \beta_{t+1} - F\beta_{t|t} - \mu. \tag{3.1.30}$$

As the innovation in predicted $\beta_{t+1|t}$ where we have some realized β_{t+1} drawn from its probability distribution. Carter Kohn smoother comprises updates to the conditional expectation that use this news

$$E[\beta_t \big| \Phi_t, \beta_{t+1}] = E[\beta_t \big| \Phi_t] + \Sigma_{\beta\zeta}\Sigma^{-1}_{\zeta\zeta}\zeta_{t+1|t} =$$
$$\beta_{t|t} + \Sigma_{\beta\zeta}\Sigma^{-1}_{\zeta\zeta}\zeta_{t+1|t} var[\beta_t \big| \Phi_t, \beta_{t+1}] = var[\beta_t \big| \Phi_t] - \Sigma_{\beta\zeta}\Sigma^{-1}_{\zeta\zeta}\Sigma_{\zeta\beta} =$$
$$\mathrm{P}_{t|t} - \Sigma_{\beta\zeta}\Sigma^{-1}_{\zeta\zeta}\Sigma_{\zeta\beta} = P_{t|t,\beta_{t+1}}.$$

Where $\beta_{t|t}$ and $P_{t|t}$ are estimated with the Kalman Filter. We obtain that:

$$\Sigma_{\zeta\zeta} = var[\beta_{t+1} - F\beta_{t|t} - \mu] = var[F\beta_t + \mu + v_{t+1} - F\beta_{t|t} - \mu]$$
$$=\mathrm{var}[\mathrm{F}(\beta_t - \beta_{t|t}) + v_{t+1}] = FP_{t|t}F' + Q$$
$$\Sigma_{\beta\zeta} = E[(\beta_t - \beta_{t\big|t})(\beta_{t+1} - F\beta_{t|t} - \mu)'] =$$
$$\mathrm{E}[(\beta_t - \beta_{t|t})(F(\beta_t - \beta_{t|t}) + v_{t+1})'] = P_{t|t}F'.$$

So using the definition to update the estimates of beta we do have:

$$\beta_{t|t,\beta_{t+1}} = \beta_{t|t} + P_{t|t}F'(FP_{t|t}F' + Q)^{-1}(\beta_{t+1} - F\beta_{t|t} - \mu) =$$
$$=\beta_{t|t} - K_{t|t}(\beta_{t+1} - F\beta_{t|t} - \mu).$$

Where:

$$K_{t|t+1} = -P_{t|t}F'(FP_{t|t}F' + Q)^{-1}. \tag{3.1.31}$$

Like the Kalman smoother, this uses the filter's estimate of $P_{t|t}$ and updates β_t using the error in predicting β_{t+1} not the data y_t. It's the beta prediction to run backward not the data. In the end we have:

$$\beta_{t|t,\beta_{t+1}} \sim N(\beta_{t|t,\beta_{t+1}}, P_{t|t,\beta_{t+1}}). \tag{3.1.32}$$

One problem with the Carter and Kohn (1994) algorithm is that if the matrix Q is singular, which is not uncommon, it means that we have more states than shocks, the algorithm must be changed as stated in

Kim and Nelson (1999) treating only those states that are shocked as observed. Now going back to where we started, our aim was to find the conditional distribution of the parameters of our model and apply Gibbs sampling. Thanks to the Kalman Smoother we do have the estimates of the betas, we can take them as parameters to develop the distribution of the other parameters. So, stating that $\theta = [H, R, Q, F, \mu]$ the steps to follow are:

- Step 1 Conditioned on θ and y we generate $B_T = (\beta_T, \beta_{T-1}, \dots \beta_1)$ the state vectors of β_T, we do need conditional distribution for all the betas
- Step 2 Conditioned on the B_T we do estimate θ
- Step 3 Iterate till convergence

3.1.6 Applied methods on FAVAR model for climate variables

Applying now what we have said to the FAVAR model, we obtain the Bayesian approach of estimating FAVAR. I used this technique in my analysis. The FAVAR model for my analysis can be written in state-space form:

$$X_t = b_i F_{1t} + b_i Y_t + \gamma_i F_{2t} + v_{i,t}, \tag{3.1.33}$$

$$Z_t = c_t + \sum_{j=1}^{P} B_j Z_{t-j} + e_t, \tag{3.1.34}$$

$$Z_t = [F_{2t}, Y_t, F_{1t}], \tag{3.1.35}$$

$$VAR(v_{i,t}) = R, VAR(e_t) = Q. \tag{3.1.36}$$

Where, X_t is a T x M matrix containing a panel of macroeconomic and economic variables, specifically $X_t = [X_1, X_2]$ where X_1 is the economic panel, X_2 is the climatic panel. While Y_t is $N \times 1$ vector of observed variables composed by $Int.Rate_t$ that actually denotes the 3M European Money Market Interest Rate, so $Y_t = [Int.Rate_t]$ and F_{1t} and F_{2t} are unobserved factors which summarize the information in the data $X_{i,t}$. F_{1t} is the economic factor extracted from X_1, F_{2t} is the climatic factor extracted from X_2.
$Z_t = [F_{2t}, Y_t, F_{1t}]$ is the vector containing both the unobserved and the observed variables, to use a shorter notation. The first equation is the observation equation of the model, while the second equation is the transition equation. To make a parallelism with Bernanke et al. (2005) they consider a shock to the interest rate in the transition equation and calculate the impulse response of each variable in their panel $X_{i,t}$. I instead assume a shock to the climatic factor F_2 and as them I compute the impulse responses of the whole panel $X_{i,t}$ and of Y_t. I assume that the lag length in the transition equation equals 2 and there are 2 unobserved factors as previously stated $F_t = [F_{1t}, F_{2t}]$. For a matter of convenience and to link to notation in (3.1.17)–(3.1.20), I further summarize the state space model by writing it in this form:

$$X_t = H\beta_t + \epsilon_t, \tag{3.1.37}$$

$$\beta_t = \mu + F\beta_{t-1} + e_t, \tag{3.1.38}$$

Where the first equation is the observation equation composed by:
X_t is the total panel, $H = [b_i, \gamma_i]$ the loadings matrix in the form:

$$H = \begin{bmatrix} 1 & 0 & \gamma_{13} & 0 & 0 & 0 \\ 0 & 1 & \gamma_{23} & 0 & 0 & 0 \\ b_{31} & b_{32} & \gamma_{33} & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ b_{NN1} & b_{NN2} & \gamma_{NN3} & 0 & 0 & 0 \end{bmatrix} \tag{3.1.39}$$

The first block is an identity (KxK) where K is the number of factors, to lock rotation and solve the problem of identification of the model due to generated regressors.
NN is the number of time series in the total panel, and γ_i is different from 0 only for fast moving variables[11].
$\beta_t = [F_{2t}, Int.Rate_t, F_{1t}, F_{2t-1}, Int.Rate_{t-1}, F_{t-1}]'$, where F_{1t} is the economic factor, and F_{2t} is the climate factor. While the second equation is the transition equation or VAR(2) equation where:

$$\beta_t = [F_{2t}, Int.Rate_t, F_{1t}, F_{2t-1}, Int.Rate_{t-1}, F_{t-1}]',$$

and

$$\beta_{t-1} = [F_{2t-1}, Int.Rate_{t-1}, F_{1t-1}, F_{2t-2}, Int.Rate_{t-2}, F_{1t-2}]',$$

that represent a VAR(2).
F is a 6x6 factor loading matrix of the form:

$$F = \begin{bmatrix} A_{11} & A_{12} & A_{13} & 0 & 0 & 0 \\ A_{21} & A_{22} & A_{23} & 0 & 0 & 0 \\ A_{31} & A_{32} & A_{33} & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{3.1.40}$$

The variance-covariance matrix for the observation equation is a NN x NN diagonal. Where NN is the number of time series in the total panel.

$$VAR(\epsilon_t) = R = \begin{bmatrix} R_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & R_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & ... & 0 & 0 & 0 \\ 0 & 0 & 0 & R_{NN} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.1.41}$$

[11]The difference between fast-moving variables and slow-moving variables is necessary for running the Gibbs Sample, I explain and justify what are the fast-moving variables in the following paragraph when I discuss the estimation of the model

The variance-covariance matrix for the VAR(2) equation is instead a 6X6 in the form:

$$VAR(e_t) = Q = \begin{bmatrix} Q_{11} & Q_{12} & Q_{13} & 0 & 0 & 0 \\ Q_{21} & Q_{22} & Q_{23} & 0 & 0 & 0 \\ Q_{31} & Q_{32} & Q_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (3.1.42)$$

Having defined all parameters of the model, to estimate it I will proceed with the estimation:

- Step 1. Set priors and starting values. First of all, I identify the total panel X, define the number of observation (T) and the number of time series (M). One important thing is that as I want to extract two factors one for the economic panel and the other one for the climate panel, the PCA to get an initial value of the factors, will be used separately on $X1$ (economic panel) and on $X2$ (climate panel). The reason behind this is that when I compute the impulse response in the last step, as I am interested in understanding the reaction of economic variables given a shock in the climate variables, it is important to have a factor that derives only from climate variables. If PCA was conducted on the total panel, then it would have not been possible to single out the effect of climate from the effect of other economic variables. Another important thing to highlight in this phase is that, in order to run the Gibbs sampler I will treat the climatic variables as slow-moving and the economic variables as fast-moving because I consider economic variables reacting contemporaneous to the climate shock. Moving on, the prior for the factor loadings in the observation equation, I assume is normal. Let $H_i = [b_{ij}, \gamma_i]$. Then $p(H_i) \sim N(H_{i0}, \Sigma_{H_i})$. The prior for the diagonal elements of R is an inverse Gamma and given by $p(R_{ii}) \sim IG(R_{ii0}, V_{R0})$. The prior for the VAR parameters μ, F and Q can be set using any of the priors for VARs for example independent Normal Inverse Wishart prior. Collecting the VAR coefficients in the matrix B and the non-zero elements of Q in the matrix Ω. This prior can be represented as $p(B) \sim N(B_0, \Sigma_B)$ and $p(\Omega) \sim IW(\Omega_0, V_0)$. The Kalman filter requires the initial value of the state vector, first row, $\beta_t = [f_{2t}, Int.Rate_t, f_{1t}, f_{2t-1}, Int.Rate_{t-1}, f_{t-1}]'$. Where f_{2t} is the climatic factor, and f_{1t} is the economic one both found with PCA so we can set β_0. Instead, for the two variance-covariance matrix they can be set equal to the identity matrix to start the algorithm.

- Step 2. Conditional on the factors F_T and R_{ii} we have to sample the factor loadings H from their conditional distributions. For each variable in X_{it}, they have a normal conditional posterior.

