Sapienza University of Rome

School of Statistical Sciences
PhD in Methodological Statistics

Thesis For The Degree Of Doctor Of Philosophy

Dynamic Factor Models: improvements and applications

Thesis Advisor
Prof. Marco Lippi

Candidate
Marco Forti
573324

Academic Year 2020/2021 (XXXII cycle)

Ad Alessia

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the last two decades data collection, aided by an increased computational capability, has considerably increased both dimension and structure of the datasets; given this, statisticians and economists may today work with time series of remarkable dimension which may come from different sources. Dealing with such datasets may not be so easy and requires the development of *ad hoc* mathematical models.
One of the targets that we need to achieve designing new models is to contrast parameters proliferation and facilitate models reduction. In order to do that, factor models represent an effective tool since they are able to synthesise information held in huge datasets in a few factors.

Among those methods, Dynamic Factor Models (DFM) represent one of the newest techniques in big data management. In particular, using DFM, it is possible to derive a representation for an infinite panel of time series which is the analogous of the Wold representation for finite dimensional ones. This is the generalised dynamic factor representation and, between its properties, has the one of keeping the time series structure of the dataset thus generally is finalised to do forecasting or structural analysis, where the object of interest are the impulse response functions.

Dynamic Factor Models are one of the possible approaches created to extract statistics out of large datasets, another interesting one surely is the utilisation of Bayesian methods while frequentist alternatives are extensively used in machine learning.

When we talk of big data we generally refer to situations in which the number of explanatory variables is large compared to the sample size.

When in this situation we could say that the researcher is in a position in which he/she is trying to estimate too much relatively to the information in her possess. This is not unfeasible but will typically lead to a very imprecise inference while it is not even doable, at least with conventional methods, when the number of explanatory variables exceeds the number of observations and when a considerable number of these observations are equal to zero. In mathematics a so composed matrix is said to be sparse.

I am not going in the details of sparsity here, for what interests us we only need to underline how sparsity requests specific tools to be approached; such tools may be statistical models in which only a relatively small number of parameters (or predictors) would play an crucial role in the description of the whole dataset. Machine learning models operate according to this logic and lasso is a clear example of that.

Lasso is a frequentist model which basically rely on a penalisation algorithm in order to regularise ("shrink") coefficients, making it possible to deal with high-dimensional data and to avoid over-fitting. By contrast this, such as other machine learning techniques, have some considerable drawbacks: first it is not possible to make any choice over the variables considered relevant for the model, second the selection procedure depends by some penalisation parameter which need to be somehow imposed, as a last point it must be noticed how such selection process would determine, one way or another, some loss of information.

On the other hand Bayesian statistics provided alternative models to handle big data inconvenient. The intuition that is behind the application of those methods in high dimensional data context is the following: dealing with a big amount of data the researcher may incur in over-parametrization/over-fitting problems, in such situation the application of a prior distribution to update the observed data would work out to restrict the parametric space.

Bayesian methods have been demonstrated effective both in case of models with many explanatory variables and in models with many dependent variables (as in the VAR case). Analytically speaking the posterior distribution will be obtained by updating observed data with prior information over the parameter of interest, while controlling the prior variance will give to researcher some control over the degree of shrinkage or, in other words, will put some numeric value over the degree of confidence that the researcher has about his beliefs.

If on one hand the Bayesian approach has demonstrated to have some interesting aspects regarding inference and forecasting performances, on the other hand such approach, and thus results, is conditional to the accuracy of the prior choice, plus this technique often imply the utilisation of MCMC processes and other computationally demanding algorithms.

For an extensive survey over Bayesian econometrics the reader can look at the scientific production by Mike West and especially, for what concerns the use of Bayesian analysis as shrinkage method, by Koop and Korobilis (2010).

Since what Dynamic Factor Model return to us is a time series, it is totally admissible to consider the application of Bayesian techniques in this environment. Several authors have studied such topic, for a complete excursus the reader can be referred to, among all, Otrok and Whiteman (1998) and Del Negro and Otrok, passing through Mike West who centred great part of his scientific production over theoretical and empirical aspects of Bayesian econometrics.

Dynamic Factor Models estimation developed faster and faster since the pioneering work of Stock and Watson (1989), in which they used factor analysis to compute coincident indicators for the business cycle and financial variables. Since factor models are used to consistently estimate common and idiosyncratic components of macroeconomic variables, DFM has been used to do predictions (see Stock and Watson (2002)) or to compute VARs and impulse response functions; the study of the link between factor modes and structural VARs has been deepened by Stock and Watson (2005b) and Forni et al. (2009).

The general idea of DFM is to reduce a large panel of data in two mutually orthogonal components: the common one, which is strongly correlated with the rest of the panel, and an idiosyncratic one, which is only mildly cross-correlated.

All the papers we refer here make use of non-parametric techniques in which factors are estimated through a procedure that calculate eigenvalues and eigenvectors from a spectral density matrix, hence essentially realising an approach that is a dynamic version of the principal component one.
Forni et al. (2005) implemented a one-sided estimation of the generalised dynamic factor models in order to improve forecasting performances, the infinite dimensionality has been introduced in the papers "The Generalized Dynamic Factor Model with infinite dimensional factor space" by Forni et al. (2015, 2017), while "Generalised Dynamic Factor Models and Volatilities" by Barigozzi and Hallin (2016, 2017) deepened the study of the volatility generated by a DFM. This last paper, in particular, applies general autoregressive conditional heteroscedastic (GARCH) processes in DFM environment and opens the way to new reflections about volatility modelling nevertheless, by now, not many researches had been conducted studies on this topic.

In consideration of this lack, the purpose of my work is to delve into the study of the idiosyncratic factors generated by a DFM, deepening, in particular, the theme of stochastic volatility.

With the expression "Stochastic Volatility" (SV hereafter) we refer to the relaxation of the homoscedasticity assumption over the variance-covariance matrix.

More in detail, while in (G)ARCH models the conditional variance is assumed to be a deterministic function of past values, in the SV models the volatility is modelled according to some aleatory process.

In formulas, the system which describes the ARCH process is the following:

$$\begin{cases} x_t = a_0 + a_1 x_{t-1} + \epsilon_t \\ \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + v_t \end{cases} \tag{1.1}$$

where $v_t = \epsilon_t^2 - \sigma_t^2$.

Te generalisation of this model ("Generalised-ARCH") impose the introduction of other variables useful to improve the performances of the model, anyway not changing the substance of the method.

On the other hand stochastic volatility, in its base version, is grounded over the following system:

$$\begin{cases} y_t = \mu + h_t + \epsilon_t \\ h_t = \phi h_{t-1} + \eta_t \end{cases} \tag{1.2}$$

where ϵ and η represent two aleatory processes which distribution should opportunely be defined. For a more complete and accurate dissertation of this theme the reader may see Asai et al. (2006) and Chib et al. (2006).

In order to identify the parameters and aleatory distributions, applications of Bayes theorem was considered: a first attempt to investigate this topic was made by Uhlig (1997) and some other followed, anyway no much literature developed from this seminar paper. Dropping Bayesian techniques, the theme of stochastic volatility is well known in literature, especially in the field of finance. Some authors recently proposed a formulation that applies SV to dynamic factor models; relatively to this topic the most interesting articles are: Koopmanp et al., Gorodnichenko and Ng (2017) and Meng.

While Bayesian methods was applied to Stochastic Volatility and dynamic factor models separately, nobody yet seems to have built a model that combine both approaches to deepening the study of volatility. This will be one of the original element proposed in my work while the adoption of Bayesian non-parametric techniques in this combined environment is, at the best of my knowledge, another element of innovation in the literature, possibly offering a new ground for future researches.

Chapter 2

Generalised Dynamic Factor Model

Volatility is an important issue in economics and financial analysis as higher-order shocks, especially second-order volatility shocks, can be identified as an important source of business cycles. Despite that, the interaction mechanism between levels and volatility is not fully understood yet. Moreover, it is well known that volatility is a manifestation of uncertainty and that level variables are affected by realised and expected volatility shocks asymmetrically; it is so central, in order to improve estimations, to separate levels and volatility and to allow for a non-linear interaction of them.

Recently, several authors investigated the way in which levels and volatility interact in a dynamic factor models framework, among all must be cited Koopmanp et al., which uses parametric techniques for DFM estimation, and Gorodnichenko and Ng (2017), which use non-parametric ones but centring their focus to VAR analysis of macroeconomic data.

Baseline models for the nonparametric analysis of big-data can be recognised in Forni et al. (2000), where they operate DFM to identify VARs and compute impulse response functions, and in Forni et al. (2003, 2005) in which they estimate the move from a two-sided to a one-sided estimation of DFM in order to do forecasting. In 2015 an infinite-dimensional representation have been proposed by Forni et al. (2015) - FHLZ (2015) hereafter - and, on this base, Barigozzi and Hallin (2016, 2017) worked on volatilities finalising their study applying Generalised Autoregressive Stochastic Heteroskedastic (GARCH) models and doing forecasting.

Our approach shares with their the non-parametric and model free method to decompose data in common and idiosyncratic component, nevertheless my focus is in deepening the study of the volatility components allowing for a time-varying behaviour of their dynamics. In particular the representation we work on is one-sided and start from a panel Y_{it} of observable stationary stochastic processes, with mean 0 and finite variance, in which we assume that such variables belong to an Hilbert space in L_2 for some probability space (Ω, F, P). The panel hence is composed by $n \times T$ levels of finite realisation and so Y will be: $Y = \{Y_{it} \mid i \in, t \in \mathbb{Z}\}$, where t stands for time and i is the cross-sectional index.

For all $n \in N$ we assume that the spectral measure on Y_n is absolutely continuous respect to the Lebesgue measure on $[-\pi, \pi]$, such that the spectral density matrix is $\Sigma_{Y;n}(\theta)$ that is Hermitian, non-negative definite and has therefore non-negative real eigenvalues for all $\theta \in [-\pi, \pi]$. Such assumptions are necessary and sufficient to guarantee a dynamic factor representation.

We can hence state that Y admits a *dynamic factor* representation, with q factors, if Y_{it} decomposes into a "*common*" component (X_{it}) and an "*idiosyncratic*" component (Z_{it}) such that:

$$Y_{it} = X_{it} + Z_{it} = \sum_{k=1}^{q} b_{ik}(L)u_{kt} + Z_{it} \tag{2.1}$$

Where u_{kt} is a q-dimensional orthonormal process, which is withe noise with mean zero, and Z_{it} is zero-mean second order stationary process whose $Z's$ are weakly cross-correlated, while the two processes are mutually orthogonal at any lead and lag.

Other element of (6.48) is $b_{ik}(L)$ which is a one-sided filter, where L denote the lag operator. This filter is defined to be square-summable $\sum_{m=1}^{\infty} b_{ikm}^2 < \infty$ for all $i \in N$ and $k = 1 \ldots q$. The one-sided filters, as demonstrated by Forni et al. (2017), can be obtained without the finite-dimensionality assumption, by adding the condition under which the common components must have rational spectral density, that is, each filter $b_{if}(L)$ in (6.48) is a ratio of polynomials in L. Forni et al. (2015, 2017) also prove that for generic values of parameters $c_{if,k}$ and $d_{if,k}$, X_{it} has the fundamental representation:

$$X_{it} = \frac{c_{i1}(L)}{d_{i1}(L)}u_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)}u_{2t} + \cdots + \frac{c_{iq}(L)}{d_{iq}(L)}u_{qt} \tag{2.2}$$

in which u_t is fundamental for X_{it}.

In order to ease the reading, from now on I will light the notation omitting the cross-section index i.

For each element of the (6.48) a cross-covariance matrices, for example between Y_t and Y_{t-k}, can be estimated; such matrices may be summarised in the form: $\Gamma_{Y;k}(\theta) = E[Y_t Y_{t-k}^{'}]$

The elements of the (6.48) may be rearranged in terms of spectral density and covariance matrices, this representation is purely illustrative, nevertheless the admissibility of such representation is essential for estimation.

$$\Gamma_{Y;k}(\theta) = \Gamma_{X;k}(\theta) + \Gamma_{Z;k}(\theta) \tag{2.3}$$

$$\Sigma_Y(\theta) = \Sigma_X(\theta) + \Sigma_Z(\theta) \tag{2.4}$$

Proceeding in order, it must be reminded that the decomposition of the panel in common and idiosyncratic components is the baseline concept in Dynamic Factor Models literature; while most of the representations are based on the assumption that the space spanned by the stochastic variable X_{it} - for t given and $i \in N$ - is finite-dimensional, Forni et al. (2015) demonstrate that such an assumption is extremely restrictive and potentially harmful so, as they do in their paper, I relax this assumption hence founding the estimation on a potentially infinite number of principal components.

The results obtained by Forni et al. (2015, 2017) rely on the singularity of the vector X_{nt}, where singularity is obtained when q is small compared to n. The q factors are the q^{th} dynamic eigenvectors of the spectral density matrix $\Sigma_{Y;n}(\theta)$: such factors are identified according to the method proposed by Hallin and Liška (2007) and are assumed to diverge as $n \to \infty$, while the $(q+1)^{th}$ one is bounded. Hallin and Lippi (2013) provide the conditions under which this assumption holds for $q < \infty$.

We can thus say that it exists a representation that may be written as a block-diagonal matrix of one-sided filters $A_n(L)$ of dimension $m(q+1) \times m(q+1)$

$$A_n(L) = \begin{bmatrix} A^1(L) & 0 & \dots & 0 & \dots \\ 0 & A^2(L) & \dots & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & A^m(L) & \\ \vdots & \vdots & & & \ddots \end{bmatrix} \tag{2.5}$$

The $(q+1) \times (q+1)$ blocks of $A^i(L)$ are so that a vector autoregressive representation can be developed, where the VAR operators $I_{q+1} - A^i(L)$ are square summable and fundamental. Such VAR representation of Y_n is allowed by the presence of a full-rank $n \times q$ matrix of constants H_n where the full-rank condition is, as said, sufficient to guarantee the fundamentalness.

The VARMA representation will thus have the form:

$$(I - A(L))Y_t = Hu_t + (I - A(L))Z_t = Hu_t + \hat{Z}_t \tag{2.6}$$

where $\hat{Z}_t$ is, by definition, equal to $(I - A(L))Z_t$ and idiosyncratic. The equation (6.53) shows well the filtering role of $A(L)$ matrix, while the common shocks u_t are loaded via H creating what we will call the level-common residuals e_t.

We have hence decomposed Y in two different components orthogonal to each other, and, as Barigozzi and Hallin (2016) - BH (2016) hereafter - show in their paper, such components may generate volatilities whose pattern may be very different from each other and from the one they show in levels. Over that, is important to notice that there is no reason to think each component to be affected by the volatility generated from it only. Saying differently: volatility generated from the common component may influence both common and idiosyncratic ones. The same is true for what concerns the volatility generated by the idiosyncratic components.

I proceed with estimation following steps in accordance to the standard DFM literature as described earlier. Running the Hallin-Liška (2007) information criterion we can get indication about the number of common shocks associated to the number of diverging eigenvalues of the spectral density matrix.

Barigozzi and Hallin, in the above cited papers published in 2016 and 2017, worked with a dataset in which $N = 90$; I, on the other hand, have the availability of a bigger dataset in which the number of firms is five times as large as their. Despite this fact the results of Hallin-Liška criterion returns me the same number of common shock ($q^T = 1$), confirming - as stated by BH - most of the empirical results on financial returns and asset pricing theory.

2.1 The level-common component

In order to estimate X_t we start by estimating the spectral density of Y_t by means of a lag-window estimator. The spectral density matrix is identified by the formula:

$$\hat{\Sigma}_Y(\theta) = \frac{1}{2\pi} \sum_{k=-T+1}^{T-1} e^{ik\theta} K\left(\frac{k}{B_T}\right) \hat{\Gamma}_{Yk} \tag{2.7}$$

in which $e^{ik\theta}$ is the autocovariance generating function, K is a kernel function and B_T is a bandwidth parameter. With the symbol $\hat{\Gamma}_{Yk}$ we define the k^{th} lag estimated autocovariance and it is equal to $\frac{1}{T}\sum_{t=|k|+1}^{T} T_t Y'_{t-|k|}$.

Once estimated the spectral density matrix we determine the number of q level-common shocks. In order to do that we use the Hallin and Liška (2007) information criterion, which is a data-driven method for the identification.

The so treated spectral density matrix will be decomposed as follows:

$$\hat{\Sigma}_X(\theta) = \sum_{k=1}^{q^T} \hat{\Pi}_Y(\theta)\hat{\Lambda}_Y(\theta)\hat{\Pi}^*_Y(\theta) \tag{2.8}$$

where $\hat{\Lambda}_Y(\theta)$ is the $q \times q$ matrix which dimensions are determined by the first q eigenvalues above calculated. $\hat{\Pi}^*_Y(\theta)$ is the $n \times q$ matrix with the corresponding eigenvectors on the column, where the asterisk denotes that the matrix in transposed, complex and conjugate.