  $H(H_i|F_t, R_{ii}) \sim N(H_i^*, V_i^*)$ $H_i^* = ({\Sigma_{H_i1}}^{-1} + \frac{1}{R_{ii}} Z_t' Z_t)^{-1} (\Sigma_{H_i}^{-1} H_{i0} + \frac{1}{R_{ii}} Z_t' X_{it})$ $V_i^* = (\Sigma_{H_i}^{-1} + \frac{1}{R_{ii}} Z_t' Z_t)^{-1}$

  Those are the Mean and the Variance of a Normal distribution, from the assumption on the prior. Where $Z_t = (F_{2t}, Int.rate_t, F_{1t})$ if the ith series X_{it} is a fast moving data series which has a contemporaneous relationship with the climatic factor and $Z_t = (F_{2t}, Int.Rate)$ if the ith series X_{it} is a slow moving variable which has no contemporaneous relationship with the climatic factor. As we have seen in the description of the model in Chapter II, as the factors and the matrix of coefficient are both estimated, the model is unidentified. According to Bernanke et al. (2005) to solve the identification problem we fix the top KxK block of b_{ij} to an identity matrix, where K is the number of factors. As I have two factors my identity matrix will be 2 by 2. As showed in the paragraph above.

- Step 3. Conditional on the Factors and the factor loadings H, it is now possible to sample the variance of the error terms of the observations equation R_{ii} from the inverse Gamma distribution with scale parameter $(X_{it} - Z_t H_i)'(X_{it} - Z_t H_i) + R_{ii0}$ with degree of freedom $T + V_{R0}$ where T is the length of the estimation sample.

- Step 4 Conditional on the factors F_t and the error covariance matrix Ω, the posterior for the VAR coefficients $B = [\mu, F]$, the coefficients in the transition equation of the model, is normal and given as

$$H(B|F_t, \Omega) \sim N(B^*, D^*),$$

where B_0, Σ_B then the parameters of the VAR will be:

$$B^* = (\Sigma_B^{-1} + \Omega^{-1} \otimes \bar{X}_t' \bar{X}_t)^{-1} (\Sigma_B^{-1} vec(B_0) + \Omega^{-1} \otimes \bar{X}_t{}' \bar{X}_t) vec(\hat{B})),$$

$$D^* = (\Sigma_B^{-1} + \Omega^{-1} \otimes \bar{X}_t' \bar{X}_t)^{-1},$$

Where $X_t = [f_{2t-1}, Int.Rate_{t-1}, f_{1t-1}, f_{2t-2}, Int.Rate_{t-2}, f_{1t-2}, 1]$ and $\hat{B}$ is the OLS estimate of B, the loadings of the transition equation.

- Step 5 Conditional on the factors F_t and the VAR coefficients B, the error covariance Ω has a inverse Wishart posterior with scale matrix $(Y_t - (X_t)B)'(Y_t - (\bar{X}_t)B) + \Omega_0$ and the degrees of freedom $T + V_0$.

- Step 6 Given the parameters H_i,R,B and Ω the model can be cast into state-space form and then the factors F_t can now be sampled by running the Carter and Kohn (1994) algorithm obtain the updates for the factors, and run it M times with the Gibbs sampling.

- Step 7 At the end I will, through the Cholesky decomposition of the variance-covariance matrix of the VAR(2), create a normal shock to the parameters of the VAR and obtain the impulse response for the variables of the total Panel.

Chapter 4

Empirical Results

4.1 Data description

In my analysis I verify whether climate change can have an impact monetary policy.

I have done a preliminary analysis mentioned in the introduction, to check empirically whether a direct impact of climate could have an effect on the ECB Refinancing Rate. Historically, central banks have never tailored monetary policy as a consequence of a climatic event, so expectations are that the ECB Refinancing rate should not be impacted directly by a climatic shock.

To verify this expectation I calibrated a VAR(2) model in this way:

$$Y = \begin{bmatrix} GHG_{Emission} \\ GDP_{EU} \\ ECB_{Rate} \end{bmatrix} = A_1 Y_{t-1} + A_2 Y_{t-2} + \epsilon_t.$$

I then supposed a shock in GHG emission and computed the impulse response of the model. From the impulse response I have obtained, (4.1) it is clear that no impact of GHG is on the ECB Refinancing rate as the impulse response is close to zero. Then, expectations are proved. What can be seen in (4.1) also is that nor the GDP's response is significantly different from zero.

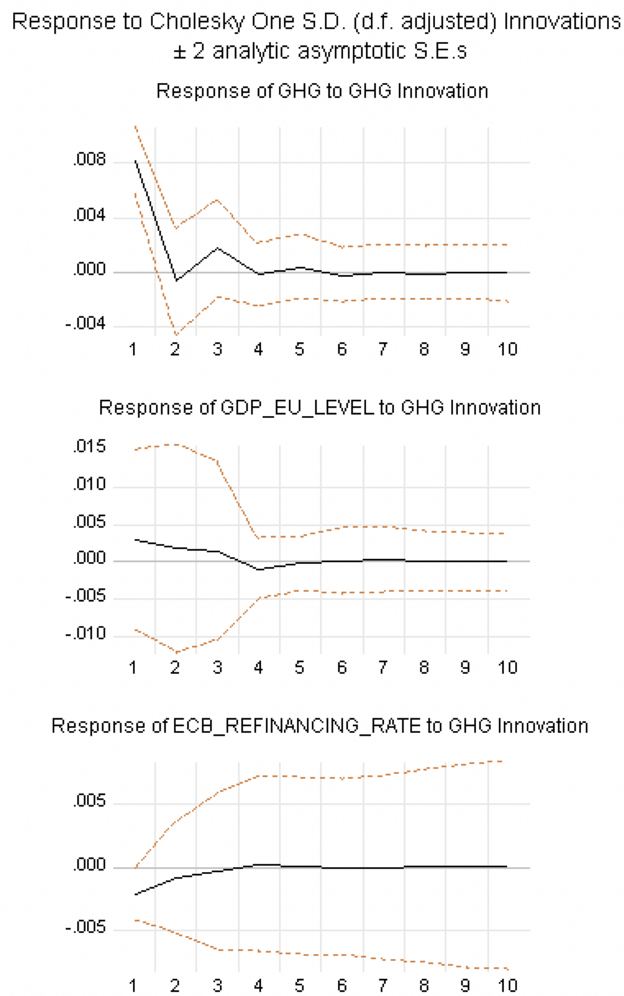


Figure 4.1: Impulse Response of EU GDP, ECB Refinancing Rate to GHG shock

According to that, to further investigate the climate impact, I used FAVAR model to check impulse responses of a wider panel of economic variables that constitute the transmission channel of monetary policy.

As stated in previous chapters, I construct, two panels the one composed by economic variables: GDPs, HCPI and Crop Prices of major euorpean countries (in order Germany, Spain, France, Italy, Netherlands) and at european level plus the 3M European money market interest rate. The other one composed by climatic variables: methane CH4 emission, carbon dioxide without land use, GHG including indirect CO2, GHG emission without land use, Hydrofluorocarbons in kilotonne of CO2, Unspecified mix of hydrofluorocarbons, nitrous oxide, nitrogen triflouride, perfluorocarbons, sulphurhexafluoride emission,temperature. All data are annual from 1990 to 2020. It is worth highlighting that finding climate related data is not an easy task. Climate change is a recent phenomenon moreover registering its impact is difficult. The granularity of the data is not good. Time series for climate related variables such as GHG Emissions starts from 1990 and ends in 2020. To use FAVAR and apply the restriction mentioned in the previous chapter, all the time series need to be stationary. As all my variables have both a trend and seasonality component due to the nature of the data, I work then with growth rate in order to have stationarity in the data. I provide here the Autocorrelation function graphs after the transformation that shows that time series are stationary and can be used for the analysis.

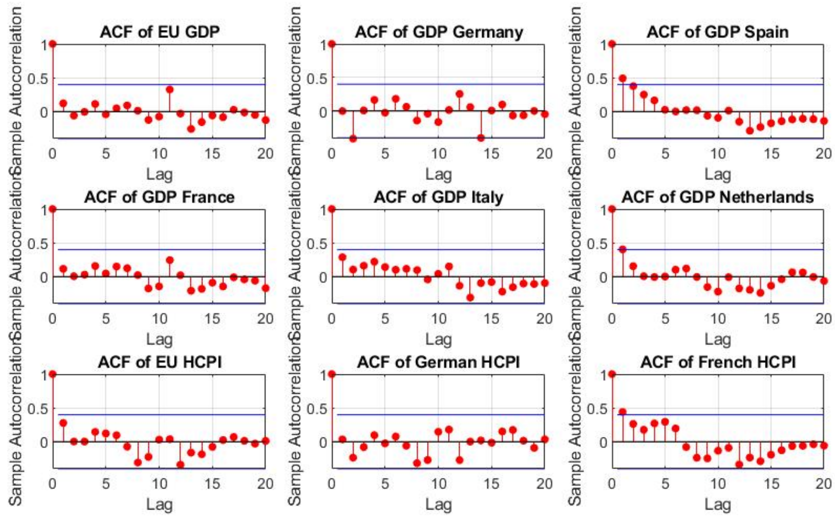


Figure 4.2: Autocorrelation function at 20 lags of GDPs and HCPIs in panel X_1

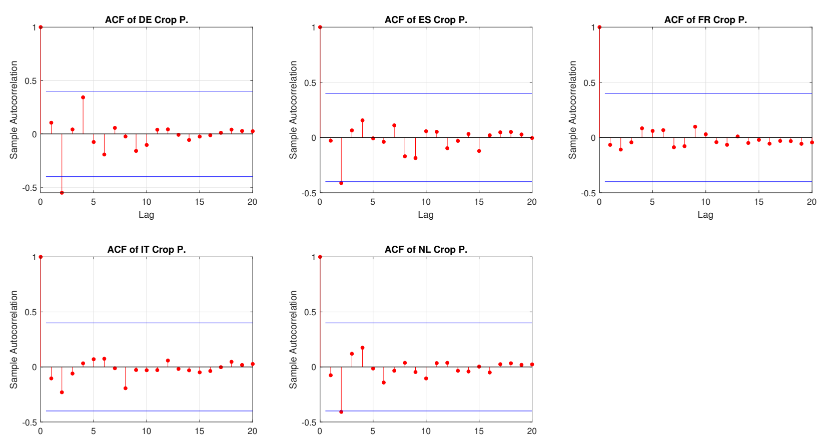


Figure 4.3: Autocorrelation function at 20 lags of Crop Prices in panel X_1

Also for the climatic panel X_2, I worked with growth rate to let them stationary.