The estimate of the covariance matrix may be obtained by Fourier transformation of $\hat{\Sigma}_X(\theta)$. As known the Fourier transformation is used to turn a function of time into its constituent frequencies and it is invertible, meaning that we can conversely move from time to frequency domain; the formulas for those conversions are:

$$\hat{\Gamma}_X = \int_{-\pi}^{+\pi} e^{ik\theta}\hat{\Sigma}_X(\theta)d\theta \tag{2.9}$$

$$\hat{\Gamma}_X = \frac{\pi}{B_T} \sum_{|h|\leq B_T} e^{ik\theta_h}\hat{\Sigma}_X(\theta_h) \tag{2.10}$$

Since we define $n = m(q+1)$ for some $m \in N$, $\hat{\Gamma}_X$ will result as a block-diagonal matrix of dimension $m(q+1) \times m(q+1)$. The estimation of each block enable us to estimate the coefficients of a VAR of dimension $(q+1)$ that will be the estimator of the autoregressive filter appearing in (6.52), which yields to the vector autoregressive representation

$$\hat{Y} = (I - \hat{A}(L))Y \tag{2.11}$$

Projecting the $\hat{Y}_t$ onto their q largest principal component provide an estimate of the level-common innovation process, formally $e_{it} = \{e_{it} \mid i = (H_n u_t)_i \mid i \in \mathbb{N}, t \in \mathbb{Z}\}$, that is: $\hat{e} = \hat{H}\hat{u}$. $\hat{H}$ is the estimator of the loadings and this may be disentangled from $\hat{u}$ imposing the identification constraint $\hat{H}\hat{H}' = I_{\hat{q}}$.

2.2 The level-idiosyncratic component

The analysis of volatility, typically, is based on the autocovariance structure of some non-linear transform of innovation processes that (6.49) does not readily offers. Forni and Lippi (2010), Forni et al. (2015) provided the frame used above to analyse the volatility coming from the common component, in a similar fashion Barigozzi and Hallin (2016, 2017) centre their paper on the componentwise residuals coming from the idiosyncratic part. In particular BH (2016) designed a two-step procedure that firstly decompose $(I - A(L))Y_t$ in $\hat{e}_t$ and $(I - A(L))Z_t$ in $\hat{v}_t$, this last one named as level-idiosyncratic residuals, then they proposed two approaches aimed to compute predictions over the level-common and level-idiosyncratic elements and to the application of GARCH techniques.

The estimation of the idiosyncratic element follows the level-common one which is, since

$\hat{\Sigma}_Y(\theta)$ and $\hat{\Sigma}_X(\theta)$ are our estimated spectral density matrix, respectively, for the data and for the common component, it imply that the one for the idiosyncratic component may be obtained as:

$$\hat{\Sigma}_Z(\theta) = \hat{\Sigma}_Y(\theta) - \hat{\Sigma}_X(\theta) \tag{2.12}$$

Given the autoregressive formulation (6.58), the estimator of $\hat{Z}$ will be:

$$\hat{Z} = (I - \hat{A}(L))\hat{Y} - \hat{e} \tag{2.13}$$

in which $\hat{\upsilon}_t$ is formally defined as: $\upsilon_{it} = \{\upsilon_{it} \mid i \mid i \in \mathbb{N}, t \in \mathbb{Z}\}$. Reminding that $\tilde{Z}_t = (I - A(L))Z_t$, the process υ_t thus become:

$$\upsilon_t = (1 - c_i(L))\hat{Z}_t \tag{2.14}$$

The representation (6.61) make use of the AR filters $c(L)$ which, as before, are one-sided, square summable and such that every root lies outside the unit circle ($c(z) = 0$). Both idiosyncratic processes are zero-mean second-order white noise and not mutually orthogonal, meaning that at this level some mild cross-correlation among them still remains.

Barigozzi and Hallin (2016, 2017) analyse both e_{it} and υ_{it} via general dynamic factor models, this mean that, as for the baseline representation (6.48), similar assumptions must take place: in particular the existence of a second-order moment for all $i \in N$ as well as spectral density continuous respect to the Lebesgue measure over $[-\pi, \pi]$. In addition, as noted by Engle and Marcucci (2006), in order to ease computations, some transformation must be applied: logarithmic proxy over square residuals would allow to analyse the panels via an additive factor models instead of imposing some positivity constraint at the moment of model estimation.

$$s_{it} = log(e_{it}^2) \tag{2.15}$$

$$w_{it} = log(\upsilon_{it}^2) \tag{2.16}$$

In a similar manner as before, those matrices rely on the assumption of the existence of the second-order moment $[E(s_{it})^2, E(w_{it})^2]$ for all $i \in N$ and for all $n \in N$ and over the assumption of a spectral density absolutely continuous with respect to the Lebesgue measure over $[-\pi, \pi]$.

Denoting with $\tilde{}$ the difference between the s_{it}, w_{it} and their expectation, the generalised dynamic factor model decomposition will return the following results:

$$\tilde{s}_{it} = \chi_{s,it} + \xi_{s,it} = \sum_{k=1}^{q} b_{s,ik}(L)\epsilon_{s,kt} + \xi_{s,it} \tag{2.17}$$

$$\tilde{w}_{it} = \chi_{w,it} + \xi_{w,it} = \sum_{k=1}^{q} b_{w,ik}(L)\epsilon_{w,kt} + \xi_{w,it} \tag{2.18}$$

In this case I will assume the existence of $q_s \in N$ such that the q_s^{th} eigenvalue of the spectral matrix diverges as $n \to \infty$, while the q_{s+1}^{th} is bounded.

Analysing the two elements $\tilde{s}_{it}$ and $\tilde{w}_{it}$ separately would imply some important drawbacks: first of all must be noted that, even though $\xi_{s,it}$ and $\xi_{w,it}$ are orthogonal respect to their common components (respectively to $\chi_{s,it}$ and $\chi_{w,it}$), nothing suggests that they may be orthogonal to each other or respect to any other component of the original panel Y_{it}. This lead us to the conclusion that $\tilde{s}_{it}$ and $\tilde{w}_{it}$ may yield some market driven component and, more important, that treating the two processes separately some loss of information may occur.

Hallin and Liška (2011), in order to deal with it, proposed to aggregate the two panels building a bigger process $\eta = \{\eta_{it} \mid i \in, t \in \mathbb{Z}\}$ which will have block structure and will live on the same assumption as before for what concerns spectral representations and eigenvalues existence.

$$\begin{cases} \tilde{s}_{it} = \chi^s_{\eta,it} + \xi^s_{\eta,it} = \sum_{k=1}^{Q} d^s_{\eta,ik}(L)\epsilon^s_{\eta,kt} + \xi^s_{\eta,it} \\ \tilde{w}_{it} = \chi^w_{\eta,it} + \xi^w_{\eta,it} = \sum_{k=1}^{Q} d^w_{\eta,ik}(L)\epsilon^s_{\eta,kt} + \xi^w_{\eta,it} \end{cases} \tag{2.19}$$

where $Q \in \mathbb{N}$ and such that $max(q_s, q_w) \leq Q \leq q_s + q_w$.

In the (2.19) is clearly visible how all the elements, except for the strongly idiosyncratic ones $\xi_{\eta,it}$ are driven by market volatility shocks $\epsilon = \{\epsilon_t = (\epsilon_{it}, \dots, \epsilon_{Qt})' \mid t \in \mathbb{Z}\}$. From here applying kernel-smoothing techniques we can identify the level-common and level-idiosyncratic volatilities, results of that will be provided in the dedicated paragraph.

2.3 The Bayesian nonparametric methods for volatility

Barigozzi and Hallin, in both their paper of 2016 and 2017, treated the level-common and the level-idiosyncratic component separately in order to compute predictions or to apply GARCH techniques. Even though they seem to obtain significant results from their estimation, the fact that the two elements are not orthogonal to each other may suggest other ways of estimation.

In GARCH models the conditional variance is a deterministic function of model parameters and past data, in stark contrast to that stochastic volatility models treat volatility as a random variable granting more flexibility and a larger space to move for further model specification.

Dealing with stochastic volatility models some difficulties may arise, in particular the estimation of the parameters composing the system is not an easy task. In order to accomplish to that I make use of Bayesian nonparametric techniques which seems the best and better performing method to synthesise the information coming from such heterogeneous data. There are several reasons that pushed me to move in this direction, a surely relevant one is that binding the analysis to some specific parametric form could be limiting, while relaxing the parametric assumptions would allow for a grater ductility and, possibly, to a bigger robustness. As known, nonparametric models rely on distributions moving over an infinite dimensional space. In this frame the difference between Bayesian and classical nonparametrics is that, while in the classical approach the infinite-dimensionality is treated as nuisance - from here the necessity of building procedures able to make inference on the finite dimensional parameter of interest - the Bayesian approach completes the model with a prior on the infinite dimensional parameter, hence developing inferences that account for a full probabilistic description of all relevant uncertainties.

Nonparametric models involve at least one infinite-dimensional parameter which is usually a function or a measure: declining those models in a Bayesian manner and considering the utilisation of mixtures may provide the necessary adaptability to correctly deal with financial time series.

Another argumentation that moved me to estimate our dataset non-parametrically comes from Mahieu and Schotman (1998): it is in their opinion that such models are able to capture features that parametric equivalent models can not fully do. In this sense Dirichlet mixture models offer a flexible alternative as it can be considered an infinite mixture of model where the data itself suggest the number of mixing components.

Dirichlet process (DP hereafter) is one of the most used prior in Bayesian nonparametrics. It configures as an extension of the k-dimensional Dirichlet distribution to a stochastic process and its success is given by the mathematical tractability that eased the derivation of several variation such as Mixture of DP (DPM) and Hierarchical DP (HPD) among the others.

Dirichlet process mixture (DPM) consist of an (up to infinite) mixture of normal distributions whose means, covariances, and mixture probabilities are estimated by applying the relatively uninformative Dirichlet process (DP) prior to the infinite number of unknowns (see Ferguson (1973) and Lo (1984)). Theoretically speaking, the DP prior essentially shrinks the number of unknowns down to just a few important mixture clusters, thus enabling us to overcome the common nonparametric problem of having more unknowns than observations.

The concept of DP prior hence comes with the Bayesian one. As said before, the focus of this paper is to study the volatility components allowing for a time-varying behaviour of them. The models that treat variances in this way are known as Stochastic Volatility (SV) models; the Bayesian framework, along with DP prior is imposed in order to take out the necessary relevancies from volatilities. Dealing with Bayesian inference in the SV models some problem may arise: it is well known, in fact, that the likelihood may often appear in an intractable form. Recently researchers tried different approaches to overcome this problem; among all, it must be cited the work by Creel and Kristensen (2015) who adopted Approximate Bayesian Computation (ABC) algorithm to avoid computing the full likelihood of all available data. As known, it is not possible to simulate directly from the posterior distribution unless it is completely parametrised by a finite number of parameter, hence ABC methods, building likelihoods based on limited information, completely rely on simulation algorithms, thing that comes with a cost in terms of computation.
Kristensen's approach offers some interesting points to reflect on, last but not not least their application to continuous-time processes, nevertheless using a mixture instead of an approximated prior seems to be more adherent to the data and coherent to our aims.

Bayesian nonparametric methods as well are not free from computational costs, nevertheless they offer on the plus side the ability to incorporate uncertainty at level of distribution function, as a consequence algorithms that eventually combine analytic derivation of the distribution function have been successfully developed in the last three decades. I will provide details about the algorithm I used as soon as models with their components will be illustrated.

Applications of Bayesian techniques to Stochastic Volatility models has been introduced by Jacquier et al. (2002) while application of nonprametric methods started to appear since early 2000's. Some of the most relevant articles are by Neal (2000) and Jensen and Maheu (2010, 2014) in which the Bayesian frame has been implemented by the adoption of Dirichlet Process Mixture (DPM).

In this paper I design a Dirichlet Process Mixture with a mass parameter $M > 0$, that somehow reports the grade of trust I have in my prior guess, and a centring distribution H that is equal to $H = N(\mu_0, \sigma_\zeta^2)$ in which μ_0 and σ_ζ^2 are the overall parameters representing, respectively, location and scale, this last one assumed to be constant. Both parameters together represent my prior assumption about the shape of the normal distribution used.

Formally speaking the DPM is described by:

$$\begin{cases} \zeta_t \mid \mu_t \sim N(\mu_t, \sigma_\zeta^2) dG_k(\mu_t, \sigma_t^2) \\ \mu_t \sim G \\ G \sim DP(MH) \end{cases} \tag{2.20}$$

As I will show later, the SV model will be formulated as a log-transformation of a Normal distribution, in consideration of this, and since the scale parameter is grater that zero, G is definable as a probability measure on the half-plane.

This normal mixture may be alternatively written in the form firstly introduced by Ferguson (1973) and Lo (1984), as:

$$f_k(y) = \int N(\zeta \mid \mu, \sigma^2) dG_k(\mu, \sigma^2) \tag{2.21}$$

$$G_k = \sum_{i=1}^{\infty} w_i \delta \mu_i, \sigma_i^2 \tag{2.22}$$

which can represent any continuous distribution in the real line. Here f_k states that this is a k-component mixture model with parameters $(\frac{M}{k}, \frac{M}{k}, \ldots \frac{M}{k})$ which is defined *infinite* as $k \to \infty$, where k is the number of clusters weighted by w_i that represents indeed the mixing weight which $\sum_{i=1}^{\infty} w_i = 1$.

The other parameter in (2.22) is δ_x that represents a Dirac measure that places mass one on any point x.

Combining the likelihood distribution, which is a Gaussian with component parameters are μ and σ_ζ, with the Dirichlet prior I get:

$$\begin{cases} \frac{n_i^{-\lambda}}{\sigma_\zeta(n_{-i,j})} exp\Big[- \frac{1}{2\sigma_\zeta^2}\Big(\bar{\eta} - h_t - \mu_i\Big)\Big] & 2 \leqslant i \leqslant k^\lambda \\ \\ \frac{M}{\sigma_\zeta(n_{-i,j})} exp\Big[- \frac{1}{2\sigma_\zeta^2}\Big(\bar{\eta} - h_t - \mu_0\Big)\Big] & 2i = k^{-\lambda} + 1 \end{cases} \tag{2.23}$$

The equation (2.23) show indeed the conditional posteriors for the indicators as the infinite limit is reached, in which M is the mass parameter of the Dirichlet process and $n_i^{-\lambda}$ is the number of observation excluding the $\lambda - th$ one. I choose to write λ in order to keep a more general notation but, in my model, this is to all intents and purposes a time index.

The main purpose of stochastic volatility models is to describe the time-varying pattern of volatility, following the formulation by Taylor (1994, 2007) - implemented by the presence of a persistence parameter ϕ - to model the volatility by an AR(1) process:

$$\begin{cases} y_t = \beta e^{\frac{h_t}{2}} \epsilon_t \\ h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_\omega \omega_t \end{cases} \tag{2.24}$$

In this formulation ϵ_t and ω_t are assumed to be a normally distributed i.i.d. stochastic processes and h_t to be the log of volatility at time t. The persistence parameter is instead modelled to be $-1 < \phi < 1$ in order to guarantee stationarity and identifiability of μ.

Linearising the above written model (2.24) I obtain $\bar{\eta}$ in which returns are in logarithmic form. h and ω are generally unobserved in Stochastic Volatility models, nevertheless I will make use of DFM results in order to get fully specified time series while estimating SV parameters as to improve the model forecasting capabilities.

$$\begin{cases} \bar{\eta_t} = h_t + \zeta_t \\ h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_\omega \omega_t \end{cases} \tag{2.25}$$

Since y is a square quantity $\bar{\eta_t}$ may be approximated with $\bar{\eta_t} = log(\eta_t + c)$ and ζ with $\zeta_t = log(\epsilon_t^2)$, this last one distributed as an F.

The presence of c is justified by A. (1996) which illustrate that in this kind of data $\bar{\eta_t}$ several times may reach zero, hence taking logs may no longer guarantee the distribution to be normal. For this, imposing an offset parameter (c) which act to modify the log transformation of the variable of interest, will produce a variable with smaller variance and more robust against non-normality. Fuller in his book states that a small fraction, about 0.02 of the average value of η would satisfy our needs, Kim et al. (1998) imposed $c = 0.001$ while more detailed arguments has been exposed by Sakaria and Griffin (2017) which demonstrated that in this kind of models for values smaller than 10^{-9} the sample mean and variance gt close to their real values.

Delatola and Griffin (2011), instead, imposed different values to c, going from $c = 10^{-7}$ to $c = 10^{-3}$, for their analysis noting that not always smaller is better but sometimes, in order to improve the predictive performances, higher values of the offset parameter must be chosen.

By modelling the returns as $\bar{\eta}_t$ instead of y we will observe several crossing of the zero line. This situation has extensively been treated in literature and often approached imposing an alternative parametrization to (2.25) with the aim of improving MCMC efficiency and statistics reliability.

Such strategy take place from the Bayesian corollary to the Basu's theorem on the independence of (complete) sufficient and ancillary statistics.