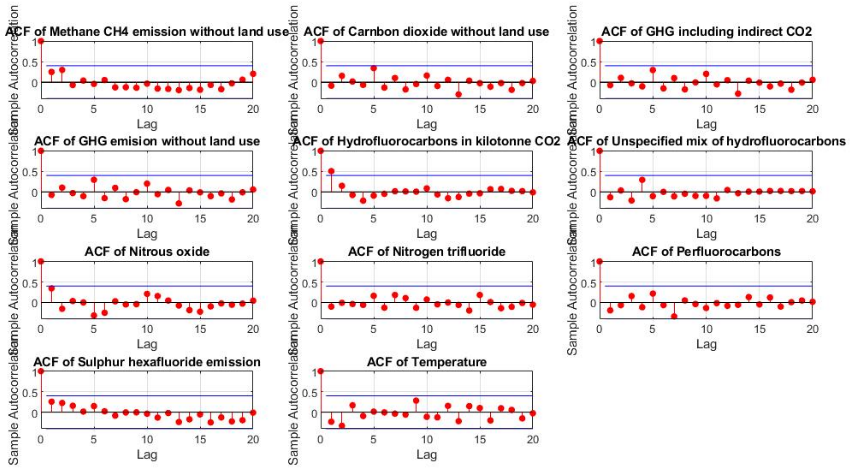


Figure 4.4: Autocorrelation function at 20 lags of Climatic Variables in panel X_2

I provide here a statistics summary of the two panels, in percentage terms, of the variables. The tables below are distinguished according to economic and climatic panel. The economic panel is composed by 18 variables as stated above (GDPs, HCPIs, Crop Prices, 3M Money Market Interest rate) while the climatic panel is composed by 11 variables that I named in the paragraph above. For a matter of space I will rename the variables with short-names.

Table 4.1: This table shows the descriptive statistics for the GDPs of major european countries in panel X_1. Results are in Percentage terms.

	EU GDP(%)	**DE GDP(%)**	**ES GDP(%)**	**FR GDP(%)**	**IT GDP(%)**	**NL GDP(%)**
Observations	25	25	25	25	25	25
Minimum	-4.49	-3.96	-9.84	-5.53	-7.77	-3.45
Maximum	6.19	5.03	8.75	5.54	6.05	7.76
Mean	3.07	2.39	3.47	2.59	2.02	3.47
Median	3.60	2.70	3.93	2.95	2.44	3.81
Variance	6.86e-02	4.71e-02	0.21	5.69e-02	9.16e-02	7.06e-02
Stdev	2.60	2.06	4.60	2.30	3.00	2.61
Skewness	-1.91	-1.36	-1.17	-1.94	-1.74	-0.71
Kurtosis	399.0	267.0	141.0	509.0	404.1	65.0

Table 4.2: This table shows the descriptive statistics for European inflation (HCPI) of major countries in EU belonging to panel X_1. Results are in in percentage terms.

	EU HCPI (%)	**DE HCPI (%)**	**ES HCPI (%)**	**FR HCPI (%)**	**IT HPCI (%)**	**NL HCPI(%)**
Observations	25	25	25	25	25	25
Minimum	0.19	0.22	-0.63	0.09	-0.19	0.11
Maximum	3.26	3.21	4.12	3.15	3.55	5.11
Mean	1.64	1.47	2.04	1.47	1.72	1.90
Median	1.91	1.55	2.34	1.75	1.96	1.72
Variance	7.4826e-05	6.0893e-03	2.0497e-02	6.4558e-03	1.0767e-02	1.2386e-02
Stdev	88.0	80.0	146.1	82.0	106.0	113.1
Skewness	-33.24	18.23	-66.01	-18.86	-33.46	85.03
Kurtosis	-81.1	-48.01	-82.00	-76.04	-71.05	16.03

Table 4.3: This table shows the descriptive statistics for Crop Prices of major European Countries belonging to panel X_1 and of the observed variable in the model (Y_t) the European 3M Money Market Interest Rate. Results must be read in percentage terms.

	DE Crop P.(%)	**ES Crop P.(%)**	**FR Crop P.(%)**	**IT Crop P.(%)**	**NL Crop P.(%)**	**EU 3M Int.Rate(%)**
Observations	25	25	25	25	25	25
Minimum	-0.39	-25.17	-25.44	-38.31	-31.37	-0.39
Maximum	0.56	45.38	55.76	43.94	87.19	6.09
Mean	2.62	1.19	23.41	2.41	3.78	2.03
Median	1.9	-1.20	-1.06	3.10	-1.68	2.09
Variance	3.83	2.10	4.56	3.37	5.67	0.04
Stdev	19.6	14.49	96.83	18.75	23.81	1.97
Skewness	82.1	112.1	451.1	22.49	196.5	23.68
Kurtosis	207.8	234.7	2131	100.09	534.4	123.09

Table 4.4: This table shows the descriptive statistics of climatic variables belonging to panel X_2. Results are in percentage terms.

	Methane CH4 emissions w/o land use(%)	**CO2 w/o land use(%)**	**GHG with indirect CO2(%)**
Observations	25	25	25
Minimum	-3.76	-8.24	-7.37
Maximum	-7.8000e-02	3.12	2.24
Mean	-1.68	-0.91	-1.00
Median	-1.81	-0.46	-0.68
Variance	3.83	2.10	4.56
Stdev	19.58	14.49	21.36
Skewness	77.08	105.87	1.0399
Kurtosis	445.43	481.84	312.24

Table 4.5: This table shows the descriptive statistics of the remaining climatic variables belonging to panel X_2. Results are in percentage terms.

	GHG w/o land use(%)	**HFCs in kilotons CO2(%)**	**Unspecified mix of HFCs(%)**	**Nitrous Oxide(%)**	**Nitrogen Triflouride(%)**
Observations	25	25	25	25	25
Minimum	-13.28	-86.25	-6.27	-48.13	-33.56
Maximum	2.24	16.77	705.41	1.85	207.05
Mean	-1.00	3.82	33.83	-1.51	8.96
Median	0.68	2.94	-2.54	-0.88	-2.83
Variance	3.37	5.67	5.9538e-02	23.19	1.27
Stdev	7.01	157.38	2.44	48.15	11.27
Skewness	-30.46	345.33	-53.70	291.42	-8.20
Kurtosis	291.44	1478.96	225.82	1267.57	299.68

4.2 Interpreting results

Lastly I compute the impulse response of the FAVAR model for the whole panel, X_{it} and Y_t so for the economic and the climatic variables, plus the observed variable, Y_t that is the 3M Money Market Interest rate.

To discuss the results obtained, I start by saying that the model I use is based on Bernanke et al. (2005) and I calibrate it using the same assumptions. I use Cholesky decomposition of the variance-covariance matrix of the VAR(2) to obtain the impulse responses of my variables ordering the variables in the following way: first the economic factor, then the interest rate and lastly the climatic factor. In this way I am indicating that the economic shocks will impact contemporaneously the climatic variables. It is, further important to highlight that, in this simple set up I do not consider extreme weather events or natural disaster associated with climate changes. Proxies used for physical risks are more associated with medium-to-long term effects of global warming, increase in GHG that actually cannot be fully captured by FAVAR. Though different in timing and immediate severity, both risks are dynamically evolving over time and interacting with each other in a non-linear way, a characteristic that this linear model cannot capture. The responses I obtained despite the direction that they take after the shock, show not a persistent and significant impact of GHG increase in economic variables. Maybe only the money market interest rate has a significant response in the short term. More in general, even the shock caused to variables that seems to react, won't last in the long term.

However, according to the results, I can say that GDP both at the aggregate (4.9) and country level (4.6),(4.8),(4.7) has a negative response in the short term, probably within a business cycle 2-8 years. At first the impact is negative, then it has also a positive rebound. Different thing happens to crop prices, that reach their peak still in the short term, but with a positive response. Physical risks associated to climate changes, (as GHG emissions increase) act as a negative supply-side shocks or as a combination of both negative supply and demand shocks through different channels. That is why prices (4.14),(4.12),(4.15) and HCPI (4.10),(4.11) response is ambiguous. It depends on the overall balance of supply and demand shocks, which may differ between individual events. In our case, as crop prices have a positive reaction, this could be linked to a predominance of supply type of adjustments. The effect of HCPI is also ambiguous as, it does have a positive impact (4.10),(4.11) and peak still in the short term, also in this case this could be lead by a supply shock ie. decrease in production. The effect on money market interest rate seems to respond positive, (4.16) so interest rates stay high at least in the short term. The reaction of Money Market interest rates in this case, could push a restrictive monetary policy reaction by central banks.

Starting from these results on the economy and trying to understand how monetary policy could be affected after an increase in GHG emissions is pretty difficult. According to the results I have obtained, if this shock will happen then, as GDP has a negative impulse response, a move by policy makers could be of decreasing interest rates, so that economy could go up again, but of course this should be balanced with the effect that expansionary monetary policy action can have on inflation. Transmission channels are also financial markets', as we said in the introduction, represented in my study by the 3M Money Market Interest Rate. As the reaction of the EU 3m Money Rate is positive, this could push a reduction in the aggregate demand as the cost of money demand would be higher. A consequence action also in this case, if the central bank want to restore equilibrium, the action would be the one of following an expansionary monetary policy.

These Results can be compared to the study of Ciccarelli and Marotta (2021). The authors in their study adopted a Structural VAR (SVAR) model to answer three different questions: if the effects of climate-related shocks were significant enough over a business cycle (2-8 years), if the effect of climate change differs in its effect on macroeconomy and if the shocks to economic variable were more demand-or supply driven. Their study is based on verifying whether an increase in GHG emissions and a technological shock (transition policies to resemble transition risk) can have an impact on Prices and Outputs of major European Countries. Actually what they proved is that the impact of physical risk is overall negative on outputs and prices, meaning a decrease in output occurs following a GHG increase and a price increase occurs following a GHG increase. While the opposite happens for the technological shock. Results are then in line with what I have obtained with the FAVAR model following a GHG shock for Crop Prices and GDPs, moreover also in their case the impact was statistically significant only in the short term.

Figure 4.5: Impulse Response of EU GDP in panel X_1

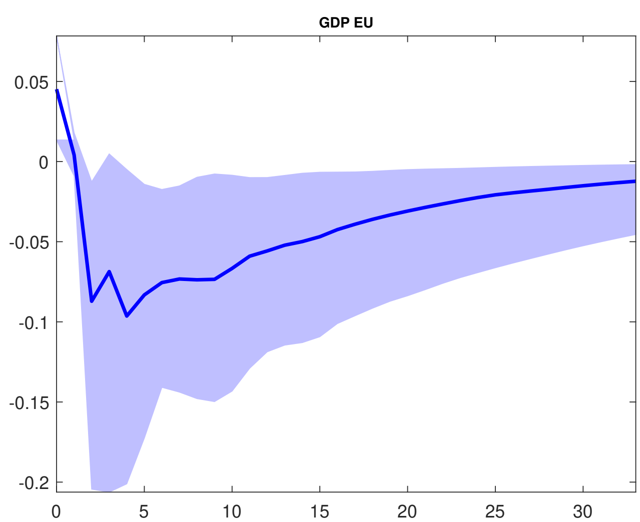
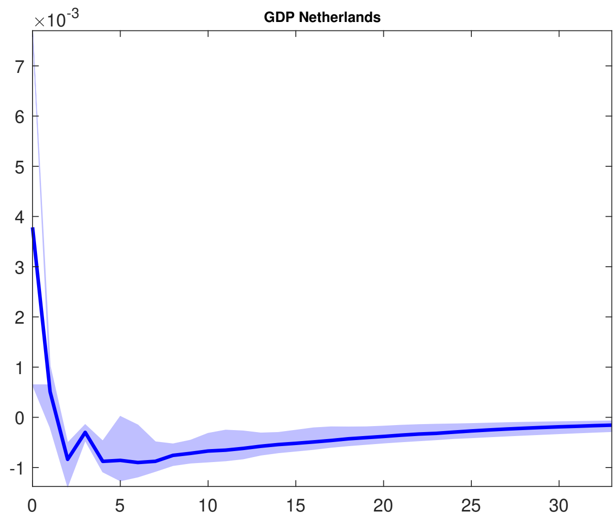