In fact, as perfectly explained by Yu and Meng (2011), the notions of sufficient statistics and ancillary statistics are mathematically equivalent to the centred and non-centred parametrization respectively. In a data augmentation scheme, as can be considered MCMC, flanking the two parametrization may boost the convergence of the algorithm while reducing the markovian dependence and solving eventual data missing problem.

A centred version of the Stochastic Volatility model I adopt may be:

$$\begin{cases} \bar{\eta_t} = \bar{h_t} + \bar{\zeta_t} \\ \bar{h_{t+1}} = \phi \bar{h_t} + \sigma_\omega \omega_t \end{cases} \tag{2.26}$$

where $\bar{\zeta_t} = \zeta_t + \mu$ and $\bar{h_t} = h_t - \mu$, this last one is normally distributed around zero: $\bar{h_t} \sim \mathcal{N}(0, \sigma_\omega^2/(1-\phi^2))$.

Facing this issue Yu and Meng (2011) propose an algorithm which interweave going back and forth between the two parametrization at each iteration of the MCMC sampler, another strategy has been adopted by Delatola and Griffin (2011) which jump from one model to the other on a probabilistic base according to the formula:

$$p(z_t) = W * centred + (1 - W) * Noncentred \tag{2.27}$$

where W is the probability the log-returns lay on the zero line.

For the computation of the full conditional posterior, for both parametrization, I will mix likelihood, which is assumed to be Normal, with priors of different shapes. The priors in question are: $p(\phi)$, $p(\mu)$, $p(\sigma_\omega^2)$ and $p(\sigma_\zeta^2)$, where the last two are the variance parameters associated to ω and ζ. Considering that all those priors are independent, Bayes' rule can be applied implying that the full posterior may be calculated as:

$$p(\mu \mid y, h, \phi, \sigma^2_\omega) \propto p(h \mid \mu, \phi, \sigma^2_\omega) p(\mu)$$

$$p(\sigma^2_\omega \mid \bar{\eta}, h, \mu, \phi) \propto p(h \mid \mu, \phi, \sigma^2_\omega) p(\sigma^2_\omega)$$

$$p(\sigma^2_\zeta \mid \bar{\eta}, h, \mu, \phi) \propto p(h \mid \mu, \phi, \sigma^2_\eta) p(\sigma^2_\zeta)$$

$$p(\phi \mid \bar{\eta}, h, \mu, \phi, \sigma^2_\eta) \propto p(h \mid \mu, \phi, \sigma^2_\eta) p(\phi)$$

Such priors are distributed as: $p(\mu) \sim \mathcal{N}(\alpha_\mu, \beta_\mu)$, $p(\sigma^2_\omega) \sim \mathcal{IG}(\alpha_{\sigma_\omega}, \beta_{\sigma_\zeta})$, $p(\sigma^2_\zeta) \sim \mathcal{IG}(\alpha_{\sigma_\zeta}, \beta_{\sigma_\zeta})$ and $p(\phi)\mathcal{N}(\alpha_\phi, \beta_\phi) I_{(-1,+1)}(\phi)$.

Updating each parameter, starting by μ, using the standard Bayesian techniques will return the full conditional posteriors that follow:

$$\propto \left\{ \frac{\sigma^2_\omega}{(n_{-i} - 1)(1-\phi)^2 + (1-\phi^2)} \right\} \left\{ \frac{h_1(1-\phi^2)}{\sigma^2_\omega} + \frac{(1-\phi)}{\sigma^2_\omega} \sum_{t=1}^{n_{-i}} (h_{t+1} - \phi h_t)^2 \right\} \tag{2.28}$$

the full conditional distribution for σ^2_ω is:

$$\propto exp\left(-\frac{1}{2} \frac{2\beta_{\sigma_\omega} + (h_1 - \mu_0)^2 (1-\phi^2) + \sum_{t=1}^{n_{-i}} (h_{t+1} - \mu_0 - \phi(h_t - \mu_0))^2}{\sigma^2_\omega} \right) \left(\frac{1}{\sigma^2_\omega} \right)^{(\alpha_{\sigma_\omega} + \frac{n_{-i}}{2}) + 1}$$

Which may alternatively written as:

$$\sigma_\omega^2 \mid h_1 \ldots h_n, \phi \sim \mathcal{IG}\left(2.5 + \frac{n_{-i}}{2}, \beta^*\right) \tag{2.29}$$

where $\beta^* = 0.025 + \frac{(h_1-\mu_0)^2(1-\phi^2)+\sum_{t=1}^{n-i}(h_{t+1}-\mu_0-\phi(h_t-\mu_0))^2}{2}$.

Similarly the conditional distribution for σ_ζ may be handled to return:

$$\sigma_\zeta^2 \sim \mathcal{IG}\left(\frac{n_{-i}+k}{2}, \frac{1}{2}\left[\sum_{i=1}^{n_{-i}}(\bar{\eta}_i - \mu_{c,i})^2 + \sum_{i=1}^{k}(\mu_i - \mu_0)^2\right]\right) \tag{2.30}$$

Lastly, the full conditional for ϕ can be made explicit as:

$$\propto exp\left\{-(\bar{\eta}_i - \frac{(h_1-\mu_0)(1-\phi^2)}{2\sigma_\omega^2} - \frac{\sum_{i=1}^{n-i}(h_{i+1}-\mu_0-\phi(h_t-\mu_0))^2}{2\sigma_\omega^2}\right\} exp\frac{-(\phi-\alpha_\phi)}{2\beta_\phi^2} I_{(-1,+1)}(\phi)$$

which can be solved into:

$$\phi^* \sim \mathcal{N}\left(\frac{\sum_{i=1}^{n-i}(h_{t+1}-\mu_0)(h_t-\mu_0)}{\sum_{i=1}^{n-i}(h_t-\mu_0)}, \frac{\sigma_\omega^2}{\sum_{i=1}^{n-i}(h_t-\mu_0)}\right) \tag{2.31}$$

The above written formulas are valid for the non-centred parametrization, the centred one works according to the same logic. Solution details for both parts will be provided in Appendix A.

As known stochastic volatility models rely on likelihoods which integrals are not analytically tractable, plus Bayesian inference is complicated by the fact that this is a non-linear state space model and the number of mixtures is unknown.

Techniques to solve this problem make use of filtering or quasi maximum likelihood methods, all of those adopting MCMC processes to update the parameters. In particular Jacquier et al. (2002) proposed a one-at-the-time updating process to revise volatility, nevertheless incurring in critiques motivated by the highly correlated nature of the samples; Jensen and Maheu (2010), by their side, proposed an alternative method in which sampling is computed by blocks demonstrating a better mixing properties.

Among all, I choose to use the algorithm created by Kim et al. (1998) in which the log-volatilities are simultaneously updated using what, since Carter and Kohn (1994), is called forward filtering backward smoothing (FFBS) algorithm. In this frame a Gibbs acceptance sampler is applied to a target distribution respect to an instrumental one.

MCMC applied to this scheme is developed to sample this density without computation of the likelihood function. The algorithm proceed by sampling each parameter as follows:

- Initialize all parameters:
- Sample $h \mid \bar{\eta}, \mu, \phi, \sigma^2_\omega, \sigma^2_\zeta, M, s$
- Sample $s \mid \bar{\eta}, \mu, \phi, \sigma^2_\omega, \sigma^2_\zeta, M, h$
- Sample $\sigma^2_\zeta, \mu, M \mid \bar{\eta}, h, \phi, \sigma^2_\omega, s$
- Sample $\phi, \sigma^2_\omega \mid \bar{\eta}, \mu, \sigma^2_\zeta, M, s, h$

In this algorithm, the first three steps are standard and quite straightforward, the last two, on the other hand, are justified by an attempt of reducing the correlation between the log-volatility h and the parameters $(\mu, \phi, \sigma^2_\omega)$. For details may be seen Delatola and Griffin

(2011), Gelfand et al. (1995) and Papaspiliopoulos et al. (2007).

Chapter 3

Data and estimation results

For doing the estimations I used data from Standard Poor's 500 (SP500) company value recorded with daily frequency from the beginning of 2008 to January 23^{th} 2020. The total sample result of a total of 3144 observations.

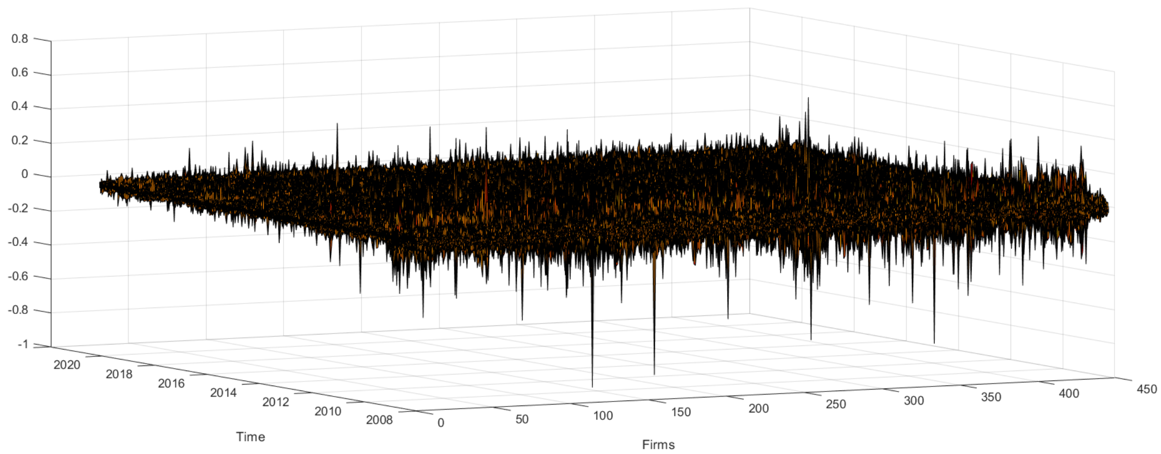

Figure 3.1: SP500

On the cross-section side the panel was originally composed by 505 observation nevertheless, since not all time series were complete in the considered period, I did not take into account those company with missing observations along the time horizon, hence remaining with a total of 446 firms. In the figure (3.1) can clearly be observed how the panel is composed, underlying how data have been taken in log-differences, so the vertical axis rate of variation is measured.

In order to check the out-of-sample forecasting performance of my method, I decided cut the last one hundred observation; this is showed in the figure (3.2) in which I only sketched the SP500 index as provided in the database and where the black vertical line indicate the moment when the cut occurred.

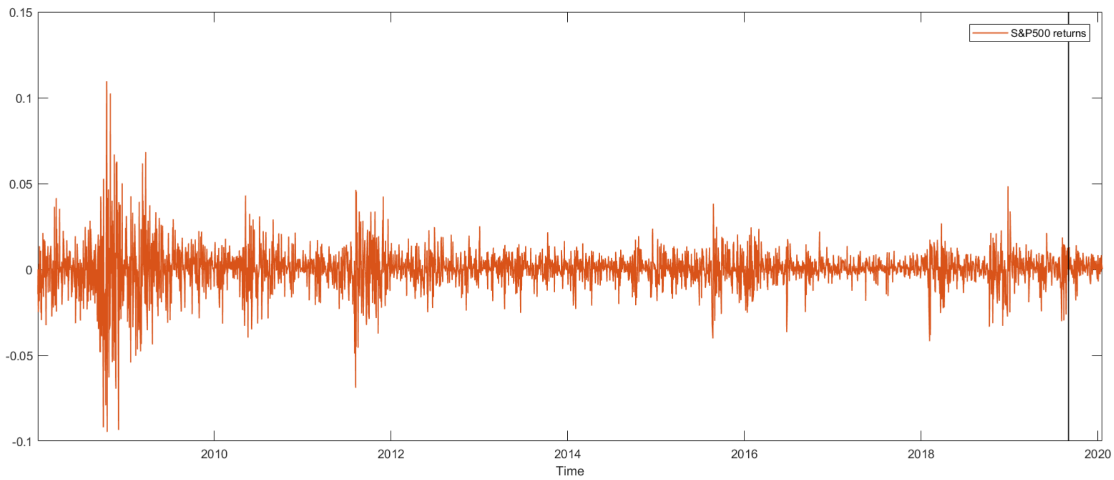

Figure 3.2: S&P500 aggregate index

I proceeded with estimation following steps in accordance to the standard DFM literature as described earlier. Running the Hallin and Liška (2007) information criterion I could get indication about the number of common shocks associated to the number of diverging eigenvalues of the spectral density matrix.

Barigozzi and Hallin, in the above cited papers published in 2016 and 2017, worked with a dataset of a cross-section dimension of $N = 90$; I, instead, have the availability of a bigger dataset in which the number of firms is five times larger than theirs. Despite of this the results coming from Hallin-Liška criterion returns me the same number of common shock ($q^T = 1$), confirming - as stated by BH - most of the empirical results on financial returns and asset pricing theory.

Once obtained level-common and level-idiosyncratic volatilities I applied kernel-smoothing techniques with a bandwidth of 15 working days - 3 calendar weeks - which produced the figures (3.3) and (3.4) respectively.

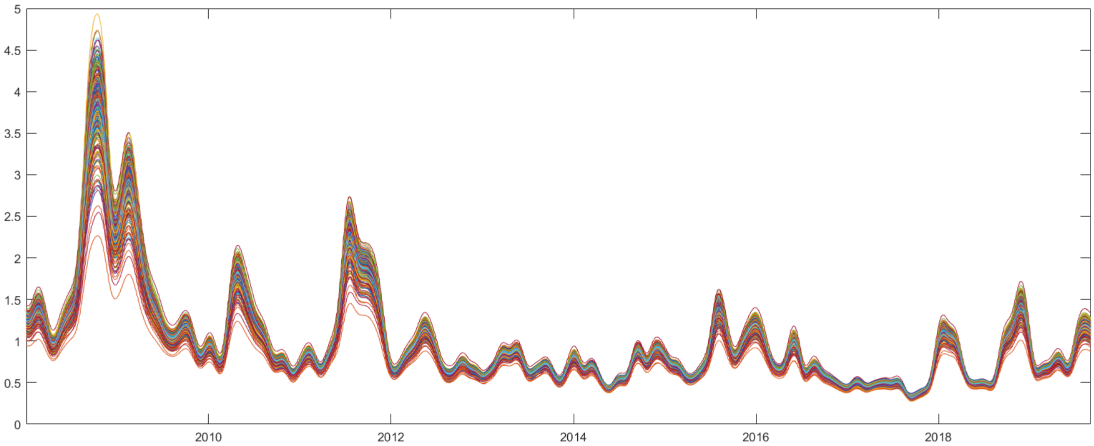

Figure 3.3: Level-common volatility

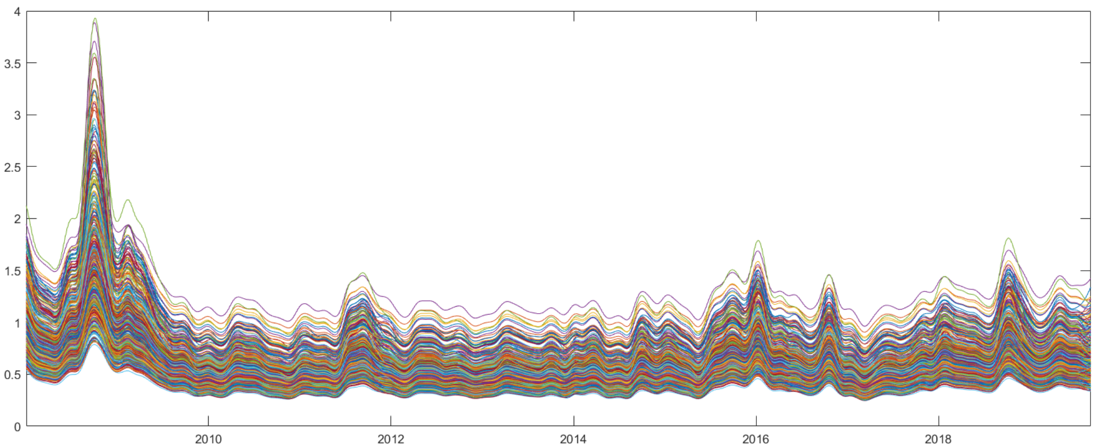

Figure 3.4: Level-idiosyncratic volatility

Observing those graphs we can clearly see how, in accordance with figure (3.2), the volatility increases in correspondence of the beginning of the financial crisis, started by the bankruptcy of Lehman brothers (autumn 2008), in 2011, when the EU member states were object of a speculative attack which required the intervention by European Central Bank, while 2016 records reports, with all probability, the effects of brexit referendum. Those results are perfectly in line with our expectation and confirm the findings in Barigozzi and Hallin (2016) who pointed out the different behaviour of the two curves according to the different nature of the shocks.

Level-common and level-idiosyncratic curves obtained from Dynamic Factor Model estimation are used as time series to be inputted in equation (2.26) as measures of volatility. In this case, in order to have a full specification of the model, only parameters need to be estimated and this is what I have done adopting a Bayesian methods. Here is another element of innovation of my approach as, differently from classical models where only σ_ω is estimated, I can actually input a measure of ω - level-idiosyncratic volatility - hence tuned by the estimation of the variance parameter. The aleatory element of the SV is generally taken as a Gaussian white noise: being able to get an accurate representation, as level-idiosyncratic volatility is, all along with σ_ω have a significant impact on the general estimation output. Operating this way, it is in fact possible to get a greater adherence to data, in particular, in periods characterised by higher volatility, hence improving the overall forecasting performance.

Looking back to figure (3.1) we can have an idea about how the panel is composed and about the heterogeneity and the complexity of the dataset; this justify the choice of a

Bayesian nonparametric approach that have been applied to the whole dataset in a procedure counting ten thousand iteration, in which the first three thousand was discharged.

The algorithm demonstrated to be quite sensitive to the starting values of the parameters however, once calibrated, it showed good predictive performance. In particular, I argued in the previous paragraphs about the choice of the offset parameter; since Griffin treated the dataset in a different manner as I did, it is not straightforward a comparison between our choices, anyway for my estimation I set it equal to $c = 2$ as it demonstrated to produce the best output. Starting values for location and scale parameters of the inverse gamma have been set equal to 2.5 and 0.025, for σ_ζ a value of 0.5 was imposed while for ϕ mean and variance was put equal to 0 and 1 according to theory.

The computation of the whole program resulted to be quite demanding in terms of time. Once the algorithm came to a solution the posterior median obtained for the persistence parameter ϕ is 0.98645, really close to its posterior mean 0.98636, implying an high level of persistence however not violating the stationarity assumptions. σ_ω posterior median have been estimated equal to 0.22569 while μ resulted to be equal to 1.526. Once again posterior means did not fall too far from their median values.

Table (3.1) reported below, displays the full set of estimated values.

Combining estimated parameter with DFM results, and coding a rolling window algorithm, I did forecasting over volatility.

For sake of clarity, I ran the forecasting procedure on a previously averaged the level-common and level-idiosyncratic volatility curves. Operating this way my results can be matched with

Parameter	Posterior median	Description
ϕ	0.98645	Persistence parameter
σ^2_ω	0.22569	Variance of the volatility equation
μ	1.526	Mean of the volatility equation
σ^2_ζ	8.5768	Variance of the returns equation
M	0.63801	Mass parameter
k	6	Number of clusters
W	0,0018566	Zero return probability

Table 3.1: Parameter estimated

the S&P500 composite index reported in figure (3.2); I made this choice because I find this to be an easier way to demonstrate my points, nevertheless the procedure may obviously be generalised to the multivariate case or some specific firm may be isolated as well.

Figure (3.5) below represents the forecasted volatility, in which it is possible to observe the evolution of the curve. As usual, the black vertical line separates the in-sample (left-hand side) forecast to the out-of-the-sample forecast (right-hand side). It is important at this stage to record a close adherence of this curve with the one described by level-common volatility reported in figure (3.3).

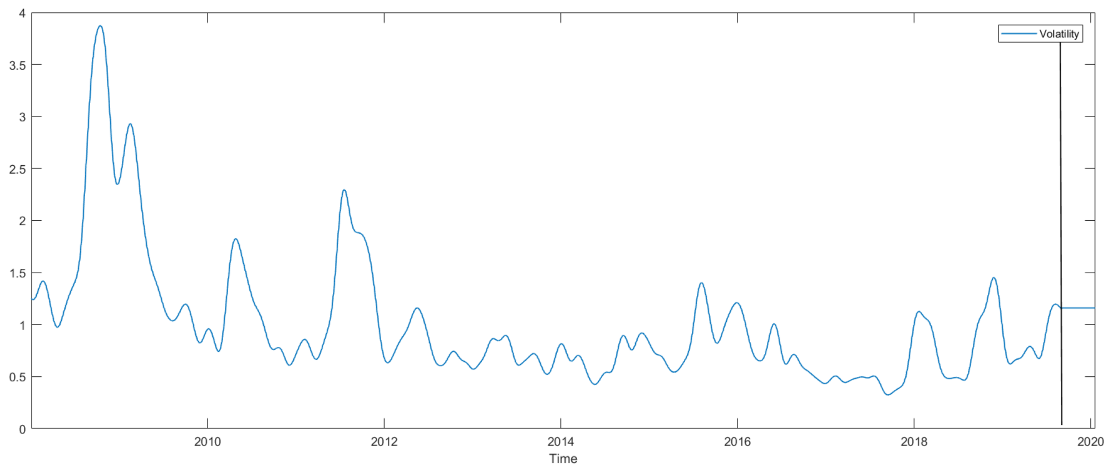

Figure 3.5: Forecasted volatility

The transformation of the forecasted log-volatility to returns produces a curve that is quite similar to the one in panel (3.2); saying differently: comparing the dynamics of S&P index with one of the generated curve we may recognise the same path made by periods characterised by a higher volatility and periods in which the market is calmer. Figure (3.6) reports both forecasted volatility and returns to observe graphically the relation that links the two curves.

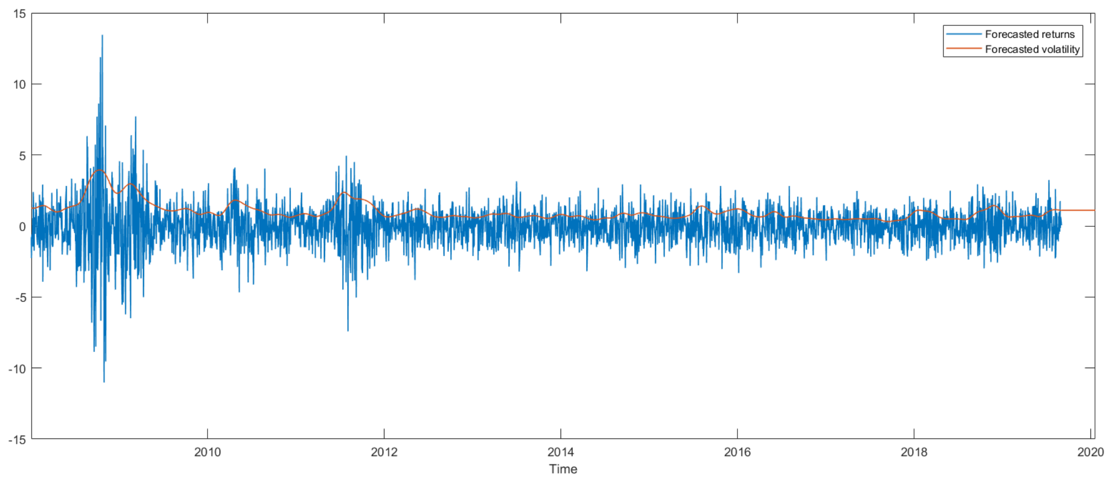

Figure 3.6: Forecasted volatility and returns

Overlapping observed and forecasted shows us the effective match of those curves highlighting the point in with they diverge the most. Figure (3.7) does that.

Looking at figure (3.7) one thing must be clarified: recalling that my model is designed by the formula (2.24), differently from the other panels where the value of β have been taken constant and equal to 1, in this last one, in order to catch the volatility clustering, I set it equal to 1.1, 2.2 or 0.35 depending by the standard deviation of the whole volatility time series. More in detail β will be equal to 0.35 if the volatility stays below the standard deviation, while it will be equal to 1.1 if the volatility overtakes the standard deviation and equal to 2.2 if it is greater than two times the standard deviation.

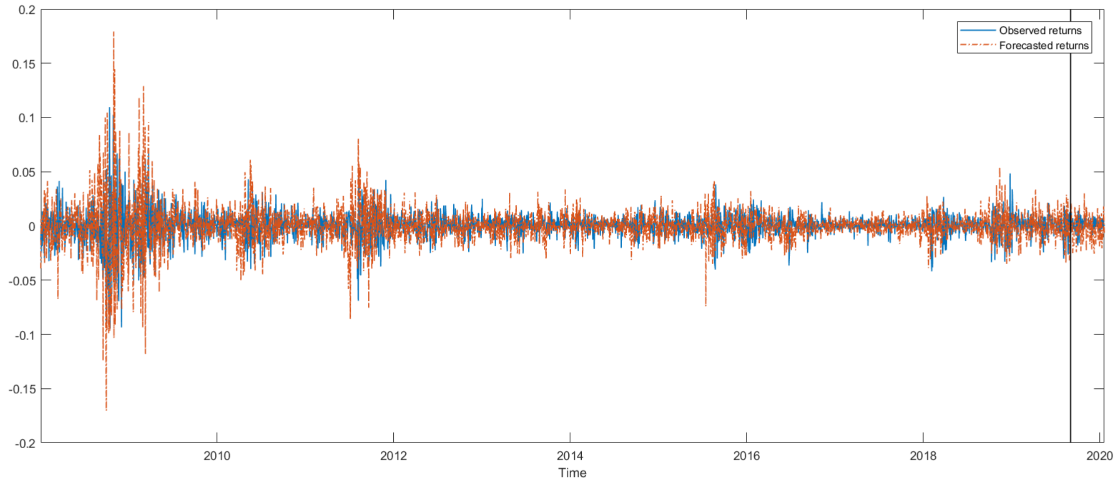

Figure 3.7: Forecasted and observed returns

Analysing the panel it is possible to confirm the model to perform pretty well as it is demonstrated by the Root Mean Square Error (RMSE) values that it produces.

To assess the forecasting performances of my model I tested it over the full sample, as to say from the beginning of the series until mid January 2020, obtaining a value slightly over 2%. Those result are confirmed by checking for the in-sample performance of the model while the out-of-the-sample forecast perform even batter dropping the RMSE value around 1.5%.

In order to check for the forecasting performance in the period of maximum volatility, I isolated the period between mid September 2008 and mid July 2009, hence the RMSE returned a value around 5% confirming a general good performance of my approach however worsening the outcome. It is so evident that the model seems to perform better in periods with less volatility, however it is still clear how the choice of β have been demonstrated to be crucial, thus a more punctual/flexible approach to its determination should be considered to gain a better adherence of model to data. This could be material for future works.

Chapter 4

Conclusions

High-dimensional financial data are generally big datasets characterised by heterogeneous time series. In order to deal with such a complex panels I adopted infinite dimensional Dynamic Factor Models (DFM), to extract volatilities, and Bayesian non-parametrics techniques to estimate the parameters of a pretty much standard Stochastic Volatility model. The final goal of my work was to produce a feasible measure of volatility and a good parameter estimation in order to get a multiperiod forecast for both volatility and returns.

The Bayesian non-parametric estimation have been developed as an infinite mixture Dirichlet process and returned parameters values in line with standard financial theory, while a two-step general Dynamic Factor Model was adopted to extract level-common and level-idiosyncratic volatility.

The applied exercises worked using S&P500 daily data spanning over 12 years while the forecasts was computed with a rolling window algorithm which returned values for both stock market volatility and returns.

The approach demonstrated to perform pretty well, however a further element of flexibility was imposed in order to get forecasts more adherent to data. Given the baseline model described in (2.24) I modelled β to vary with the volatility as to intercept the that phenomenon of volatility clustering that, as in general as in my case, characterise the financial time series.

The results showed good adherence of the forecasted volatility and forecasted returns with actual realisations demonstrating good overall performances of my approach. Several experiments was conducted splitting the samples and running running separated tests: in any case the model offered good forecasting properties both in-sample and out-of-sample however suffering periods with a bigger variability. This suggest the adoption of different, more accurate way to model β opening the for future works.

The general results surely was positive and offer my approach as an effective way to deal with large, high-frequency datasets.

Chapter 5

Introduction

In the last two decades data collection, aided by an increased computational capability, has considerably increased both dimension and structure of the datasets; given this, statisticians and economists may today work with time series of remarkable dimension which may come from different sources. Dealing with such datasets may not be so easy and requires the development of *ad hoc* mathematical models.

One of the targets that we need to achieve designing new models is to contrast parameters proliferation and facilitate models reduction. In order to do that, factor models represent an effective tool since they are able to synthesise information held in huge datasets in a few factors.

Among those methods, Dynamic Factor Models (DFM) represent one of the newest techniques in big data management. In particular, using DFM, it is possible to derive a representation for an infinite panel of time series which is the analogous of the Wold representation for finite dimensional ones. This is the generalised dynamic factor representation and, between its properties, has the one of keeping the time series structure of the dataset thus generally is finalised to do forecasting or structural analysis, where the object of interest are the impulse response functions.

Dynamic Factor Models are one of the possible approaches created to extract statistics out of large datasets, another interesting one surely is the utilisation of Bayesian methods while frequentist alternatives are extensively used in machine learning.

When we talk of big data we generally refer to situations in which the number of explanatory variables is large compared to the sample size. When in this situation we could say that the researcher is in a position in which he/she is trying to estimate too much relatively to the information in his/her possess. This is not unfeasible but will typically lead to a very imprecise inference while it is not even doable, at least with conventional methods, when the number of explanatory variables exceeds the number of observations and when a considerable number of these observations are equal to zero. In mathematics a so composed matrix is said to be *sparse*.

I am not going in the details of the sparsity concept here, for what interests us we only need to underline how sparsity requests specific tools to be approached; such tools may be statistical models in which only a relatively small number of parameters (or predictors) would play an crucial role in the description of the whole dataset. Machine learning models operate according to this logic and lasso (Least Absolute Shrinkage and Selection Operator) is a clear example of that while other examples may be traced in ridge regressions, support vector machine and stochastic search models.

As an example, Lasso is a frequentist model which basically rely on a penalisation algorithm in order to regularise ("shrink") coefficients, making it possible to deal with high-dimensional data and to avoid over-fitting. By contrast this, just like other machine learning techniques, have some considerable drawbacks: first it is not possible to make any choice over the variables considered relevant for the model, second the selection procedure depends by some penalisation parameter which need to be somehow imposed, as a last point it must be noticed how such selection process would determine, one way or another, some loss of information.

Somehow in contrast with the concept of sparsity there is the one of *density*. Among the ones that work over density it is possible to enumerate DFM and Bayesian methods which provide an alternative approach to handle big data inconvenient. The intuition that is behind the application of those methods in high dimensional data context is the following: dealing with a big amount of data the researcher may incur in over-parametrization/over-fitting problems, in such situation the application of a prior distribution to update the observed data would work out to restrict the parametric space.

Bayesian methods have been demonstrated effective both in case of models with many explanatory variables and in models with many dependent variables (as in the VAR case). Analytically speaking the posterior distribution will be obtained by updating observed data with prior information over the parameter of interest, while controlling the prior variance will give to researcher some control over the degree of shrinkage or, in other words, will put some numeric value over the degree of confidence that the researcher has about his beliefs.

If on one hand the Bayesian approach has demonstrated to have some interesting aspects regarding inference and forecasting performances, on the other hand such approach, and thus results, is conditional to the accuracy of the prior choice, plus this technique often imply the utilisation of MCMC processes and other computationally demanding algorithms.

Sparsity-based techniques have become widely used in econometrics, for an extensive survey, with a particular focus on policy evaluation, may be seen Crato and Paruolo (2019), however the adaptability and the performances demonstrated by density-based approaches pushed in this direction. For a quick, although intensive, survey over the use of density in econometrics, with particular reference to Bayesian econometrics, the reader can be referred to, among all, Otrok and Whiteman (1998), Del Negro and Otrok and in particular to great part of the scientific production by Mike West, while, for what concerns the use of Bayesian analysis as shrinkage method Koop and Korobilis (2010) is seminar.

Dynamic Factor Models estimation developed faster and faster since the pioneering work of Stock and Watson (1989), in which they used factor analysis to compute coincident indicators for the business cycle and financial variables. Since factor models are used to consistently estimate common and idiosyncratic components of macroeconomic variables, DFM has been used to do predictions (see Stock and Watson (2002)) or to compute VARs and impulse response functions; the study of the link between factor modes and structural VARs has been deepened by Stock and Watson (2005a) and Forni et al. (2009).

The general idea of Dynamic Factor Models is to reduce a large panel of data in two mutually orthogonal components: the common one, which is strongly correlated with the rest of the panel, and an idiosyncratic one, which is only mildly cross-correlated.

Still about DFM, all the papers we refer here make use of non-parametric techniques in which factors are estimated through a procedure that calculate eigenvalues and eigenvectors from a spectral density matrix, hence essentially realising an approach that is a dynamic version of the principal component one.
Forni et al. (2005) implemented a one-sided estimation of the generalised dynamic factor models in order to improve forecasting performances, the infinite dimensionality, instead, have been introduced in the papers "The Generalized Dynamic Factor Model with infinite dimensional factor space" by Forni et al. (2015, 2017), while deepening the study of the volatility generated by a DFM, Barigozzi and Hallin (2016, 2017) proposed a method to isolate level-common and level-idiosyncratic volatility and Barigozzi et al. (2019) used the block structure of factors - as introduced by Hallin and Liška (2011) - to investigate the interaction those elements.