Figure 4.6: Impulse Response of DE GDP in panel X_1



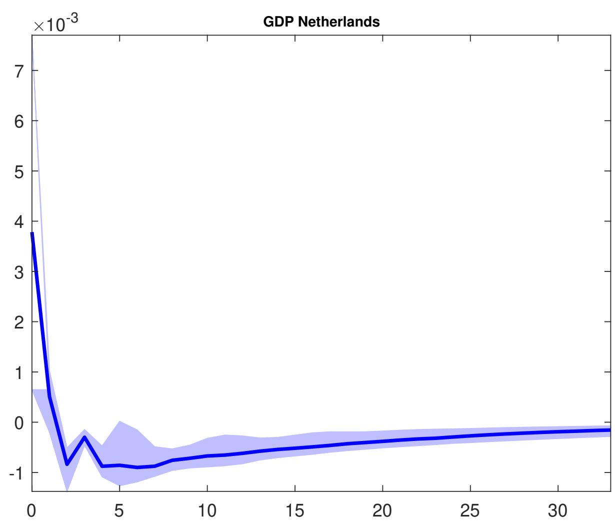
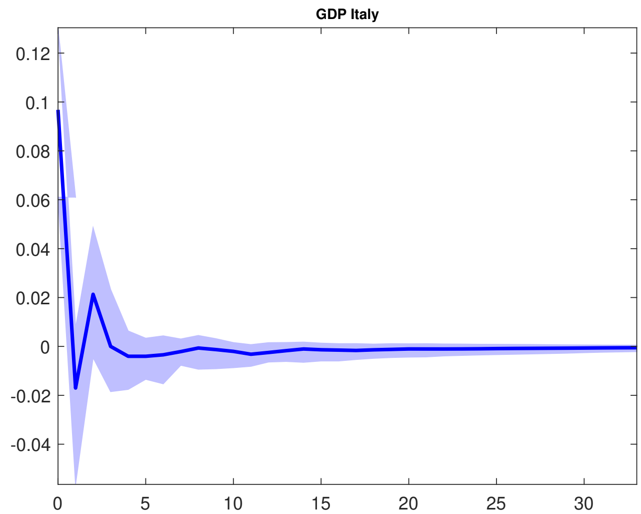


Figure 4.7: Impulse Response of NL GDP in panel X_1

Figure 4.8: Impulse Response of IT GDP in panel X_1



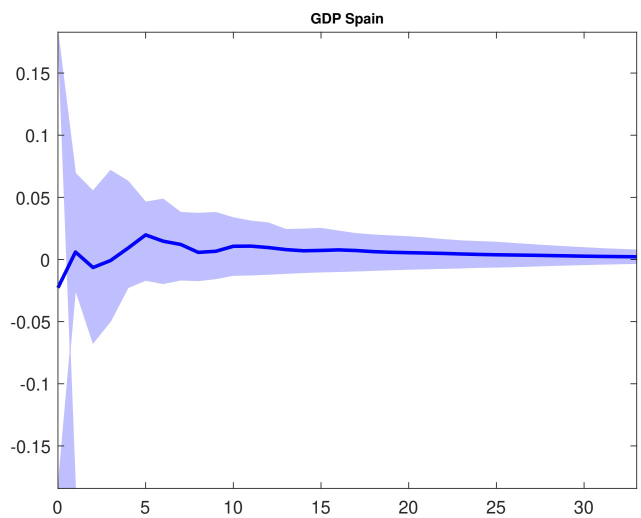
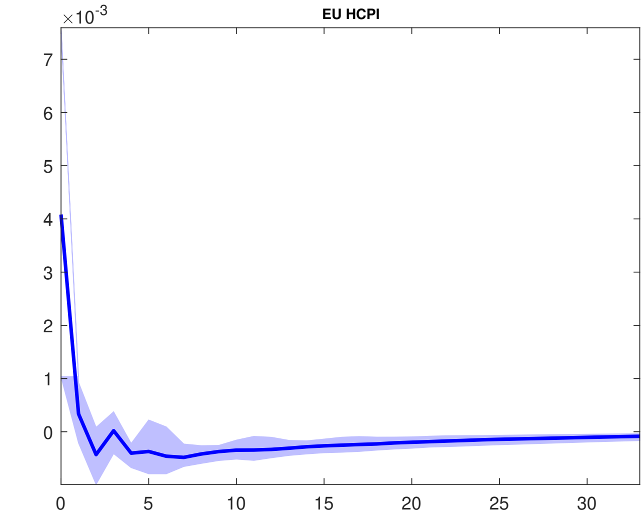


Figure 4.9: Impulse Response of ES GDP in panel X_1

Figure 4.10: Impulse Response of EU HCPI in panel X_1



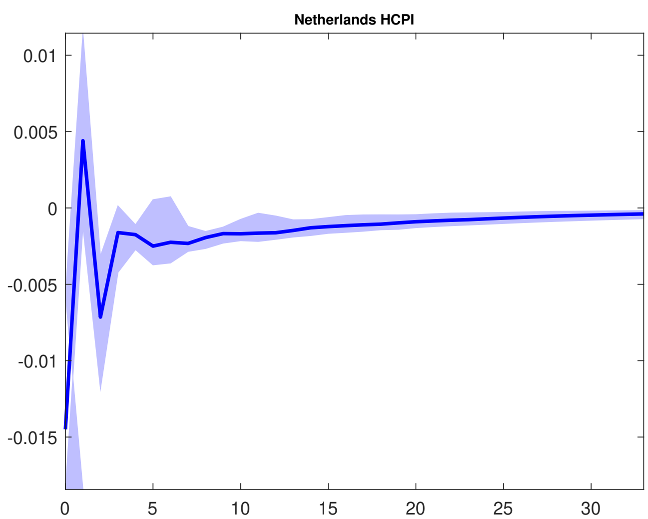
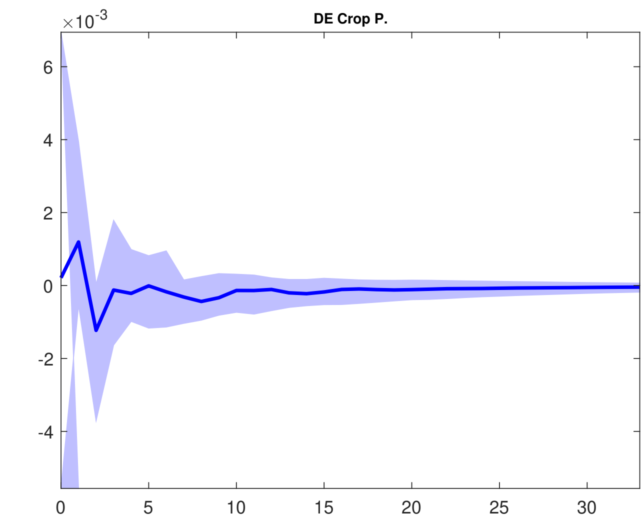


Figure 4.11: Impulse Response of NL HCPI in panel X_1

Figure 4.12: Impulse Response of DE Crop Prices in panel X_1



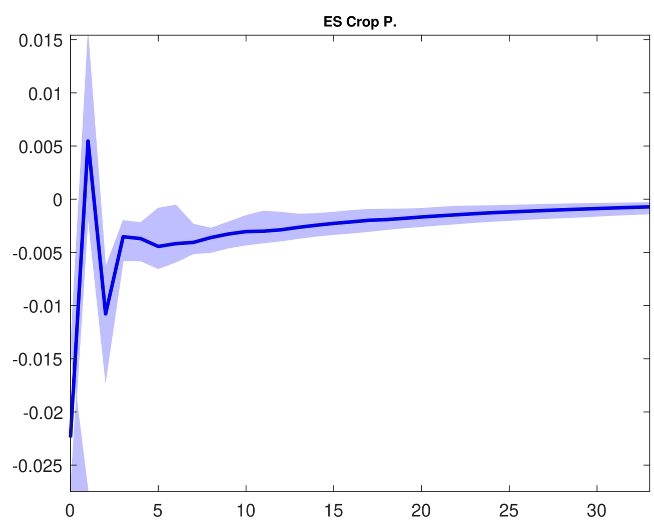
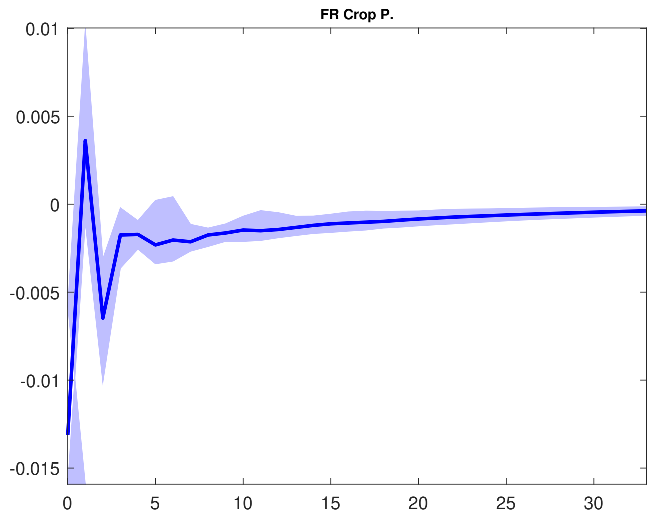


Figure 4.13: Impulse Response of ES Crop Prices in panel X_1

Figure 4.14: Impulse Response of FR Crop Prices in panel X_1



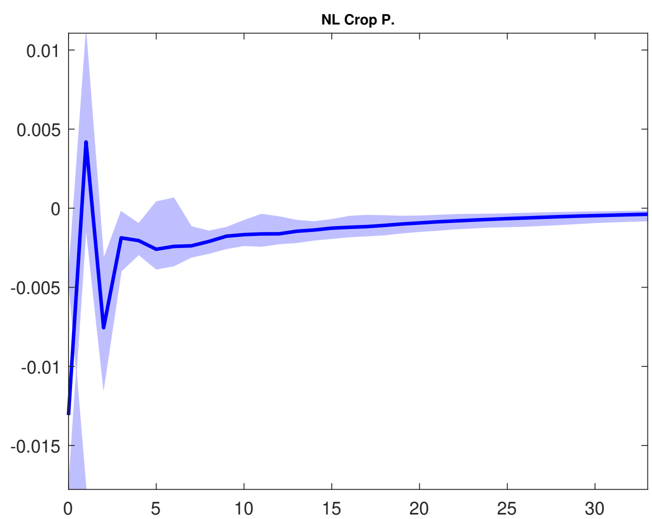


Figure 4.15: Impulse Response of NL Crop Prices in panel X_1

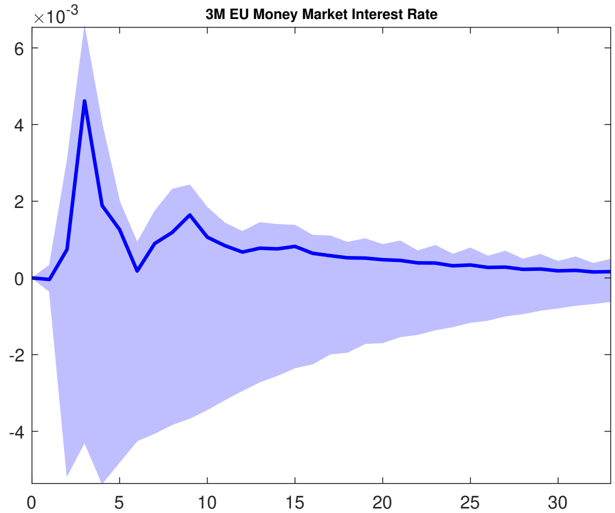


Figure 4.16: Impulse Response of 3M Money Market Interest Rate observable variable in FAVAR Y_t