Talking about modelling in economics, Dynamic Stochastic General Equilibrium (DSGE) models are an interesting instrument. Such a models basically reproduce an economic system and thus may be adopted to do policy analysis or forecasting. Once a DSGE is opportunely linearised, it behave all in all like a VAR. Taking advantage of this (Structural)-VAR have been pulled over to DSGEs in order to do counterfactual checking and validating this last ones

In the literature there are examples in which factor models are applied all along with VARs: they are known as Factor Augmented VAR (FAVAR) (see Bernanke et al. (2004) for an example), however all of those DFM rely over parametric versions, no much have been done non-parametrically.

Since my approach rely over non-parametric, infinite-dimension dynamic factor model, here may already be traced an element of innovation of this work. Further detail of my contributions will be depicted later on.

The great success of the DSGE models is that they combine the macroeconomic dynamics of the time series with a rigorous microeconomic foundations. Those characteristics opened the way to the estimation of as to become an effective tool for Central Banks. Seminar is the paper by Smets and Wouters (2002) which became the reference model for the European Central Bank which uses similar fashion models to analyse the economy of the Eurozone as a whole. Over this baseline, other bigger and more complex models took form, among all must be cited Adolfson et al. (2007) which enlarged the model accounting for the international market and introduced Bayesian techniques for parameter estimation.

The international markets are a crucial factor in every economy as they may influence agents in several ways. Despite the importance of this element the variables composing such a block have always been taken only theoretically, in other words there is not yet, at the best of my knowledge, an effective way to bring model to data in this market.

My approach aim for filling this gap by using non-parametric Dynamic Factor Models techniques. In particular I will put in play an infinite dimensional generalised dynamic factor representation, firstly introduced by Forni et al. (2015) and Forni et al. (2017), in order to estimate the rest of the world time series to plug into a small open economy DSGE. In order to do that a block-wise representation of the DFM is requested (see Hallin and Liška (2011)) hence the computation of Impulse Response Functions (*IRF*) is straightforward.

Since I calibrate the model for the Italian economy, to get a more punctual estimation of the foreign sector time series dynamic factor models will be an essential tool. In particular those models allow me to correctly identify the rest of the world respect to Italy instead of using a general world time series as estimated by official statistics.

The model validation, with particular concerns of the foreign sector, is another target of my study.

Validation through VAR is typically carried out by comparing the empirical impulse response functions with the theoretical ones obtained by DSGE. Fundamentalness is a weak point of VAR analysis; however the adoption of a Dynamic Factor Models significantly simplify the identification problem, providing a satisfactory solution and offering an additional support for the adoption of dynamic factor analysis in this framework.

Chapter 6

The model

The crisis of the recent years made clear how business cycle movements no longer abstracts from international facts; the inclusion, in the most recent models, of international markets capable to simulate the propagation mechanism of rest of the world dynamics is hence crucial.

For my work I start by the now standard set-up of the large-scale New Keynesian Dynamic Stochastic General Equilibrium (DSGE) model characterised by a Small Open Economy (SOE hereafter) framework, developed by Adolfson et al. (2007) and Christiano et al. (2011). In this framework the foreign sector is exogenous with respect the domestic economy and its evolution is described by a Structural Vector AutoRegressive system (SVAR) and, in order to bring model to data, I will estimate foreign variables through Dynamic Factor Models (DFM) hence matching the DSGE's Impulse Response Functions (IRF) and the DFM's ones with the aim of validating the theoretical model. IFRs generated by dynamic factor models enjoy of peculiar properties which encourage the adoption of this tool in this framework, details of which will be exposed in the apposite chapter.

Opening the economy we allow households to make savings both in domestic and foreign bonds while the final goods are produced by specific firms that combine domestic homogeneous goods with imported goods. In order to allow for incomplete exchange rate pass-trough import and export by including price rigidities.

The other main difference in this kind of models are in fiscal and monetary authority block, which must take into account the foreign variables, and in the presence of a trade balance sector which describes the net foreign asset position of our economy.

6.1 Households

The economy is populated by a continuum of households indexed by $j \in (0,1)$, which attain utility from consumption, leisure and money holding. The households possess the economy's physical capital and determine the rate at which this stock is accumulated or utilised.
The generic j^{th} preferences are given by:

$$U_t = E_0^j \sum_{t=0}^{\infty} \beta^t \left[\zeta_t^c \log\left(C_{j,t} - bC_{j,t-1}\right) + A_q \frac{(\frac{Q_{j,t}}{z_t P_t})^{1-\sigma_q}}{1-\sigma_q} + \zeta_t^h A_L \frac{(h_{j,t})^{1+\sigma_L}}{1+\sigma_L} \right] \quad (6.1)$$

In the utility function b captures habit in consumption while ζ^c and ζ^h are positive parameters which bring in the function the preference shocks effects and are assumed to be an AR(1) i.i.d. process, hence having the generic i^{th} form:

$$\zeta_t^i = \rho_{\zeta^i} \hat{\zeta}_{t-1}^i + \xi_{\zeta^i,t}$$

The last term of the utility accounts for the utility coming from leisure, where σ_L denote the inverse Frisch elasticity, while the second element is the real cash balance scaled by a permanent technology shock z_z.

Particular attention must be payed to the first term of (6.1): that is the aggregate consumption $C_{j,t}$ and, since we are in a open economy, it is given by a Constant Elasticity of Substitution *CES* index of imported ($C^m_{j,t}$) and domestically produced ($C^d_{j,t}$) goods.

$$C_{j,t} = \left[(1-\omega_c)^{\frac{1}{\eta_c}}(C^d_{j,t})^{\frac{\eta_c-1}{\eta_c}} + \omega_c^{\frac{1}{\eta_c}}(C^m_{j,t})^{\frac{\eta_c-1}{\eta_c}}\right]^{\frac{\eta_c}{\eta_c-1}} \tag{6.2}$$

where η_c is the elasticity of substitution between domestic and foreign goods.

In order to determine the aggregate price level we can maximize (6.2) subject to the following budged constraint:

$$P^c_t C_{j,t} = P_t C^d_{j,t} + P^{m,c}_t C^m_{j,t} \tag{6.3}$$

where P^c_t, P_t and $P^{m,c}_t$ are aggregate consumer price index (CPI), the prices of domestically produced goods and the prices of imported goods respectively.

Solving the first order condition, and after some manipulation, it is possible to define the demands for domestically produced and imported consumption goods:

$$C^d_{j,t} = (1-\omega_c)\left[\frac{P_t}{P^c_t}\right]^{-\eta_c} C_{j,t} \tag{6.4}$$

$$C^{m}_{j,t} = \omega_c \left[\frac{P^{m,c}_t}{P^c_t} \right]^{-\eta_c} C_{j,t} \tag{6.5}$$

Here, as in all the formulas below, I define, ω as the generic share of imports and η as the elasticity of substitution between domestic and imported goods; those parameters may refer to consumption or investment at the occurrence.

Aggregate consumer price index will thus be:

$$P^c_t = \left[(1-\omega_c)(P^{1-\eta_c}_t) + \omega_c (P^{m,c}_t)^{1-\eta_c} \right]^{\frac{1}{1-\eta_c}} \tag{6.6}$$

Similarly, investment may be defined by a *CES* which aggregates domestic and imported components. Before that, recalling how households own physical capital k_t and setting an exogenous capital destruction rate δ_k, let's define the law of motion according to which it accumulates:

$$k_t = (1-\delta_k)k_{t-1} + \Gamma_t \left(1 - S_k \left(\frac{I_{k,t}}{I_{k,t-1}} \right) \right) I_{k,t} \tag{6.7}$$

As in Christiano et al. (2005), I assume that in the deterministic steady-state there are no capital adjustment costs, hence the installation technology is $\left(1 - S_k \left(\frac{I_{k,t}}{I_{k,t-1}} \right) \right)$ with $(S_k(1) = S'_k(1) = 0)$ and a function that is concave in the neighbourhood of that deterministic steady-state $S''_k(1) = 1/k_k > 0)$. Γ_t is the investment-specific technology shock that follows an autoregressive process given by:

$$\hat{\Gamma}_t = \rho_{\Gamma_t} \hat{\Gamma}_{t-1} + \xi_{\Gamma,t} \tag{6.8}$$

Remembering the definition of ω_i and η_i, we are now ready to specify the *CES* function as follows:

$$I_{j,t} = \left[(1-\omega_i)^{\frac{1}{\eta_i}}(I^d_{j,t})^{\frac{\eta_i-1}{\eta_i}} + \omega_i^{\frac{1}{\eta_i}}(I^m_{j,t})^{\frac{\eta_i-1}{\eta_i}}\right]^{\frac{\eta_i}{\eta_i-1}} \tag{6.9}$$

Operating as for the consumption prices, we can maximize (6.9) obtaining the demand function of domestic and imported investment:

$$I^d_{j,t} = (1-\omega_i)\left[\frac{P_t}{P^i_t}\right]^{\eta_i} I_{j,t} \tag{6.10}$$

$$I^m_{j,t} = \omega_i\left[\frac{P^{m,i}_t}{P^i_t}\right]^{\eta_i} I_{j,t} \tag{6.11}$$

Thus the aggregate investment price index will be:

$$P^i_t = \left[(1-\omega_i)(P_t^{1-\eta_i}) + \omega_i(P^{m,i}_t)^{1-\eta_i}\right]^{\frac{1}{1-\eta_i}} \tag{6.12}$$

In this model we assume intermediate goods prices to face nominal rigidity defined by a Calvo (1983) scheme however, since those are generated at firm level, we will expose the details of this in the next sections.

The households choices the level of $C_{j,t}$, $M_{j,t+1}$, Δ_t, $\bar{K}_{j,t+1}$, $I_{j,t}$, $u_{j,t}$, $Q_{j,t}$, $B^*_{j,t+1}$ and $h_{j,t}$ that maximise its utility (6.1) facing the following budget constraint:

$$
\begin{cases}
P_t^c C_{j,t}(1+\tau_t^c) + P_t^i I_{j,t} + M_{j,t+1} + S_t B_{j,t+1}^* + P_t\Big(a(u_{j,t})\bar{K}_{j,t} + P_{k',t}\Delta_t\Big) \leq \\
R_{t-1}(M_{j,t} - Q_{j,t}) + Q_{j,t} + (1-\tau_t^k)\Pi_t + \frac{(1-\tau_t^y)}{(1+\tau_t^w)} W_{j,t} h_{j,t} \\
+(1-\tau_t^k) R_t^k u_{j,t} \bar{K}_{j,t} + R_{t-1}^* \Phi\big(\frac{A_{t-1}}{z_{t-1}}, \tilde{\phi}_{t-1}\big) S_t B_{j,t}^* \\
-\tau_t^k \Big[(R_{t-1}-1)(M_{j,t} - Q_{j,t}) + \Big(R_{t-1}^* \Phi\big(\frac{A_{t-1}}{z_{t-1}}, \tilde{\phi}_{t-1}\big) - 1\Big) S_t B_{j,t}^* + B_{j,t}^*(S_t - S_{t-1})\Big] + TR_t + D_{j,t}
\end{cases}
\tag{6.13}
$$

The right side of the inequality represents the households disposal assets and the left side how they use it. Observing this part of the equation it is possible to see how households use their wealth to buy consumption ($P_t^c C_{j,t}$), investments ($P_t^i I_{j,t}$) and financial assets that they hold in form of cash ($M_{j,t}$) and foreign bonds ($B_{j,t+1}^*$) which is multiplied by the nominal exchange rate (S_t) which is given in terms of domestic currency needed to buy a unit of foreign currency, thus an increase of S_t imply a depreciation of the exchange rate and vice versa. The last term express the presence of physical capital and its utilisation cost.

Moving to the right hand side of the (6.13) must be underlined the presence of profits Π_t and $D_{j,t}$ which is the household's net cash income from participating in the state contingent securities at time t. In this framework taxes and (lamp-sum) transfers balance each other: if this last one is represented by TR_t the taxation terms are composed by different elements as τ^c for what concerns consumption and τ^k as the taxation to capital-income. τ^w and τ^y are related to the job market and stand for pay-roll and labour income tax respectively.

The computation of the first order conditions involve the using of the Lagrange multiplier

(v_t), hence need to be observed how, after the combination of the f.o.c. for domestic and foreign bond holding, is it possible to obtain the modified uncovered interest rate parity condition which, in its log-linearized form, is equal to:

$$\hat{R}_t - \hat{R}^*_t = E_t\Delta\hat{S}_{t+1} - \hat{a}_t\tilde{\phi}_a + \hat{\tilde{\phi}}_t$$

where it is assumed that the premium on foreign bond holdings follow the function $\Phi(a_t, \tilde{\phi}_t) = exp\Big(- \tilde{\phi}_a(a_t - \bar{a}) + \tilde{\phi}_t\Big)$ in which $\tilde{\phi}_t$ is a (zero mean) risk premium shock. The real net foreign asset position of the domestic economy is defined as:

$$A_t = \frac{S_t B^*_{t+1}}{P_t} \tag{6.14}$$

$\Phi(a_t, \tilde{\phi}_t)$ lives on the assumption to be strictly decreasing in a_t so that domestic households are charged of a premium on the foreign interest rate if the home country is a net borrower ($B^*_{t+1} < 0$) hence $\hat{a}_t$ enters in the interest rate parity condition because of the imperfect integration of the international financial markets.

6.2 Firms

The intermediate domestic firms produce a differentiated good which they sell to a final good producer. In order to put in place production, intermediate producers use capital and labour inputs and, since this is an open economy setting, intermediate firms may be distinguished as domestic, importing and exporting.

Importing and exporting firms operates in the world market and buy an homogeneous international final goods that are differentiated by brand naming before they sell to domestic households. Exporting firms operate with a similar scheme buying final domestic goods and differentiate it before sell it abroad

6.2.1 Final good firms

The final good producer aggregates intermediate goods into an homogeneous final good according to that Dixit and Stiglitz (1977) aggregator:

$$Y_t = \left(\int_0^1 Y_{i,t}^{\frac{1}{\lambda_{d,t}}} di \right)^{\lambda_{d,t}} \tag{6.15}$$

Domestic-produced goods compete with the imported goods bought by the importing while the exporting firms place a fraction of the final product abroad.

In the above written formula (6.15) $Y_{i,t}$ denote the intermediate goods, with $i \in (0,1)$ while $\lambda_{d,t}$ represents the degree of substitutability hence determining the mark-up for the domestic good market which is assumed to follow a persistent stochastic process:

$$\lambda_{d,t} = (1 - \rho_{\lambda_{d,t}})\lambda_d + \rho_{\lambda_{d,t}}\lambda_{d,t-1} + \epsilon_{\lambda_d t}$$

where $\epsilon_{\lambda_d t}$ is an *i.i.d.* shock.

Te final good producer choose the bundle of goods that minimises the cost of producing Y_t, taking all intermediate goods prices $P_{i,t}$, final domestic goods prices P_t and the quantity of intermediate goods $Y_{i,t}$ as given. The unit price of the unit output is equal to its unit cost

P_t. The profit maximisation leads to the following first order condition:

$$\frac{Y_{i,t}}{Y_t} = \left(\frac{P_t}{P_{i,t}}\right)^{\frac{\lambda_{d,t}}{\lambda_{d,t}-1}} \tag{6.16}$$

By integrating (6.15) and using (6.16) it is possible to obtain the (6.17) which trace the relation that link between final and intermediate goods prices.

$$P_t = \left(\int_0^1 P_{i,t}^{\frac{1}{1-\lambda_{d,t}}} di\right)^{1-\lambda_{d,t}} \tag{6.17}$$

6.2.2 Intermediate firms

All firms combine labour and capital using the same *CRTS* production function and are afflicted by the same aggregate technology shock z_t which is permanent.

$$Y_{i,t} = \left(z_t H_{i,t}\right)^{1-\alpha} K_{i,t}^{\alpha} \epsilon_t - z_t \phi \tag{6.18}$$

It is evident how the function articulates in a Cobb-Doulas fashion in which $K_{i,t}$ represents the capital stock which, since we allow for variable capital utilisation, may differ from physical capital, and $H_{i,t}$ the labour inputs. The last term to present is ϵ_t which is the neutral and stationary one that follows the following univariate representation.

In order to ensure that profits are zero in steady state, the fixed cost is assumed to grow at the same rate as consumption, investment, the real wage, and output do in steady state. If this would not be true it would imply the presence of a monopoly power with a consequent increase in profits and the limitation in the movement in and out of the market. The process for the permanent technology level z_t is exogenously given by:

$$\frac{z_t}{z_{t-1}} = \mu_{z,t} \tag{6.19}$$

$$\mu_{z,t} = (1 - \rho^{\mu_z})\mu_z + \rho^{\mu_z}\mu_{z,t-1} + \varepsilon_{z,t} \tag{6.20}$$

$$\hat{\epsilon}_t = \rho^{\epsilon}\hat{\epsilon}_{t-1} + \varepsilon_t^{\epsilon} \tag{6.21}$$

where $\hat{\epsilon}_t = (\epsilon_t - 1)/1$, $E(\epsilon_t) = 1$ and $\varepsilon_t^{\epsilon} \sim N(0, 1)$.

The firm choice over factor utilization is made by the cost minimization problem constrained by the production function. Each firm must hence choose $K_{i,t}$, $H_{i,t}$, assuming prices $(P_{i,t})$ as given, according to the following formula:

$$W_t R_t^f H_{i,t} + R_t^k K_{i,t} + \lambda_t P_{i,t}\left[Y_{i,t} - \left(z_t^{1-\alpha} H_{i,t}\right)^{1-\alpha} K_{i,t}^{\alpha}\epsilon_t - z_t^u\phi\right] \tag{6.22}$$

where λ_t is the Lagrangian operator which may be interpreted as the nominal cost of producing one additional unit of the domestic good (nominal marginal cost).

The remaining terms take the usual meaning, among those I introduce R_f^k for the gross nominal rate of interest payed by firms which reflects the assumption according of a fraction (v_t) of the intermediate firm wage financed in advance.

$$R_t^f = \nu_t R_{t-1} + 1 - \nu_t$$

Still talking of wage the (6.22) reports the terms $H_{i,t}$ representing the labour input with its relative (aggregate) price W_t to be corresponded. The resulting first order conditions thus are:

$$W_t R_t^f = (1-\alpha)\lambda_t P_{i,t} \left[z_t^{1-\alpha} H_{i,t}^{-\alpha} K_{i,t}^{\alpha} \epsilon_t \right] \tag{6.23}$$

$$R_t^k = \alpha \lambda_t P_{i,t} \left[z_t^{1-\alpha} H_{i,t}^{1-\alpha} K_{i,t}^{\alpha-1} \epsilon_t \right] \tag{6.24}$$

Since the i^{th} firm is a monopolist in the production of i^{th} good - namely monopolist competitor - it may set its price and it will do it in consideration of a probability ξ_d of not being able to reoptimize its price. This setup was firstly proposed by Calvo (1983) and introduce in the model the so called nominal frictions and obligate the firms to face another optimization problem:

$$\max_{P_t^{new}} = E_t \sum_{s=0}^{\infty} (\beta\xi_d)^s \upsilon_{t+s} \Bigg[\Big((\pi_t \pi_{t+1} \dots \pi_{t+s-1})_d^{\kappa} (\bar{\pi}_{t+1}^c \bar{\pi}_{t+2}^c \dots \bar{\pi}_{t+s}^c)^{1-\kappa_d} P_t^{new} \Big) Y_{i,t+s} \\ - MC_{i,t+s}(Y_{i,t+s} + z_{t+s}\phi) \Bigg] \tag{6.25}$$

where $(\beta\xi_d)^s \upsilon_{t+s}$ is the stochastic discount factor and $MC_{i,t+s}$ is the firm's marginal cost that comes manipulating (6.23) and (6.24) and has the form:

$$mc_t = \left(\frac{1}{1-\alpha}\right)^{1-\alpha}\left(\frac{1}{\alpha}\right)^{\alpha}\left(r_t^k\right)^{\alpha}\left(\bar{w}_t R_t^f\right)^{1-\alpha}\frac{1}{\epsilon_t} \tag{6.26}$$

while, following the same procedure, the equilibrium rental rate of capital will be:

$$r_t^k = \frac{\alpha}{1-\alpha}\bar{w}_t \mu_{z,t} R_t^f k_t^{-1} H_{i,t}^j \tag{6.27}$$

Despite the (6.25) (and FOCs) must be noticed how (6.26) are no longer in nominal terms (nor non-stationary), this is because I applied the following transformation before end to the final solution:

$$R_t^k = \frac{R_t^k}{P_t}$$

$$\bar{w}_t = \frac{W_t}{P_t z_t}$$

$$k_{t+1} = \frac{K_{t+1}}{z_t}$$

$$\bar{k}_{t+1} = \frac{\bar{K}_{t+1}}{z_t}$$

the same transformations brought to (6.27).

Combining (6.25) with (6.17) we obtain the average price at period t:

$$\begin{aligned} P_t &= \left[\int_0^{\xi_d}\left(P_{t-1}(\pi_{t-1})^{\kappa_d}(\bar{\pi}_t^c)^{\kappa_d}\right)^{\frac{1}{1-\lambda_{t,d}}} + \int_{\xi_d}^1\left(P_t^{new}\right)^{\frac{1}{1-\lambda_{t,d}}} di\right]^{1-\lambda_{t,d}} \\ &= \left[\xi_d\big(P_{t-1}(\pi_{t-1})^{\kappa_d}(\bar{\pi}_t^c)^{\kappa_d}\big)^{\frac{1}{1-\lambda_{t,d}}} + (1-\xi_d)\big(P_t^{new}\big)^{\frac{1}{1-\lambda_{t,d}}} di\right]^{1-\lambda_{t,d}} \end{aligned} \tag{6.28}$$

which, after some manipulation, once log-linearized brings to:

$$(\hat{\pi}_t - \hat{\bar{\pi}}_t^c) = \frac{\beta}{1 + k_d \beta}\left(\hat{\pi}_{t+1}^d - \rho_\pi \hat{\bar{\pi}}_t^c\right) + \frac{k_d}{1 + k_d \beta}\left(\hat{\pi}_{t-1}^d - \hat{\bar{\pi}}_t^c\right) - \frac{k_d \beta (1 - \rho_\pi)}{1 + k_d \beta}\hat{\bar{\pi}}_t^c + \frac{(1 - \xi_d)(1 - \beta \xi_d)}{\xi_d (1 + k_d \beta)}\left(\hat{mc}_{t,d} + \hat{\lambda}_{t,d}\right) \quad (6.29)$$

which is the Phillips curve.

6.3 Export and import

The structure of import and export is characterised by price frictions and competitive monopoly. Importers purchase an homogeneous foreign good, which they turn into a specialized input and sell to domestic retailers, which retailers are of three types: the first one uses imports to produce consumption goods, the second one uses imports to produce investment goods while the last one uses imports as an input into the production of specialized exports.

Exporters, on the other hand, use homogeneous goods derived from imports and homogeneous domestically produced goods to realise specialized export goods which are sold abroad to foreign citizens.

I will get into the details of importers and exporters in the following subsections, however in this context it is important to underline how pricing frictions and exchange rate play a crucial role. By the import point of view it is important to model, in line with the empirical evidences (see Christiano et al. (2011) and Burlon et al. (2018)), the time that exchange rate shocks take to pass into domestic prices. In this case, in order to complete the overview, must be stressed out how imported goods must be combined with domestic inputs and so the final prices will be a combination of the two factors. It is even important to underline how consumption does not expand rapidly because of the assumption of habit persistence in preferences and so do investments because of the assumption that there are adjustment costs associated with changing the investment flows.

Another important thing to note is that, since part of the import is used in the production of goods aimed for export, the responses of expansionary or contractionary monetary policy, both domestic and foreign, will have effects on the demand of exports and imports as well, hence dampening the responses of those variables. Under a model perspective it is therefore better to keep an asymmetrical structure for the import of goods destined to the export sector; in order to do so, a low price frictions in those goods is supposed.

It is intuitive now how exchange rate would work in a similar manner as prices do. The real exchange rate is in fact defined as:

$$x_t = S_t \frac{P_t^*}{P_t^c} \tag{6.30}$$

This formula is on one hand being composed by prices (domestic and foreign) and affected by the above exposed rigidities, on the other hand it illustrates the role of the nominal exchange rate S. Now, since the model features limit nominal exchange rate pass-through, it is straightforward how a slow response of S lessens the response of exports to monetary policy shocks.

The above written formula (6.30) in its log-linear form becomes:

$$\hat{x}_t = \hat{S}_t + \hat{P}^*_t - \hat{P}^c_t \tag{6.31}$$

6.3.1 Importing firms

Here now I get into the details of imports. As previously exposed, foreign firms sell a homogeneous good to domestic importers which convert the homogeneous good into a specialised input that monopolistically supply to domestic retailers.

Importing firms are of three types of: the ones that produce goods used in the production of intermediate goods aimed to consumption, the ones that produce goods used in the production of intermediate goods aimed to investment and the ones that produce goods used in the production of intermediate goods destined to export. Importers buy homogeneous goods in the world market at price P^*_t and, since they are subject to price frictions à la Calvo (1983), an incomplete price pass-through arises.

The demand for imported consumption goods is given by:

$$C^m_{i,t} = \left(\frac{P^{m,c}_{i,t}}{P^{m,c}_t}\right)^{-\frac{\lambda^{m,c}_t}{\lambda^{m,c}_t - 1}} C^m_t \tag{6.32}$$

where $C^m_{i,t}$ is the output of the i^{th} specialised producer. The formula accounts for two different prices, both taken as given: $P^{m,c}_{i,t}$ which denote the price of the i^{th} intermediate output, and $P^{m,c}_t$ which denote the price index of the imported consumption C^m_t. This last one defined as the final import consumption good, which is a function having a *CES* structure, and composed by a continuum of i differentiated imported consumption goods, each supplied by a different firm.

$$C_t^m = \left(\int_0^1 (C_{i,t}^m)^{\frac{1}{\lambda_t^{m,c}}} di \right)^{\lambda_t^{m,c}} \tag{6.33}$$

Similarly, the final imported investment good is composed by a continuum of differentiated imported investment goods as follows:

$$I_t^m = \left(\int_0^1 (I_{i,t}^m)^{\frac{1}{\lambda_t^{m,i}}} di \right)^{\lambda_t^{m,i}} \tag{6.34}$$

hence the demand for the differentiated imported investment goods will be given by:

$$I_{i,t}^m = \left(\frac{P_{i,t}^{m,i}}{P_t^{m,i}} \right)^{-\frac{\lambda_t^{m,i}}{\lambda_t^{m,i}-1}} I_t^m \tag{6.35}$$

As for the final goods producers, λ_t^j, with j denoting the consumption or investment, is a time-varying mark-up following an AR(1) process. Once again it take values in the range $1 \leq \lambda_t^j < \infty$ and have the explicit form of:

$$\lambda_t^{m,c} = (1 - \rho_{\lambda^{m,c}})\lambda^{m,c} + \rho_{\lambda^{m,c}}\lambda_{t-1}^{m,c} + \varepsilon_{\lambda^{m,c},t} \tag{6.36}$$

$$\lambda_t^{m,i} = (1 - \rho_{\lambda^{m,i}})\lambda^{m,i} + \rho_{\lambda^{m,i}}\lambda_{t-1}^{m,i} + \varepsilon_{\lambda^{m,c},t} \tag{6.37}$$

As we mentioned, the aggregate price index accounts for nominal rigidities and is identical for imported consumption goods and investment goods, here denoted by a generic j taking values ($j = [c, i]$):

$$
\begin{aligned}
P_t^{m,j} &= \left[\int_0^1 \left(P_{i,t}^{m,j}\right)^{\frac{1}{1-\lambda_t^{m,j}}} di \right]^{1-\lambda_t^{m,j}} \\
&= \left[\xi_{m,j} \left(P_{t-1}^{m,j} (\pi_{t-1}^{m,j})^{k^{m,j}} (\bar{\pi}_t^c)^{1-k^{m,j}} \right)^{\frac{1}{1-\lambda_t^{m,j}}} + (1-\xi_{m,j}) \left(P_{new,t}^{m,j}\right)^{\frac{1}{1-\lambda_t^{m,j}}} \right]^{1-\lambda_t^{m,j}}
\end{aligned}
\tag{6.38}
$$

Combining of the two demand functions (6.32) and (6.35) in firms' optimisation problem and computing the first derivatives with respect to consumption and investment goods, it is possible to use the result to get the Phillips curve for imported consumption goods and imported investment goods:

$$
\begin{aligned}
(\hat{\pi}_t^{m,j} - \hat{\bar{\pi}}_t^c) = & \frac{\beta}{1+\kappa_{m,j}\beta} \left(\hat{\pi}_{t+1}^{m,j} - \rho_\pi \hat{\bar{\pi}}_t^c \right) + \frac{\kappa_{m,j}}{1+k_{m,j}\beta} \left(\hat{\pi}_{t-1}^{m,j} - \hat{\bar{\pi}}_t^c \right) \\
& - \frac{\kappa_{m,j}\beta(1-\rho_\pi)}{1+\kappa_{m,j}\beta} \hat{\bar{\pi}}_t^c + \frac{(1-\xi_{m,j})(1-\beta\xi_{m,j})}{\xi_{m,j}(1+\kappa_{m,j}\beta)} \left(\hat{mc}_t^{m,j} + \hat{\lambda}_t^{m,j} \right)
\end{aligned}
\tag{6.39}
$$

where $\hat{mc}_t^{m,j} = \hat{P}_t^* + \hat{s}_t - \hat{P}_t^{m,j}$ and $\hat{\lambda}_t^{m,j}$ is, as always, the mark-up shock. About this last one must be underlined that there is a close correspondence between substitution elasticity and mark-up shocks, in particular I want to point out how such a shock can originate either in households' willingness to substitute among goods or in importing firms' price setting behaviour. For details Adolfson et al. (2007) provide a good technical explanations.

6.3.2 Exporting firms

The exporting firms buy final domestic goods and resell them in the international market after having differentiated them by brand naming according to the following demand function:

$$X_{i,t} = \left(\frac{P^x_{i,t}}{P^x_t}\right)^{-\frac{\lambda^x_t}{\lambda^x_t - 1}} X_t \tag{6.40}$$

where $P^x_{i,t}$ is denominated in foreign currency and, in order to allow for an incomplete pass-through, I assume it to be sticky.

Te above written formula comes from the optimization of the production function $\left[\int_0^1 \left(X_{i,t}\right)^{\frac{1}{\lambda^x_t}} di\right]^{\lambda^x_t}$ where the specialised intermediate goods $X_{i,t}, i \in (0,1)$ are used to produce X_t.

As for the import sector λ^x_t is a mark-up element following an exogenous autoregressive process of order one

$$\lambda^x_t = (1 - \rho_{\lambda^x})\lambda^x + \rho_{\lambda^x}\lambda^x_{t-1} + \varepsilon_{\lambda^x,t} \tag{6.41}$$

Once again with $1 \leq \lambda^x_t < \infty$

In order to illustrate the relation that links exporter demand with the foreign market let's see the total demand by for the domestic export:

$$X_t = \left(\frac{P^x_t}{P^*_t}\right)^{-\eta_f} Y^*_t \tag{6.42}$$

Here P_t^x is the export price index while asterisks denote the rest of the world variables: Y_t^* and P_t^* respectively for foreign GDP and foreign currency price of homogeneous goods.

Complete the set of equation the equation for P_t^x, which is obtained by the combination of production and demand function:

$$P_t^x = \left[\int_0^1 \left(P_{i,t}^x\right)^{\frac{1}{1-\lambda^x}} di \right]^{1-\lambda^x}$$

Following the same steps as we did for the import sector we obtain the aggregate export inflation equation:

$$\left(\hat{\pi}_t^x - \hat{\bar{\pi}}_t^c\right) = \frac{\beta}{1+\kappa_x\beta}\left(\hat{\pi}_{t+1}^x - \rho_\pi \hat{\bar{\pi}}_t^c\right) + \frac{\kappa_x}{1+\kappa_x\beta}\left(\hat{\pi}_{t-1}^x - \hat{\bar{\pi}}_t^c\right) - \frac{\kappa_x\beta(1-\rho_\pi)}{1+\kappa_x\beta}\hat{\bar{\pi}}_t^c + \frac{(1-\xi_x)(1-\beta\xi_x)}{\xi_x(1+\kappa_x\beta)}\left(\hat{mc}_t^x + \hat{\lambda}_t^x\right) \quad (6.43)$$

where $\hat{mc}_t^x = \hat{P}_t^* + \hat{s}_t - \hat{P}_t^x$.

6.4 Wage setting

Households supply specialised labour ($h_{j,t}$) to firms at a price (wage (W_t)) that may be updated with a probability ξ_w according to a Calvo's scheme - the same used for other markets exposed in the previous chapters - that is:

$$W_{j,t+1} = \left(\pi_t^c\right)^{\kappa_w}\left(\bar{\pi}_{t+1}^c\right)^{1-\kappa_w}\mu_{z,t+1}W_{j,t} \quad (6.44)$$

where, once again, $\mu_{z,t+1} = \frac{z_{t+1}}{z_t}$.

The optimization problem faced by j^{th} household configures by inserting the 6.44 to the households utility function 6.1 and the first order condition hence become:

$$\sum_{s=0}^{\infty}(\beta\xi_w)^s h_{j,t+s}\left[-\zeta^h_{t+s}A_L(h_{j,t+s})^{\sigma_L}+\frac{W^{new}_t}{z_tP_t}\frac{z_{t+s}v_{t+s}P_{t+s}}{\lambda_w}\frac{(1-\tau^y_{t+s})}{(1+\tau^w_{t+s})}\frac{(\frac{P^c_{t+s-1}}{P^c_{t-1}})^{\kappa_w}(\bar{\pi}^c_{t+1}\ldots\bar{\pi}^c_{t+s})^{1-\kappa_w}}{\frac{P^d_{t+s}}{P^d_t}}\right]=0 \tag{6.45}$$

In which must be noticed the presence of ξ_w.