I have done one more check to verify further the impact of climate in the economy. I compute the impulse response of the extracted factors and the ECB rate. As I take into account only three variables, the model I used is again a VAR(2) model calibrated in this way:

$$Y_t = \begin{bmatrix} F_2 \\ R_{ECB} \\ F_1 \end{bmatrix} = A_1 Y_{t-1} + A_2 Y_{t-2} + \epsilon.$$

Where A_1 and A_2 are the loading matrix. I included the factors extracted from the panel X_{it} so, F_1 and F_2 together with the ECB refinancing rate in a VAR(2) and I checked the impulse responses by shocking the climatic factor F_2. What I have obtained is the figure below:

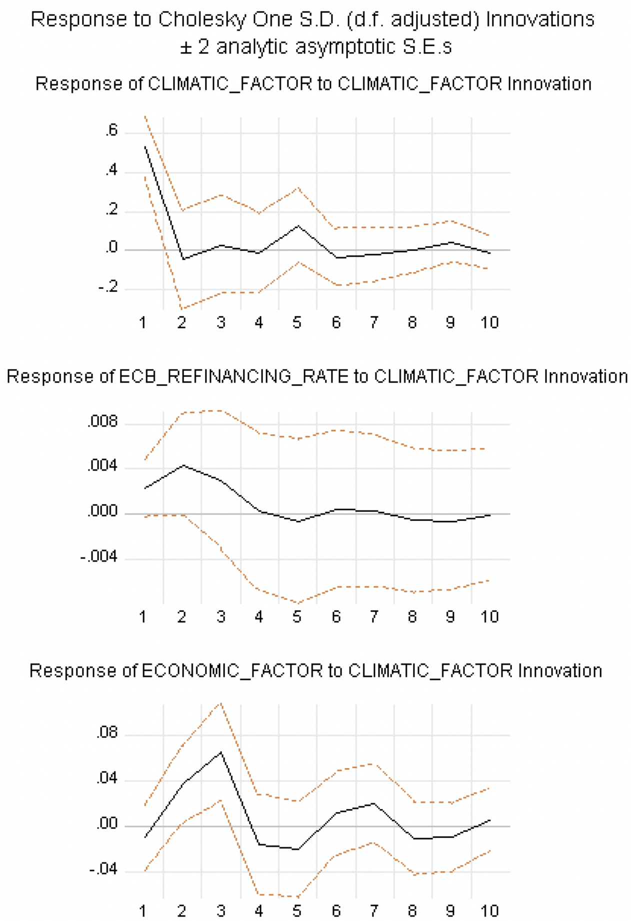


Figure 4.17: Impulse Response for VAR(2) with Economic and Climatic factor

Also in this case the responses are not significant even though the response of the economic factor to the climatic factor innovation, seems close to be significant. As said before, the economic factor is the one extracted from panel X_1 that has GDP, Crop Prices and HCPI. As those are completely different variables that should react differently to climatic shocks, this is the reason why the impulse response is not clearly going up or down. This results states that economy react to GHG innovation, but it is too early to clearly state the impact.

Anyway, this analysis is a starting point that can be further developed by using other factors as climate risk proxies and other model that can capture the dynamic of climate, for example TV-FAVAR even if it will be challenging with so short time series, or mixed-frequency VAR, due to lack of observation. Moreover, further trials can also be done with FAVAR model by changing the assumptions.

Conclusion

In my thesis I developed a novel application of FAVAR within climate change.

My analysis concern with the study of the effect that climate related variables (physical risk, increase of GHG emissions) have on economic variables (GDPs,Prices,HCPI) and monetary policy represented by the ECB refinancing rate. The aim is to address and validate the theories about the response of official rates and monetary policy leeway accounting for climate change. I found that, as stated in the introduction, monetary policy can only be affected indirectly by climate change, as historically no reaction of the central banks as been required following a shock in GHG, no clear evidence can be found in this regard. Looking at the reaction of the other economic variables though, to understand whether monetary policy is affected is important to first understand if the shock caused by climate is demand or supply side driven. In my application we saw that for prices and HCPI a shock could have been caused by supply adjustments (shrink in production) same for GDP (even if in this case also a demand side shock can be argued). In that case as a shrink in production is thought, central banks can take action to foster production by decreasing the official interest rate to maintain price stability and foster economy. Even though the effect on inflation should be regarded too.

The impulse response I have obtained for the panel shows different impact of climate change. My results can be influenced by the fact that the assumptions I used for the FAVAR model are too strict and more suitable for short-term shocks rather than climatic shocks that have a long term impact, nonetheless the granularity of the data is still scarce as time series starts in 1995 till 2020, which are only 25 observations.

However, the study I am doing contribute to the debate about the implication of climate change in the economy. At this stage it is difficult to validate theories on whether climate has a meaningful impact on the economy, and it is even more difficult to assess reaction of central banks to this kind of shocks. Climate is a still open topic, that needs further study. According to other econometric literature on climate see Ciccarelli and Marotta (2021) and Liam, Sánchez, and Höcherl (Liam et al.) no clear or statistically significant results have been obtained on aggregate variables such as GDP, Inflation, Prices nor for physical risk nor for transition risk – also using different models such as simple VAR or SVAR model. By narrowing the analysis to specific sectors and specific effects, instead, results can actually be obtained see Christensen et al. (2021). Authors focused on understanding whether climate, intermittency and demand of wind energy impacted the CO_2 emissions. They showed, by using cointegration approach, that accounting for climate has a mitigating effect on estimated abatement potential of wind power justified by a long-run equilibrium relation.

Bibliography

Amengual, D. and M. W. Watson (2007). Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business & Economic Statistics 25*(1), 91–96.

Amisano, G. and C. Giannini (2012). *Topics in structural VAR econometrics*. Springer Science & Business Media.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bai, J., K. Li, and L. Lu (2016). Estimation and inference of FAVAR models. *Journal of Business & Economic Statistics 34*(4), 620–641.

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

Bai, J. and S. Ng (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics 25*(1), 52–60.

Bai, J. and S. Ng (2013). Principal components estimation and identification of static factors. *Journal of econometrics 176*(1), 18–29.

Bai, J. and P. Wang (2014). Identification theory for high dimensional static and dynamic factor models. *Journal of Econometrics 178*(2), 794–804.

Berkowitz, J. and L. Kilian (2000). Recent developments in bootstrapping time series. *Econometric Reviews 19*(1), 1–48.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly journal of economics 120*(1), 387–422.

Blake, A. P., H. Mumtaz, et al. (2012). Applied Bayesian econometrics for central bankers. *Technical Books*.

Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81*(3), 541–553.

Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica: Journal of the Econometric Society*, 1305–1323.

Choi, I. (2012). Efficient estimation of factor models. *Econometric Theory 28*(2), 274–308.

Christensen, B. J., N. Datta Gupta, and P. Santucci de Magistris (2021). Measuring the impact of clean energy production on co 2 abatement in Denmark: Upper bound estimation and forecasting. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 184*(1), 118–149.

Ciccarelli, M. and F. Marotta (2021). Demand or supply? An empirical exploration of the effects of climate change on the macroeconomy.

Di Persio, L. and M. Frigo (2016). Gibbs sampling approach to regime switching analysis of financial time series. *Journal of Computational and Applied Mathematics 300*, 43–55.

Doz, C., D. Giannone, and L. Reichlin (2011). A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics 164*(1), 188–205.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics 82*(4), 540–554.

Geweke, J. (1977). The dynamic factor analysis of economic time series. *Latent variables in socio-economic models*.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *Review of economics and statistics 80*(2), 218–230.

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

Kim, C.-J. and C. R. Nelson (1999). Has the us economy become more stable? A Bayesian approach based on a markov-switching model of the business cycle. *Review of Economics and Statistics 81*(4), 608–616.

Liam, Y. W., A. E. Sánchez, and B. J. Höcherl. The Macroeconomics of Fighting Climate Change: Estimating the Impact of Carbon taxes, public and private r&d.

Mira, A. and D. J. Sargent (2003). A new strategy for speeding Markov chain Monte Carlo algorithms. *Statistical Methods and Applications 12*(1), 49–60.

Sims, C. A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European economic review 36*(5), 975–1000.

Sims, C. A. (1993). A nine-variable probabilistic macroeconomic forecasting model. In *Business cycles, indicators, and forecasting*, pp. 179–212. University of Chicago press.

Stock, J. H. and M. W. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association 97*(460), 1167–1179.

Stock, J. H. and M. W. Watson (2005). Implications of dynamic factor models for VAR analysis.

Stock, J. H. and M. W. Watson (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of Macroeconomics*, Volume 2, pp. 415–525. Elsevier.

Appendix I

4.3 Proof that the PC estimator minimizes the variance paragraph 1.1.2

It can be shown that the PC estimator efficiently minimize the variance according to Bai and Wang (2014) we know that (equation above dei min) can be rewritten as:

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\lambda_i' F_t)^2 = min_{\Lambda,F}\frac{1}{NT}tr[(X-F\Lambda')(X-F\Lambda')']$$

Where X is a matrix TxN. By developing the trace we obtain:

$$min_{\Lambda,F}\frac{1}{NT}tr[XX'] - 2\frac{1}{NT}tr[X\Lambda F'] + \frac{1}{NT}tr[F\Lambda'\Lambda F']$$

Using the trace derivatives rules we know:

$$\frac{\partial tr(XA)}{\partial X=A}' \frac{\partial tr(XBX'C)}{\partial X} = C'XB' + CXB$$

We obtain:

$$min_{\Lambda,F}\frac{1}{NT}tr[XX'] - 2\frac{1}{NT}tr[X\Lambda F'] + \frac{1}{NT}tr[F\Lambda'\Lambda F'] = min_{\Lambda,F} - 2\frac{1}{NT}tr[\Lambda F'X] + \frac{1}{NT}tr[\Lambda F'F\Lambda']$$

Taking the derivative with respect to Λ equal to zero we have:

$$-2\frac{1}{NT}X'F + \frac{2}{NT}\Lambda F'F = 0 \Longleftrightarrow \Lambda' = X'F(F'F)^{-1} \longleftrightarrow \Lambda = \frac{(X'F)}{T}\frac{F'F}{T}^{-1} = \frac{X'F}{T}$$

Under the normalization condition:

$$\frac{F'F}{T} = I_K$$

By substitution of :

$$\Lambda = \frac{(X'F)}{T} in - 2\frac{1}{NT}tr[\Lambda F'X] + \frac{1}{NT}tr[\Lambda F'F\Lambda']$$

obtaining:

$$min_F - \frac{1}{NT^2}tr[X'FF'X] + \frac{1}{N}tr[\Lambda\Lambda'] = min_F - \frac{1}{NT^2}tr[X'FF'X] =$$

$$max_F tr[X'FF'X] = max_F[F'XXF] = max_F[F'C\Lambda C'F]$$

Where the last equality is true because a symmetric matrix can be decomposed as $C\Lambda C'$ with C matric of eigenvectors and Λ is the ordered diagonal matrix of eigenvalues. This maximization problem is solved when F is an eigenvector of $X'X$, that is $C = [C_1 : C_2]$ so that $F = C_1$.