$h_{j,t}$ in the objective function can be argued in form of the demand function for labour of the individual household, that is:

$$h_{j,t} = \left(\frac{W_{j,t}}{W_t}\right)^{\frac{\lambda_w}{1-\lambda_w}} H_t \tag{6.46}$$

where λ_w is the wage mark-up.

Since specialised labour ($h_{j,t}$) is supplied by households to domestic firms we need to clarify firm's production function. In particular they operate combining supplied labour with labour contractors into a homogeneous labour service and it happens according to the following production function:

$$H_t = \left[\int_0^1 (h_{j,t})^{\frac{1}{\lambda_w}} dj\right]^{\lambda_w} \tag{6.47}$$

where, as always, $1 \leq \lambda_w < \infty$

6.5 Foreign economy

My model configures as an small open economy, which means that the economy is influenced by the rest of the world without having being able to alter international variables. The variables that defines the rest of the world economy are price, interest rate and output however estimating those variables is not an easy task. Actually World Bank supply world time series, however throughout the adoption of Dynamic Factor Model - *(DFM)* - I propose an alternative way of estimation of those time series, taking advantage of all the features that characterise this approach.

What I do is to consider, instead of the whole world, only the rest of the world respect to Italy, hence making the estimation free of this bias and formally more accurate. This realises in a panel of hundred of time series from which to extract the factors that identify the variables of interest.

6.5.1 Dynamic Factor Model

The Dynamic Factor Models, in nonparametric form, find their baseline version in Forni et al. (2000), where they apply *DFM* to identify VARs and compute impulse response functions, and in Forni et al. (2003, 2005) in which they move from a two-sided to a one-sided estimation of *DFM* in order to improve the forecasting performances.

Forni et al. (2015, 2017) firstly allowed DFM to span the factor space in every possible dimension, hence proposing an infinite-dimensional representation, which is the one I choose to work with in this paper. This representation, starting from a panel Y_{it} of observable stationary stochastic processes, with mean 0 and finite variance, assume that such variables belong to an Hilbert space in L_2 for some probability space (Ω, F, P). The panel hence is composed by $n \times T$ levels of finite realisation and so Y will be: $Y = \{Y_{it} \mid i \in, t \in \mathbb{Z}\}$, where t stands for time and i is the cross-sectional index.

For all $n \in N$ we assume that the spectral measure on Y_n is absolutely continuous respect to the Lebesgue measure on $[-\pi, \pi]$, such that the spectral density matrix is $\Sigma_{Y;n}(\theta)$ that is Hermitian, non-negative definite and has therefore non-negative real eigenvalues for all $\theta \in [-\pi, \pi]$. Such assumptions are necessary and sufficient to guarantee a dynamic factor representation.

We can hence state that Y admits a *dynamic factor* representation, with q factors, if Y_{it} decomposes into a "*common*" component (X_{it}) and an "*idiosyncratic*" component (Z_{it}) such that:

$$Y_{it} = X_{it} + Z_{it} = \sum_{k=1}^{q} b_{ik}(L)u_{kt} + Z_{it} \tag{6.48}$$

Where u_{kt} is a q-dimensional orthonormal process, which is withe noise with mean zero, and Z_{it} is zero-mean second order stationary process whose $Z's$ are weakly cross-correlated, while the two processes are mutually orthogonal at any lead and lag.

Other element of (6.48) is $b_{ik}(L)$ which is a one-sided filter, where L denote the lag operator. This filter is defined to be square-summable $\sum_{m=1}^{\infty} b_{ikm}^2 < \infty$ for all $i \in N$

and $k = 1 \dots q$. The one-sided filters, as demonstrated by Forni et al. (2017), can be obtained without the finite-dimensionality assumption, by adding the condition under which the common components must have rational spectral density, that is, each filter $b_{if}(L)$ in (6.48) is a ratio of polynomials in L. Forni et al. (2015, 2017) also prove that for generic values of parameters $c_{if,k}$ and $d_{if,k}$, X_{it} has the fundamental representation:

$$X_{it} = \frac{c_{i1}(L)}{d_{i1}(L)} u_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)} u_{2t} + \cdots + \frac{c_{iq}(L)}{d_{iq}(L)} u_{qt} \tag{6.49}$$

in which u_t is fundamental for X_{it}.

In order to ease the reading, from now on I will light the notation omitting the cross-section index i.

For each element of the (6.48) a cross-covariance matrices, for example between Y_t and Y_{t-k}, can be estimated; such matrices may be summarised in the form: $\Gamma_{Y;k}(\theta) = E[Y_t Y_{t-k}^{'}]$

The elements of the (6.48) may be rearranged in terms of spectral density and covariance matrices, this representation is purely illustrative, nevertheless the admissibility of such representation is essential for estimation.

$$\Gamma_{Y;k}(\theta) = \Gamma_{X;k}(\theta) + \Gamma_{Z;k}(\theta) \tag{6.50}$$

$$\Sigma_Y(\theta) = \Sigma_X(\theta) + \Sigma_Z(\theta) \tag{6.51}$$

Proceeding in order, it must be reminded that the decomposition of the panel in common and idiosyncratic components is the baseline concept in Dynamic Factor Models literature; while most of the representations are based on the assumption that the space spanned by the stochastic variable X_{it} - for t given and $i \in N$ - is finite-dimensional, Forni et al. (2015) demonstrate that such an assumption is extremely restrictive and potentially harmful so, as they do in their paper, I relax this assumption hence founding the estimation on a potentially infinite number of principal components.

The results obtained by Forni et al. (2015, 2017) rely on the singularity of the vector X_{nt}, where singularity is obtained when q is small compared to n. The q factors are the q^{th} dynamic eigenvectors of the spectral density matrix $\Sigma_{Y;n}(\theta)$: such factors are identified according to the method proposed by Hallin and Liška (2007) and are assumed to diverge as $n \to \infty$, while the $(q+1)^{th}$ one is bounded. Hallin and Lippi (2013) provide the conditions under which this assumption holds for $q < \infty$.

We can thus say that it exists a representation that may be written as a block-diagonal matrix of one-sided filters $A_n(L)$ of dimension $m(q+1) \times m(q+1)$

$$A_n(L) = \begin{bmatrix} A^1(L) & 0 & \dots & 0 & \dots \\ 0 & A^2(L) & \dots & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & A^m(L) & \\ \vdots & \vdots & & & \ddots \end{bmatrix} \tag{6.52}$$

The $(q+1) \times (q+1)$ blocks of $A^i(L)$ are so that a vector autoregressive representation can be developed, where the VAR operators $I_{q+1} - A^i(L)$ are square summable and fundamental. Such VAR representation of Y_n is allowed by the presence of a full-rank $n \times q$ matrix of constants H_n where the full-rank condition is, as said, sufficient to guarantee the fundamentalness.

$$(I - A(L))Y_t = Hu_t + (I - A(L))Z_t = Hu_t + \hat{Z}_t \tag{6.53}$$

where $\hat{Z}_t$ is, by definition, equal to $(I - A(L))Z_t$ and idiosyncratic. The equation (6.53) shows well the filtering role of $A(L)$ matrix, while the common shocks u_t are loaded via H creating what we will call the level-common residuals e_t.

We have hence decomposed Y in two different components orthogonal to each other, and, as Barigozzi and Hallin (2016) show in their paper, such components may generate volatilities whose pattern may be very different from each other and from the one they show in levels. Over that, is important to notice that there is no reason to think each component to be affected by the volatility generated from it only. Saying differently: volatility generated from the common component may influence both common and idiosyncratic ones. The same is true for what concerns the volatility generated by the idiosyncratic components.

6.5.2 The block structure

The interdependency of level-common and level-idiosyncratic component have been introduced by Hallin and Liška (2011). In this paper they propose a block-wise structure of the dataset and investigate the dynamic interrelations within and between the blocks.

In particular they show that all the components span the generic Hilbert space $\mathcal{H}$ and thus every single component configures as a partition of it. As long as such decomposition exist and is unique, it is possible to identify the source of the shocks that drive the whole panel and their components and, using the same definition as in Barigozzi et al. (2019), it is even possible to classify those shocks as "global" and "local", referring as shocks that drive the intersection between the common spaces and the ones that only pool a part respectively.

The concepts of global and local shocks will not be exploited in this paper, however those definitions pass trough the identification of the level-common and level-idiosyncratic components, which will be instrumental for my work. Details and computation procedures will be exposed in the next paragraphs.

The level-common component

In order to estimate common component X_t, we start by estimating the spectral density of Y_t by means of a lag-window estimator. The spectral density matrix is identified by the formula:

$$\hat{\Sigma}_Y(\theta) = \frac{1}{2\pi} \sum_{k=-T+1}^{T-1} e^{ik\theta} K\left(\frac{k}{B_T}\right) \hat{\Gamma}_{Yk} \tag{6.54}$$

in which $e^{ik\theta}$ is the autocovariance generating function, K is a kernel function and B_T is a bandwidth parameter. With the symbol $\hat{\Gamma}_{Yk}$ we define the k^{th} lag estimated autocovariance and it is equal to $\frac{1}{T}\sum_{t=|k|+1}^{T} T_t Y'_{t-|k|}$.

Once estimated the spectral density matrix we determine the number of q level-common shocks. In order to do that we use the Hallin and Liška (2007) information criterion, which is a data-driven method for the identification.

The so treated spectral density matrix will be decomposed as follows:

$$\hat{\Sigma}_X(\theta) = \sum_{k=1}^{q^T} \hat{\Pi}_Y(\theta)\hat{\Lambda}_Y(\theta)\hat{\Pi}_Y^*(\theta) \tag{6.55}$$

where $\hat{\Lambda}_Y(\theta)$ is the $q \times q$ matrix which dimensions are determined by the first q eigenvalues above calculated. $\hat{\Pi}_Y^*(\theta)$ is the $n \times q$ matrix with the corresponding eigenvectors on the column, where the asterisk denotes that the matrix in transposed, complex and conjugate.

The estimate of the covariance matrix may be obtained by Fourier transformation of $\hat{\Sigma}_X(\theta)$. As known the Fourier transformation is used to turn a function of time into its constituent frequencies and it is invertible, meaning that we can conversely move from time to frequency domain; the formulas for those conversions are:

$$\hat{\Gamma}_X = \int_{-\pi}^{+\pi} e^{ik\theta}\hat{\Sigma}_X(\theta)d\theta \tag{6.56}$$

$$\hat{\Gamma}_X = \frac{\pi}{B_T} \sum_{|h|\leq B_T} e^{ik\theta_h}\hat{\Sigma}_X(\theta_h) \tag{6.57}$$

Since we define $n = m(q+1)$ for some $m \in N$, $\hat{\Gamma}_X$ will result as a block-diagonal matrix of dimension $m(q+1) \times m(q+1)$. The estimation of each block enable us to estimate the coefficients of a VAR of dimension $(q+1)$ that will be the estimator of the autoregressive filter appearing in (6.52), which yields to the vector autoregressive representation

$$\hat{Y} = (I - \hat{A}(L))Y \tag{6.58}$$

Projecting the $\hat{Y}_t$ onto their q largest principal component provide an estimate of the level-common innovation process, formally $e_{it} = \{e_{it} \mid i = (H_n u_t)_i \mid i \in \mathbb{N}, t \in \mathbb{Z}\}$, that is: $\hat{e} = \hat{H}\hat{u}$. $\hat{H}$ is the estimator of the loadings and this may be disentangled from $\hat{u}$ imposing the identification constraint $\hat{H}\hat{H}' = I_{\hat{q}}$.

The level-idiosyncratic component

The analysis of volatility, typically, is based on the autocovariance structure of some non-linear transform of innovation processes that (6.49) does not readily offers. Forni and Lippi (2010), Forni et al. (2015) provided the frame used above to analyse the volatility coming from the common component, in a similar fashion Barigozzi and Hallin (2016, 2017) centre their paper on the componentwise residuals coming from the idiosyncratic part. In particular BH (2016) designed a two-step procedure that firstly decompose $(I - A(L))Y_t$ in $\hat{e}_t$ and $(I - A(L))Z_t$ in $\hat{v}_t$, this last one named as level-idiosyncratic residuals, then they proposed two approaches aimed to compute predictions over the level-common and level-idiosyncratic elements and to the application of GARCH techniques.

The estimation of the idiosyncratic element follows the level-common one which is, since $\hat{\Sigma}_Y(\theta)$ and $\hat{\Sigma}_X(\theta)$ are our estimated spectral density matrix, respectively, for the data and for the common component, it imply that the one for the idiosyncratic component may be obtained as:

$$\hat{\Sigma}_Z(\theta) = \hat{\Sigma}_Y(\theta) - \hat{\Sigma}_X(\theta) \tag{6.59}$$

Given the autoregressive formulation (6.58), the estimator of $\hat{Z}$ will be:

$$\hat{Z} = (I - \hat{A}(L))\hat{Y} - \hat{e} \tag{6.60}$$

in which $\hat{\upsilon}_t$ is formally defined as: $\upsilon_{it} = \{\upsilon_{it} \mid i \mid i \in \mathbb{N}, t \in \mathbb{Z}\}$. Reminding that $\tilde{Z}_t = (I - A(L))Z_t$, the process υ_t thus become:

$$\upsilon_t = (1 - c_i(L))\hat{Z}_t \tag{6.61}$$

The representation (6.61) make use of the AR filters $c(L)$ which, as before, are one-sided, square summable and such that every root lies outside the unit circle ($c(z) = 0$). Both idiosyncratic processes are zero-mean second-order white noise and not mutually orthogonal, meaning that at this level some mild cross-correlation among them still remains.

Barigozzi and Hallin (2016, 2017) analyse both e_{it} and υ_{it} via general dynamic factor models, this mean that, as for the baseline representation (6.48), similar assumptions must take place: in particular the existence of a second-order moment for all $i \in N$ as well as spectral density continuous respect to the Lebesgue measure over $[-\pi, \pi]$. In addition, as noted by Engle and Marcucci (2006), in order to ease computations, some transformation must be applied: logarithmic proxy over square residuals would allow to analyse the panels via an additive factor models instead of imposing some positivity constraint at the moment of model estimation, moreover some attention mus be payed to the fact that e_{it} and υ_{it}

are the white noise residual obtained from two mutually orthogonal components, then by construction they are uncorrelated, therefore justifying our definition of

$$s_{it} = log[(e_{it} + \upsilon_{it})^2] - Elog[(e_{it} + \upsilon_{it})^2] \tag{6.62}$$

which is an $N \times T$ panel of centred volatility proxies.

In a similar manner as before, this matrix rely on the assumption of a spectral density absolutely continuous with respect to the Lebesgue measure over $[-\pi, \pi]$, hence the generalised dynamic factor model decomposition will return the following representation:

$$s_{it} = \chi_{s,it} + \xi_{s,it} = \sum_{k=1}^{q} b_{s,ik}(L)\epsilon_{s,kt} + \xi_{s,it} \tag{6.63}$$

In this case I will assume the existence of $q_s \in N$ such that the q_s^{th} eigenvalue of the spectral matrix diverges as $n \to \infty$, while the q_{s+1}^{th} is bounded.

s_{it} can be decomposed using GDFM as to obtain a moving average representation for the common components of the form:

$$\chi_t = D(L)R^{-1}R\epsilon_t$$

where the matrix R is a square one of dimension $Q_s \times Q_s$ which can be determined by imposing appropriate exogenous restrictions. Among all possible choices we restrict our search to orthogonal transformations as is customary in structural VAR models.

With this framework every block may be analysed separately and jointly at the occurrence, as it is possible for the intersection of those blocks; this is the point of disentangling the panel into level- and volatility-common panels. However in this paper I only rely on a common representation, hence imposing a structure to the so obtained VAR in order to maintain a comparability with the theoretical model coming from the DSGE.

According to the above exposed procedures, common and idiosyncratic component may be extracted out from each block, which may be seen as a partition of the space spanned by Y where each block is allowed to intersect with the others. The point in which all the three shocks intersect identify the "global" shock area and indicate an ideal zone in which a shock occurring in one subpanel is pervasive to the whole economy. Intersections, instead, are the zones in which partitions overlap. The study of global and local shocks have been the core of Barigozzi et al. (2019), however the adoption of this representation in conjunction with DSGE will be left for future works.

6.5.3 Structural VAR (DSGE)

As said, the bock structure is essential to study the international inter-dependence of the economies or, in other words, the way in which international macroeconomic variables have effects to each other.

The topic offers many points of interest, however, in order to guarantee a better comparability with impulse response functions generated by DSGE, once extracted the series, another decomposition were imposed, details of which will be provided in the dedicated chapter.

In the model, international variables are exogenous to the ones of the small domestic economy and their evolution is described by a fourth-order structural VAR, where contemporaneous correlations are defined by the structure of the stochastic component matrix B reported below:

$$F_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ b_{3,1} & b_{3,2} & 1 \end{bmatrix} \tag{6.64}$$

where $X_t^* = [\pi_t^* \;\; \hat{y}_t^* \;\; R_t^*]'$.