Appendix II

4.4 Proof that the Static form of DFMs is equal to the Dynamic for of DFMs

We can actually show that the dynamic and the static form can converge:

$$\begin{cases} X_t = \Lambda F_t + \epsilon_t \\ F_t = \Phi F_{t-1} + v_t \end{cases} \quad Static form$$

$$\begin{cases} X_t = \Lambda_1 F_t + \Lambda_2 F_{t-1} + \epsilon_t \\ F_t = \Phi F_{t-1} + v_t \end{cases} \quad Dynamic form$$

starting from the dynamic form by substitution we can write:

$$X_t = \Lambda_1 \Phi F_{t-1} + \Lambda_1 v_t + \Lambda_2 F_{t-1} + \epsilon_t$$

suppose $\Lambda_1 \Phi = \tilde{\Lambda}_1$ then we have:

$$\tilde{\Lambda}_1 F_{t-1} + \Lambda_2 F_{t-1} + (\epsilon_t + \Lambda_1 v_t)$$

$$(\tilde{\Lambda}_1 + \Lambda_2) F_{t-1} + (\epsilon_t + \Lambda_1 v_t) \rightarrow X_t = \Lambda_0 F_{t-1} + \xi_t$$

From the static form, if we substitute the second equation into the first one, we obtain:
$X_t = \Lambda \Phi F_{t-1} + \Lambda v_t + \epsilon_t$ that is equal to: $X_t = \Lambda_0 F_{t-1} + \xi_t$
so we have proved that dynamic form of DFM and Static form of DFM can converge.

Appendix III

4.5 MATLAB Code

The code used for this analysis has been developed by Blake et al. (2012) for Bank of England. I have adapted it to fit my study.

```
clear
clc
addpath('functions')
FILENAME=('Paneltot.xlsx');
Sheetname='Foglio5';
[data junk]=xlsread(FILENAME,Sheetname);
Filename1=('Paneltot.xlsx');
Sheetname1='Foglio6';
[index k]=xlsread(Filename1,Sheetname1);
Filename2=('Paneltot.xlsx');
Sheetname2='Foglio4';
[Int b]=xlsread(Filename2,Sheetname2);
Filename3=('NAMES.XLS');
Sheetname3=('Sheet2');
[junnk names]=xlsread(Filename3,Sheetname3);
%%
X1=data(3:end-1,2:18); %economic panel
X2=data(3:end-1,19:end); %climatic panel
X=[X1,X2]; % panel completo
int=Int(1:end-1,2);
index=index(1:end,2);
KK=1;  %number of factors per il panel di economic
LL=1; % number of factors per il panel di clima
L=2;  %number of lags in the VAR
N=KK+LL+1; %number of Variables in var K factors plus the interest rate
NN=cols(X);% size of the panel
PP=cols(X1); % size of the economic panel
MM=cols(X2); % size of the climatic panel
T=rows(X);
%step 1 of the algorithm set starting values and priors
```

Figure 4.18: Part I FAVAR Code

Firstly I loaded the data (Paneltot) and the index. According to what said in Chapter 4, the panel loaded contains both economic and climatic variables, but to study the climatic impact on the economic variables, it is needed to separate the different variables, this is why I create an Economic X_1 panel and Climatic X_2 one, as showed in the figure. Then I continue with the Step I of the estimation. I set the Initial

Values and I do PCA of the two panels. I obtain pmat and pmat2 that are respectively the economic and the climatic factor

```
%get an intial guess for the factor via principal components
pmat=extract(X1,KK); %fattore economico
pmat2=extract(X2,LL); %fattore climatico
beta0=[pmat(1,:) int(1) pmat2(1,:) zeros(1,N)];  %state vector S[t-1/t-1]
ns=cols(beta0);
P00=eye(ns);  %P[t-1/t-1]
rmat=ones(NN,1); %arbitrary starting value for the variance of the idiosyncratic component
Sigma=eye(N);  %arbitrary starting value for the variance of VAR errors
%%
% Trovare i coefficienti dell'observation equation
reps=500;
burn=10;
mm=1;
correlazioni=[];
for i=1:PP
    correlazioni(i,:)=corr(pmat2,X1(1:end,i));
end
correlazioni=[correlazioni;corr(pmat2,int)];
%%
for m=1:reps
%gibbs sampling


%step 2 sample factor loadings
fload=[];
floadr=[];
error=[];
for i=1:NN
    y=X(:,i);
    if index(i)==0
        x=[pmat int];
    else
        x=[pmat int pmat2];
    end
    M=inv(x'*x)*(x'*y);
    V=rmat(i)*inv(x'*x);
    %draw
    ff=M+(randn(1,cols(x))*cholx(V))';
```

Figure 4.19: Part II FAVAR Code

After defining the initial values and the number of repetition for the Gibbs sampling, matrix H (fload) is found. In this matrix I will distinguish between slow-moving variables that will be regressed on $x = [pmat, int]$ that are the economic factor and the interest rate, while fast moving variables are regressed on $x = [pmat, int, pmat2]$ that are the economic, the interest rate and the climatic factor. An extra step is done here to compute the correlation among variables.

```
    %save
    if index(i)==0;
        fload=[fload;ff'];
        floadr=[floadr;0];
    else
           fload=[fload;ff(1:end-1)'];
        floadr=[floadr;ff(end)];
    end
    error=[error y-x*ff];
end

%for identification top K by K block of fload is identity
fload(1:KK+LL,1:KK+LL)=eye(KK+LL);
%for identification top K by 1 block of Floadr is zero
floadr(24:24+KK+LL-1,1)=zeros(KK,1);

%step 3 sample variance of the idiosyncratic components from inverse
%wishart
%%
rmat=[];
for i=1:NN
    rmati= IG(0,0,error(:,i));
    rmat=[rmat rmati];
end
```

Figure 4.20: Part III FAVAR Code

```
%%
%step 4 sample VAR coefficients
Y=[pmat int pmat2];
X3=[lag0(Y,1) lag0(Y,2) ones(rows(Y),1)];
Y=Y(2:end,:);
X3=X3(2:end,:);

M=vec(inv(X3'*X3)*(X3'*Y));  %conditional mean
V=kron(Sigma,inv(X3'*X3)); %conditional variance
chck=-1;                   %make sure VAR is stationary
while chck<0
beta=M+(randn(1,N*(N*L+1))*cholx(V))';  %draw for VAR coefficients
S=stability(beta,N,L);
if S==0
    chck=10;
end
end
beta1=reshape(beta,N*L+1,N);

errorsv=Y-X3*beta1;

%sample VAR covariance
scale=errorsv'*errorsv;
Sigma=iwpQ(T,inv(scale));
```

Figure 4.21: Part IV FAVAR Code

Next step is to initialize the VAR equation. I am conducting a VAR(2) so X_3 has two lags as described in chapter IV. I then check for stability of the VAR.

```
%%
%matrix of factor loadings
H=zeros(NN,(KK+LL+1)*L);
H(1:rows(fload),1:KK+LL+1)=[fload floadr];
H(rows(floadr)+1,KK+1)=1;
%matrix R
R=diag([rmat 0]);
%vector MU
MU=[beta1(end,:)';zeros(3,1)]';
%matrix F
F1=[beta1(1:N,:)';eye(3)];
F=[F1,zeros(6,3)];
%matrix Q
Q=zeros(rows(F),rows(F));
Q(1:N,1:N)=Sigma;
%%
```

Figure 4.22: Part V FAVAR Code

Having all the parameters, I can now start the Carter and Kohn algorithm described in chapter 3. Where beta2 is the outcome of the algorithm, meaning, the update of beta_tt. Initial values for the vector beta are were saved in the vector beta_0, defined in figure one of the MATLAB Code.

```
%Carter and Kohn algorithm to draw the factor
beta_tt=[];          %will hold the filtered state variable
ptt=zeros(T,ns,ns);    % will hold its variance
% %%%%%%%%%%%Step 6a run Kalman Filter
i=1;
x=H;
%Prediction
beta10=MU+beta0*F';
p10=F*P00*F'+Q;
yhat=(x*(beta10)')';
eta=[X(i,:) int(i,:)]-yhat;
feta=(x*p10*x')+R;
%updating
K=(p10*x')*inv(feta);
beta11=(beta10'+K*eta')';
p11=p10-K*(x*p10);
beta_tt=[beta_tt;beta11];
ptt(i,:,:)=p11;
for i=2:T
    %Prediction
beta10=MU+beta11*F';
p10=F*p11*F'+Q;
yhat=(x*(beta10)')';
eta=[X(i,:) int(i,:)]-yhat;
feta=(x*p10*x')+R;
%updating
K=(p10*x')*inv(feta);
beta11=(beta10'+K*eta')';
p11=p10-K*(x*p10);
ptt(i,:,:)=p11;
beta_tt=[beta_tt;beta11];
end
```

Figure 4.23: Part VI FAVAR Code

```
% Backward recursion to calculate the mean and variance of the distribution of the state
%vector
beta2 = zeros(T,ns);   %this will hold the draw of the state variable
jv=1:3; %index of state variables to extract
wa=randn(T,ns);
f=F(jv,:);
q=Q(jv,jv);
mu=MU(jv);
i=T;  %period t
p00=squeeze(ptt(i,jv,jv));
beta2(i,jv)=beta_tt(i:i,jv)+(wa(i:i,jv)*cholx(p00));   %draw for beta in period t from N(beta_tt,ptt)
%periods t-1..to .1
for i=T-1:-1:1
pt=squeeze(ptt(i,:,:));
bm=beta_tt(i:i,:)+(pt*f'*inv(f*pt*f'+q)*(beta2(i+1:i+1,jv)-mu-beta_tt(i,:)*f')')';
pm=pt-pt*f'*inv(f*pt*f'+q)*f*pt;
beta2(i:i,jv)=bm(jv)+(wa(i:i,jv)*cholx(pm(jv,jv)));

end
pmat=beta2(:,1);   %update the factor for economic
pmat2=beta2(:,3);  % update the factor for climate
%%
if m>burn
 %compute impulse response
    A0=cholx(Sigma);
    yhat=zeros(36,N);
vhat=zeros(36,N);
vhat(3,1:N)=[0 0 1]; %shock al fattore climatico

for i=3:36
 yhat(i,:)=[yhat(i-1,:) yhat(i-2,:) 1]*[beta1(1:N*L,:);zeros(1,N)]+vhat(i,:)*A0;
end

yhat1=yhat*H(:,1:KK+LL+1)';  %impulse response for the panel

irfmat(mm,1:36,1:NN+1)=(yhat1);
```

Figure 4.24: Part VII FAVAR Code

Last step is to compute the shock to the VAR(2) equation trough the Cholesky decomposition and to plot the impulse response.

```
end
%%
irf=prctile(irfmat,[50 16 84],1);

names{41}='Interest Rate';
figure(1)
j=1
for i=1:size(irf,3)
subplot(3,10,j)
plotx1(squeeze(irf(:,3:end,i))');
title(strcat('\fontsize{8}', names(i)))
j=j+1
axis tight
end
```

Figure 4.25: Part VIII FAVAR Code

Here the confidence interval for the impulse response are defined $[50, 16, 84]$.

Chapter 5

Summary

Climate change is a hot and important topic nowadays consensus is growing that if not stopped it will at best costly for the economy, at worst disastrous for human society. Proof of this is for example the PG&E (Pacific Gas and Electric) case, that has been defined by the Wall street Journal as the first corporation subject to climate bankruptcy. Basically, PG&E declared bankruptcy after being hit with what it was then estimated to be 30$ billion in liabilities tied to tragic wildfires in 2017 and 2018, which took 86 lives. Heat waves, and so wildfires, increased in number due to increasing temperature within the country, but the company did not take this risk into consideration ending up with not having the necessary capital to repay clients when the event happened.
Another proof of the importance of climate risk, concerns bank's capital ratios. Banks that for instance have corporate portfolios with large clients exposed to high emission sectors i.e the motor engine sector, are subject to transition risk due to emission intensity of their clients. Policy measures that for example, make petrol and diesel engines less attractive, will reduce demand for cars with these engines leading companies relying on this technology to higher PDs (Probability of defaults). If a bank is financing clients facing this issue then its capital ratio will be affected directly. If the financial sector is affected so is the real economy.

Climate change could have a number of implications for the conduct of monetary policy as a result of its potential impact on the macro economy and financial markets. Two aspects are important to consider. The first aspect concerns the possible implications of climate change and mitigation policies for the ability of central banks to fulfill their price stability mandate. The second aspect concerns the extent to which central banks themselves can play a supporting role in mitigating the risks associated with climate change, while staying within their mandate. Related to this, an additional question arise with regards to the contributions that central banks can make to support the green transition. The way through which climate change affects monetary policy should be by impacting the transmission channel of monetary policy. Climate change can in fact impact both financial markets or GDP, prices etc. For what concern the financial markets side, what could happen is that climate change brings to a sudden re-pricing of climate-related assets or to re-evaluate – in negative- the collaterals that banks own for their loans. A change in value in banks' collaterals could lead to capital and liquidity shortage that weaken banks' ability to channel funds to the real economy. If the financial system is weakened the transmission of monetary policy is impaired. Second thing, that can happen is a dampening in interest rates. If climate-related factors were to cause r to fall further, the policy rate could hit the ELB more often thus limiting the monetary policy space for conventional tools. Another effect caused by climate related variables is that they could correct the identification of shocks relevant for the medium-term inflation outlook. This would make it more difficult to assess the monetary policy stance and potentially increase the prevalence of output and price stabilization trade-offs for central banks that focus on price stability. Uncertainty surrounding the magnitude of the effects of climate change and the horizon over which they will play out in the economy will further complicate the assessment of appropriate monetary policy actions. Uncertainty may also destabilize the expectation-formation process of economic agents, in particular with regard to inflation expectations. Moreover, the effects of physical and transition risks could be asymmetric and heterogeneous complicating even more the behavior of a central bank. So, to sum up, the interest rate channel, the credit channel, the asset price channel, the exchange rate channel and the expectation channel can all be affected by climate change variables. Moreover, they can be affected differently according to the physical or transition risk.

Knowledge on that, central banks have not yet taken the impact of climate change into account in the design of their monetary policy strategy. However, some consideration may already be done to redesign future monetary policy. One problem caused by the climate change is that it shapes economic and financial trends over horizons that exceed the traditional monetary policy horizon. This means that some of the climate change shocks may call for the medium-term policy horizon to be lengthened to take account of the impacts of possibly repeated and correlated climate and/or transition shocks to price stability if, for example, they lead to trade-offs between inflation and employment. At the same time, the credibility of the central bank may be compromised if the time horizon is extended too far into the future and inflation targets are missed too often. In this case, clear communication about the policy intention of the central bank will be essential to mitigate credibility losses. The challenges for monetary policy are contained if the transition is orderly and spread over decades. As we have said before, climate change may put downward pressure on the drivers of r due to shocks on the demand or supply side, by reducing consumption from the demand side or by reducing production due to extreme weather events for the supply side. In any case what derives from these shocks is that the ECB will have a lower policy leeway in taking decision, so the monetary policy could become less effective as already impacted by the climate change.

I use factor models to check these theories, more in particular Factor-augmentend vector autoregressive (FAVAR) model, that in the climatic econometrics literature was never used before to assess the impact of climate change. FAVAR model is actually an extension of Dynamic Factor Models and Vector Autoregressive models Kilian and Lütkepohl (2017), Stock and Watson (2016). Factor models are useful in this case because they allow to summarize a wide concept expressed by several variables within just one or few factors of interest, moreover FAVAR overcomes the limit of VAR models ie. omitted variables bias and small-scale studies. Actually I used VAR to check, just for a few variables, as a preliminary assessment, whether climate can have a direct impact on monetary policy. I hereby used a VAR(2) model with variables Y_t=[EU GDP, ECB Refinancing rate, GHG emission] to compute the impulse response of a shock in GHG emission. The results I have obtained perfectly match with theory meaning that the the impulse response of the ECB Refinancing Rate is close to zero pointing out that no direct climate effect exist on monetary policy. Historically, in fact, the ECB has never tailored monetary policy to address a climate-related event. I provide details of this study in the last chapter.

Starting from this result, as I showed that no direct impact of climate exist on monetary policy, represented by ECB Refinancing Rate, I check then the impact of climate change in a variety of economic variables, that represent transmission channels for monetary policy. To do so I use a FAVAR model.
The FAVAR model I use is based on Bernanke et al. (2005) model, where the authors measured monetary policy shock on a vast economic panel. Actually as I am checking for a different shocks I re-calibrate the model to conduct my analysis. I create two different panels one for economic variables and another one for climatic variables.

The FAVAR model for climate variable has been estimated in this way:

$$X_t = b_{i1}F_2 + b_{i2}Int.Rate_t + \gamma_i F_1 + v_{i,t} \tag{5.0.1}$$

$$Z_t = c_t + \sum_{j=1}^{P} B_j Z_{t-j} + e_t \tag{5.0.2}$$

$$Z_t = [F_{2t}, Int.Rate_t, F_{1t}] \tag{5.0.3}$$

$$VAR(v_{i,t}) = R, VAR(e_t) = Q \tag{5.0.4}$$

X_t is the total panel from which I will extract the unobserved factors, then the variable to add to the estimated factors considered as the only observed variable, is the 3m Money Market EU Interest Rate. $X_{i,t}$, that is composed by $X_{i,t} = [X_1, X_2]$ where X_1, is the economic panel, and X_2 is the climatic panel, is linked to the unobserved F_{1t} (economic factor), F_{2t} (climatic factor) and observed variable as showed in the first equation of the system, $Int.Rate_t$. We can further summarize the state space model by writing it in this form:

$$X_{i,t} = H\beta_t + \epsilon_t \tag{5.0.5}$$

$$\beta_t = \mu + F\beta_{t-1} + e_t \tag{5.0.6}$$

Where the first equation is the observation equation composed by:

$X_{i,t}$ is the total panel, $H = [b_i, \gamma_i]$ the loadings matrix in the form:

$$H = \begin{bmatrix} 1 & 0 & \gamma_{13} & 0 & 0 & 0 \\ 0 & 1 & \gamma_{23} & 0 & 0 & 0 \\ b_{31} & b_{32} & \gamma_{33} & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ b_{NN1} & b_{NN2} & \gamma_{NN3} & 0 & 0 & 0 \end{bmatrix} \quad (5.0.7)$$

The first block is an identity (KxK) where K is the number of factors, to lock rotation and solve the problem of identification of the model due to generated regressors. NN is the number of time series in the total panel, and γ_i is different from 0 only for fast moving variables. $\beta_t = [F_{2t}, Int.Rate_t, F_{1t}, F_{2t-1}, Int.Rate_{t-1}, F_{t-1}]'$, where F_{1t} is the economic factor, and F_{2t} is the climate factor.

While the second equation is the transition equation or VAR(2) equation where:

$\beta_t = [F_{2t}, Int.Rate_t, F_{1t}, F_{2t-1}, Int.Rate_{t-1}, F_{t-1}]'$ and

$\beta_{t-1} = [F_{2t-1}, Int.Rate_{t-1}, F_{1t-1}, F_{2t-2}, Int.Rate_{t-2}, F_{1t-2}]'$ that represent a VAR(2).

F is a 6x6 factor loading matrix of the form:

$$F = \begin{bmatrix} A_{11} & A_{12} & A_{13} & 0 & 0 & 0 \\ A_{21} & A_{22} & A_{23} & 0 & 0 & 0 \\ A_{31} & A_{32} & A_{33} & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad (5.0.8)$$

The variance-covariance matrix for the observation equation is a NN x NN diagonal. Where NN is the number of time series in the total panel.

$$VAR(\epsilon_t) = R = \begin{bmatrix} R_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & R_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & ... & 0 & 0 & 0 \\ 0 & 0 & 0 & R_{NN} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (5.0.9)$$

The variance-covariance matrix for the VAR(2) equation is instead a 6X6 in the form:

$$VAR(e_t) = Q \begin{bmatrix} Q_{11} & Q_{12} & Q_{13} & 0 & 0 & 0 \\ Q_{21} & Q_{22} & Q_{23} & 0 & 0 & 0 \\ Q_{31} & Q_{32} & Q_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (5.0.10)$$

Having defined all parameters of the model, to estimate it we will follow the same steps described in chapter III, that I report here applied to my analysis.

- Step 1. Set priors and starting values. First of all, I identify the total panel X and define the number of observation and the number of time series. One important thing is that as we want to extract two factors one for the economic panel and the other one for the climate panel, the PCA to get an initial value of the factors, will be used separately on $X1$ (economic panel) and on $X2$ (climate panel). The reason behind this is that when we will compute the impulse response in the last step, as we are interested in understanding the reaction of economic variables given a shock in the climate variables, it is important to have a factor that derives only from climate variables. If PCA was conducted on the total panel, then it would have not been possible to single out the effect of climate. Another important thing to highlight in this phase is that, in order to run the Gibbs sampler I will treat the climatic variables as slow-moving and the economic variables as fast-moving. Moving on, the prior for the factor loadings in the observation equation, I assume is normal. Let $H_i = [b_{ij}, \gamma_i]$. Then $p(H_i) \sim N(H_{i0}, \Sigma_{H_i})$. The prior for the diagonal elements of R is an inverse Gamma and given by $p(R_{ii}) \sim IG(R_{ii0}, V_{R0})$. The prior for the VAR parameters μ, F and Q can be set using any of the priors for VARs for example independent Normal Inverse Wishart prior. Collecting the VAR coefficients in the matrix B and the non-zero elements of Q in the matrix Ω. This prior can be represented as $p(B) \sim N(B_0, \Sigma_B)$ and $p(\Omega) \sim IW(\Omega_0, V_0)$. The Kalman filter requires the initial value of the state vector, first row, $\beta_t = [f_{2t}, Int.Rate_t, f_{1t}, f_{2t-1}, Int.Rate_{t-1}, f_{t-1}]'$. Where f_{2t} is the climatic factor, and f_{1t} is the economic one both found with PCA so we can set β_0. Instead, for the two variance-covariance matrix they can be set equal to the identity matrix to start the algorithm.
- Step 2. Conditional on the factors F_T and R_{ii} we have to sample the factor loadings H from their conditional distributions. For each variable in X_{it}, they have a normal conditional posterior. $H(H_i|F_t, R_{ii}) \sim N(H_i^*, V_i^*)$ $H_i^* = ({\Sigma_{H_i}}_1^{-1} + \frac{1}{R_{ii}} Z_t' Z_t)^{-1} (\Sigma_{H_i}^{-1} H_{i0} + \frac{1}{R_{ii}} Z_t' X_{it})$ $V_i^* = (\Sigma_{H_i}^{-1} + \frac{1}{R_{ii}} Z_t' Z_t)^{-1}$ Those are the Mean and the Variance of a Normal distribution, from our assumption on the prior. Where $Z_t = (F_{2t}, Int.rate_t, F_{1t})$ if the ith series X_{it} is a fast moving data series which has a contemporaneous relationship with the economic factor and $Z_t = (F_{2t}, Int.Rate)$ if the ith series X_{it} is a slow moving variable which has no contemporaneous relationship with the economic factor. As we have seen in the description of the model in Chapter II, as the factors and the matrix of coefficient are both estimated, the model is unidentified. According to Bernanke et al. (2005) to solve the identification problem we fix the top KxK block of b_{ij} to an identity matrix, where K is the number of factors. As we do have two factors our identity matrix will be 2 by 2. As showed in the paragraph above.
- Step 3. Conditional on the Factors and the factor loadings H it is now possible to sample the variance of the error terms of the observations equation R_{ii} from the inverse Gamma distribution with scale parameter $(X_{it} - Z_t H_i)'(X_{it} - Z_t H_i) + R_{ii0}$ with degree of freedom $T + V_{R0}$ where T is the length of the estimation sample.

- Step 4 Conditional on the factors F_t and the error covariance matrix Ω, the posterior for the VAR coefficients $B = [\mu, F]$, the coefficients in the transition equation of the model, is normal and given as

$$H(B|F_t, \Omega) \sim N(B^*, D^*)$$

$$\text{where } B_0, \Sigma_B$$

$$\text{then the parameters of the VAR will be:}$$

$$B^* = (\Sigma_B^{-1} + \Omega^{-1} \otimes \bar{X}_t' \bar{X}_t)^{-1} (\Sigma_B^{-1} vec(B_0) + \Omega^{-1} \otimes \bar{X}_t{}' \bar{X}_t) vec(\hat{B}))$$

$$\text{D}^* = (\Sigma_B^{-1} + \Omega^{-1} \otimes \bar{X}_t' \bar{X}_t)^{-1}$$

  Where $X_t = [f_{2t-1}, Int.Rate_{t-1}, f_{1t-1}, f_{2t-2}, Int.Rate_{t-2}, f_{1t-2}, 1]$ and $\hat{B}$ is the OLS estimate of B, the loadings of the transition equation.

- Step 5 Conditional on the factors F_t and the VAR coefficients B, the error covariance Ω has a inverse Wishart posterior with scale matrix $(Y_t - (X_t)B)'(Y_t - (\bar{X}_t)B) + \Omega_0$ and the degrees of freedom $T + V_0$.

- Step 6 Given the parameters H_i,R,B and Ω the model can be cast into state-space form and then the factors F_t can now be sampled by running the Carter and Kohn (1994) algorithm obtain the updates for the factors, and run it M times with the Gibbs sampling.

- Step 7 At the end I will, through the Cholesky decomposition of the variance-covariance matrix of the VAR(2), create a normal shock to the parameters of the VAR and obtain the impulse response for the variables of the total Panel.

The results I have obtained can be summed up in this way n this simple set up I do not consider extreme weather events or natural disaster associated with climate changes, in fact proxies used for physical risks are more associated with medium-to-long term effects of global warming, increase in GHG that cannot be fully captured by FAVAR. Though different in timing and immediate severity, both risks are dynamically evolving over time and interacting with each other in a non-linear way, a characteristic that this linear model cannot capture. The responses I obtained despite the direction that they take after the shock, show not a persistent and significant impact of GHG increase in economic variables . Maybe only the money market interest rate has a significant response in the short term. More in general, even the shock caused to variables that seems to react, won't last in the long term.

However, to discuss the results I obtain in figure 4.1, I can say that GDP both at the aggregate and country level, has a negative response in the short term, probably within a business cycle 2-8 years. At first the impact is negative, then it has also a positive rebound. Different thing happens to crop prices, that reach their peak still in the short term, but with a positive response. Physical risks associated to climate changes, (as GHG emissions increase) act as a negative supply-side shocks or as a combination of both negative supply and demand shocks through different channels. That is why prices and HCPI response is ambiguous. It depends on the overall balance of supply and demand shocks, which may differ between individual events. In our case, as crop prices have a positive reaction, this could be linked to a predominance of supply type of adjustments. The effect of HCPI is also ambiguous as, it does have a positive impact and peak still in the short term, also in this case this could be lead by a supply shock. The effect on money market interest rate seems to respond positive, so interest rates stay high at least in the short term. The reaction of Money Market interest rates in this case, could push a restrictive monetary policy reaction by central banks.

Starting from these results on the economy and trying to understand how monetary policy could be affected after an increase in GHG emissions is pretty difficult. Monetary policy is affected by mainly two kind of risks as we said in Chapter I: macroeconomic ones or financial ones. Historically, despite huge climate events (i.e Floods, heart-quakes), that caused a reaction by the central bank (i.e. foster consumption), no clear other climate event pushed a reaction by the central bank. Climate change may only affect monetary policy through transmission channels, meaning that a shock on GDP or Inflation or Interest rates, caused by climate related variables, can lead to a move by policy makers. For example in this case, as GDP has a negative impulse response, a move by policy makers could be of decreasing interest rates, so that economy could go up again. Transmission channels are also financial markets' movements influenced by ESG (Environmental Social and Governance) scores. A change in financial instruments linked with climate could push in different direction the economy asking a central bank to take action. As the reaction of the EU 3m Money Rate.

To further test, climate change impact on the economy I have also calibrated a VAR(2) model

$$Y = [F_2, R_{ECB}, F_1] = c + A_1 Y_{t-1} + A_2 Y_{t-2}. \tag{5.0.11}$$

where F_2 is the climatic factor extracted with PCA and Gibbs Sampling in FAVAR code, and F_1 is the economic factor estimated with PCA and Gibbs in FAVAR code to obtain the following impulse responses.

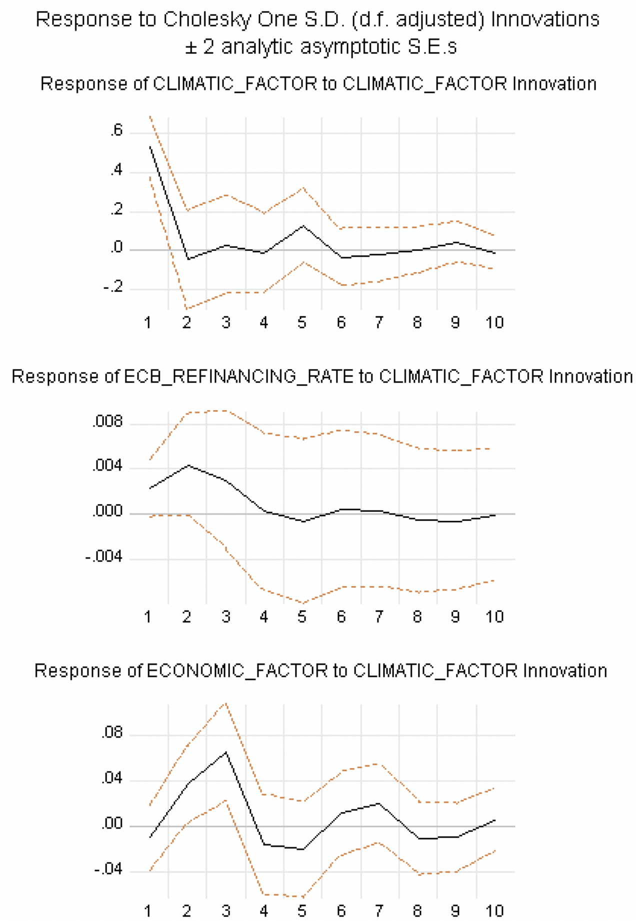


Figure 5.1: Impulse Response for VAR(2) with Economic and Climatic factor

Also in this case the responses are not significant even though the response of the economic factor to the climatic factor innovation, seems close to be significant. As said before, the economic factor is extracted from a panel that has GDP, Crop Prices and HCPI –that is why the impulse response is not clearly up or down. This can be due to shocks in the supply side, as we have seen in the crop prices response with FAVAR (i.e prices increase or GDP decrease). It states that economy react to GHG innovation, but it is too early to clearly state the impact.

Anyway, this analysis is a starting point that can be further developed by using other factors as climate risk proxies and other model that can capture the dynamic of climate, for example TV-FAVAR even if it will be challenging with so short time series, or mixed-frequency VAR, due to lack of observation. Moreover, further trials can also be done with FAVAR model by changing the assumptions. the results I obtained contribute to the debate about the implication of climate change in the economy. At this stage it is difficult to validate theories on whether climate has a meaningful impact on the economy, and it is even more difficult to assess reaction of central banks to this kind of shocks. Climate is a still open topic, that needs further study. According to other econometric literature on climate see Ciccarelli and Marotta (2021) and Liam, Sánchez, and Höcherl (Liam et al.) no clear or statistically significant results have been obtained on aggregate variables such as GDP, Inflation, Prices nor for physical risk nor for transition risk – also using different models such as simple VAR or SVAR model. By narrowing the analysis to specific sectors and specific effects, instead, results can actually be obtained see Christensen et al. (2021). Authors focused on understanding whether climate, intermittency and demand of wind energy impacted the CO_2 emissions. They showed, by using cointegration approach, that accounting for climate has a mitigating effect on estimated abatement potential of wind power justified by a long-run equilibrium relation.