The assumptions on the contemporaneous correlations matrix B are consistent with the hypothesis that output and inflation do not respond contemporaneously to the other shocks in the system, hence the SVAR system adds three linear stochastic equations - with four lags each - to the economic and stochastic relations of the domestic economy model.

6.6 Net Foreign asset

The Net Foreign Assets (NFA), in aggregate level, evolve according to the following formula:

$$S_t B_{t+1}^* - R_{t-1}^* \Phi(a_{t-1}, \tilde{\phi}_{t-1}) S_t B_t^* = S_t P_t^x (C_t^x + I_t^x) - S_T P_t^* (C_t^m + I_t^m) \tag{6.65}$$

I get in the detail of the term $R_{t-1}^* \Phi(a_{t-1}, \tilde{\phi}_{t-1})$, already present in (6.13): this is the risk-adjusted gross nominal interest rate over foreign bond holding, where $a_t = \frac{S_t B^*_{t+1}}{P_t z_t}$ is the stationarized real term definition of NFA. In other words if if $B^*_{t+1} < 0$ the domestic economy is a net borrower and households are charged by a premium on foreign asset while the inverse is true if $B^*_{t+1} > 0$ hence the economy is a net lender. The so defined risk-premium assure the steady-state of the model to be unique and independent of the initial net foreign asset and capital stock of the economy.

Scaling the 6.65 by multiplying by $\frac{1}{P_t z_t}$ and using $\frac{C^x_t}{z_t} + \frac{I^x_t}{z_t} = \left[\frac{P^x_t}{P^*_t}\right]^{-\eta_f} \frac{Y^*_t}{z^*_t}\frac{z^*_t}{z_t}$ a_t may be expressed as:

$$a_t = (mc^x_t)^{-1}(\gamma^{x,*}_t)^{-\eta_f} y^*_t \tilde{z}t^* - (\gamma^f_t)^{-1}(c^m + i^m_t) + R^*_{t-1}\Phi(a_{t-1}, \tilde{\phi_{t-1}})\frac{a_{t-1}}{\pi_t \mu_{z,t}}\frac{S_t}{S_{t-1}} \tag{6.66}$$

which once log-linearized becomes:

$$\begin{aligned} \hat{a}_t &= -y^* \hat{mc^x_t} - \eta_f y^* \hat{\gamma}^{x,*}_t + y^* \hat{y}^*_t + y^* \hat{\tilde{z}}^*_t \\ &+ \hat{\gamma}^f_t (c^m + i^m_t) - c^m \bigg(-\eta_c (1-\omega_c)(\gamma^{c,d})^{-(1-\eta_c)} \hat{\gamma}^{mc,d}_t + \hat{c}_t \bigg) \\ &+ i^m \bigg(-\eta_i (1-\omega_i)(\gamma^{i,d})^{-(1-\eta_i)} \hat{\gamma}^{mi,d}_t + \hat{i}_t \bigg) + \frac{R}{\pi \mu_z} \hat{a}_{t-1} \end{aligned} \tag{6.67}$$

6.7 Fiscal and Monetary policy

In this economy spends resources on government consumption of the final domestic good, G_t and we assume that there is not government debt, hence the resulting fiscal surplus/deficit plus the seigniorage are assumed to be transferred back to the households TR_t in a lump-sum fashion.

$$TR_t + P_t G_t = R_{t-1}(M_{t+1} - M_t) + \tau_t^c P_t^c C_t + \frac{(\tau_t^y + \tau_t^\omega)W_t}{1 + \tau_t^\omega} H_t$$
$$+ \tau_t^k \Big[(R_{t-1} - 1)(M_t Q_t) + R_t^k u_t \bar{K}_t + \Big(R_{t-1}^* \Phi(a_{t-1}, \tilde{\phi}_{t-1} - 1) \Big) S_t B_t^* + \Pi_t \Big] \tag{6.68}$$

In order to finance its spending the government collects tax revenues resulting from taxes on capital income τ_t^k, labor income τ_t^y, consumption τ_t^c, and pay-roll τ_t^w. Following Adolfson et al. (2007) we define

$$\tau_t = [\tau_t^k \quad \tau_t^y \quad \tau_t^c \quad \tau_t^w \quad G_t]$$

where G_t is the public expenditure and the fiscal policy follows a Vector Autoregressive representation given by:

$$\Gamma_{0\tau_0} = \Gamma(L)\tau_{t-1} + \epsilon_{\tau,t} \tag{6.69}$$

with $\epsilon_{\tau,t} \sim N(0, \Sigma_\tau)$

The monetary authority, instead, is supposed to follow a Taylor rule according to which the Central Bank adjust the short-term interest rate in response to the CPI inflation rate deviation from a time-varying inflation target $\hat{\pi}_t^c - \hat{\bar{\pi}}_t^c$ (omitting indirect taxes (τ_t^c)), the output gap (measured as actual minus trend output) and the real exchange rate.

About the response to inflation deviation

Thus, monetary policy is approximated with the following (log-linearized) instrument rule:

$$\hat{R}_t = \rho_R \hat{R}_{t-1} + (1-\rho_R)\big(\hat{\bar{\pi}}_t^c + r_\pi(\hat{\pi}_{t-1}^c - \hat{\bar{\pi}}_t^c) + r_y \hat{y}_{t-1} r_x \hat{x}_{t-1}\big) + r_{\Delta_\pi}\Delta\hat{\pi}_t^c + r_{\Delta_y}\Delta\hat{y}_t + \varepsilon_{R,t} \quad (6.70)$$

where $\varepsilon_{R,t}$ is an uncorrelated monetary policy shock.

6.8 Market clearing conditions

Combining government's and households' budget constraints with zero profit condition of final goods producers and employment agencies yields the aggregate resource constraint:

$$C_t^d + I_t^d + G_t + C_t^x + I_t^x \leq Y_t \bar{K}_t$$

where Y_t is defined as we did before: $Y_t = \epsilon_t z_t^{1-\alpha} H_t^{1-\alpha} K_t^\alpha - z_t\phi - a(u_t)\bar{K}_t$.

Scaling the formula to be real and opportunely substituting the variables with the ones above defined we get:

$$\begin{aligned}(1-\omega_c)\left[\frac{P_t^c}{P_t}\right]^{\eta_c} c_t + (1-\omega_i)\left[\frac{P_t^i}{P_t}\right]^{\eta_i} i_t + g_t + (1-\omega_c)\left[\frac{P_t^x}{P_t^*}\right]^{-\eta_f} y_t^* \frac{z_t^*}{z_t} \\ \leq \epsilon_t\left(\frac{1}{\mu_{z,t}}\right) K_t^\alpha H_t^{1-\alpha} - \phi - a(u_t)\bar{k}_t \frac{1}{\mu_{z,t}}\end{aligned} \quad (6.71)$$

In (6.71) z_t^* represents the technology shock and, as the other variables, have been stationarized by scaling with z_t getting $\tilde{z}_t^* = \frac{z_t^*}{z_t}$. $\tilde{z}_t^*$ is by definition a stationary shock that measure the degree of asymmetry in permanent shocks to technological progress between foreign and domestic economy and, as for the other shocks in the model, once log-linearized ($\tilde{z}_t^* = 1$ in steady state), follows an AR(1) process of the form:

$$\hat{\tilde{z}}_{t+1}^* = \rho_{\tilde{z}^*} \hat{\tilde{z}}_t^* + \epsilon_{\tilde{z}^*,t+1} \tag{6.72}$$

Money market, on the other hand, clears at the following condition:

$$\nu W_t H_t = \mu_t M_t - Q_t \tag{6.73}$$

which, once standardised, becomes:

$$\nu \bar{\omega}_t H_t = \frac{\mu_t \bar{m}_t}{\pi_t \mu_{z,t}} - q_t \tag{6.74}$$

Chapter 7

Data and estimation results

The data have been taken from OECD database, for the period between 1995 and 2019 with quarterly frequency. Since the theoretical model is calibrated on Italian economy, the rest of the world time series have been extracted from every country except for Italy; in particular I used gdp, interest rate and inflation (consumer prices) for every country that could provide a complete time series for the considered period. Those variables have a size, respectively, of 71, 132 and 75 time series, summing up to 278 variables in cross-section and a total of 27522 observations once the longitudinal dimension is considered.

As showed by Hallin and Liška (2011), the block structure of DFM is essential if some dependence between blocks must be accounted. In particular they define the mathematical background upon which a decomposition in sub-panels may be built. Such decomposition was further investigated by Barigozzi et al. (2019), who constructed an intersection of the common spaces of all blocks to allow shocks to be pervasive. Since I work in a structural VAR environment, I do not need such a decomposition to identify shocks, however I rely on a block structure of the panel in order to extract the common factors representing the desired series. More importantly, since the rest of the world variables are exogenous in the model, then I can estimate a VAR for this section alone.

Validation through VAR is typically carried out by comparing the empirical impulse response functions (*IRF*) with the theoretical ones obtained by DSGE. In order to guarantee that the information provided by the VAR is *sufficient* I rely on the concept of *fundamentalness* which says that a VAR is fundamental if and only if it is informationally sufficient for all the structural shocks[1]. In other words, the structural representation is fundamental if all the shocks can be recovered from a VAR.

This gives the intuition of how fundamentalness is a weak point of VAR analysis; however the adoption of a DFM approach significantly simplify the identification problem, providing a satisfactory solution (see Forni et al. (2009) and Forni et al. (2000) for further details) and offering another (theoretical) support for the adoption of dynamic factor analysis in this framework.

Since fundamentalness of structural shocks can be assumed in the DFM framework, identification is reduced to the choice of a matrix H such that economically motivated restrictions on the matrix $B_n(L)H$ are fulfilled.
Identification can be achieved by maximizing or minimizing an objective function involving $B_n(L)H$. An alternative is to impose zero restrictions either on the impact effects or the long-run effects or both.

[1]See Lippi and Reichlin (1993) to deepen the link between sufficient information and fundamentalness

Forni et al. (2015, 2017) show that, letting the factor space be of infinite dimensions, the model return good results, outperforming all the other factor models both doing IRF or forecasting. As in their experiment, I identify the "structural" shocks and the corresponding impulse–response functions by imposing a Cholesky identification scheme on the first q variables. The results are reported in the figure 7.1 below.

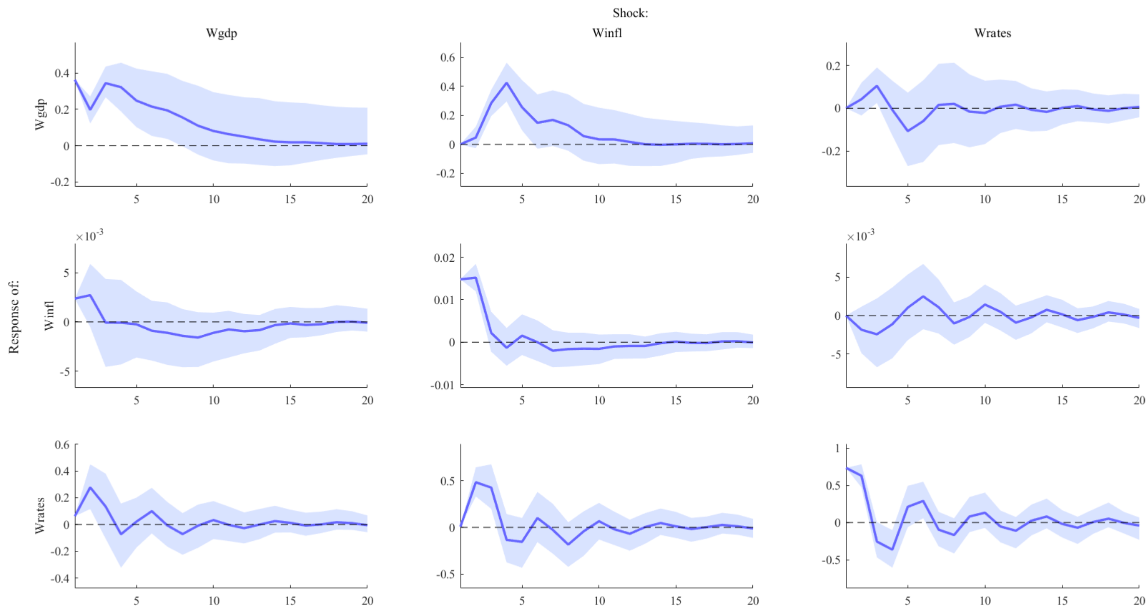

Figure 7.1: S-DFM Impulse Response Function

On the other hand, the theoretical model was calibrated for the Italian economy setting the discount factor parameter β equal to 0.996, the depreciation rate $\delta = 0.025$ and the capital share in production $\alpha = 0.333$, this in line with the information provided by Beqiraj and Tancioni (2014) and Annicchiarico et al. (2013), this last ones even indicate a value of 0.25 for the tax over labour income τ^y. The habit parameter b is set equal to 0.7 as suggested by Smets and Wouters (2007) while the imported share of investment and consumption (ω_i, ω_c) was imposed equal to 0.52 and 0.285 respectively, this is to match the ratio of imports over gdp and investment (gross capital formation) over imports.

The persistence parameter for the inflation ρ_π is 0.975 while the money growth rate is related to the steady state level of inflation and its hence set equal to 1.01 (per quarter) and the cash to money ratio parameter A_q was set equal to 0.3776 following the findings of Adolfson et al. (2007).

Central, for the aims of this work, is the role played by the parameters linked to the foreign market. In particular, the elasticity of substitution parameter of domestic and foreign consumption goods (η_c and η_i) and the elasticity of substitution for investment η_f, which will all be set equal to 1.5 as done in the papers by Beqiraj and Tancioni (2014) and Christiano et al. (2011). Belong to this group the indexation parameters κ_x, $\kappa_{m,i}$ and $\kappa_{m,c}$, all set equal to 0.5, and the time varying mark-up parameters λ_d, $\lambda_{m,c}$ and $\lambda_{m,i}$ all initially fixed at 1.2.

The complete set of parameter values are summarised in the table B.1 in the appendix.

The first thing to notice is an oscillating behaviour of interest rate. In the first place must be noticed that the interest rate appeared to be most volatile among the estimated series, the behaviour of the IRF surely is a consequence of the lag structure imposed, that, in order to keep the comparability with the DSGE, are modelled to be 4. Under a more economic point of view not surprises the positive response to a shock of both gdp and inflation.

This waving walk of the interest rate have a direct impact over both gdp and inflation that seem to acquire the same pattern once hit by an unexpected shock in R. To underline here the negative effect that such a shock has over inflation at the moment that it occurs.

No surprises for what concern the movement generated by gdp, while to notice a little lag in the response of gdp when the shock comes from inflation. Figure 7.1 above illustrates this all.

Applying the same shock to the international sector of the theoretical DSGE model, the responses to a shock in the rest of the world gdp is reported in the 7.2 below:

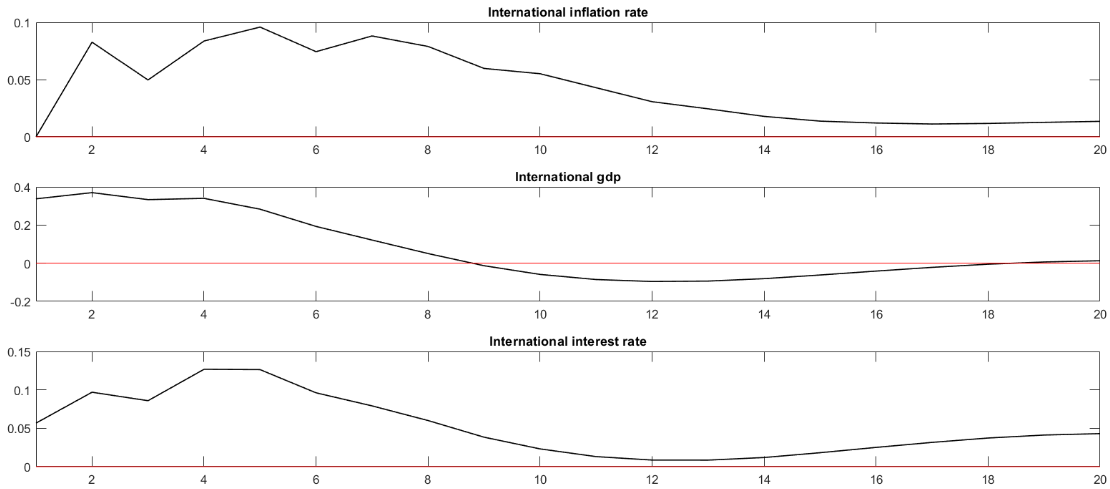

Figure 7.2: DSGE response to a shock on international gdp

Here it is possible to observe a pattern similar to the one obtained using estimated data, however loosing the cyclical behaviour that characterised the empirical analysis.

Way more different is the walk showed by the system once hit by an orthogonal shock over prices. The model, in fact, offer positive idiosyncratic response, which demonstrate to be quite persistent, and another positive response by the side of interest rate: this last one which tend to reabsorb way faster however never really coming back to the steady state. The behaviour of gdp shows to be pretty different having an initial positive response, then evolving on a negative field after nine periods.

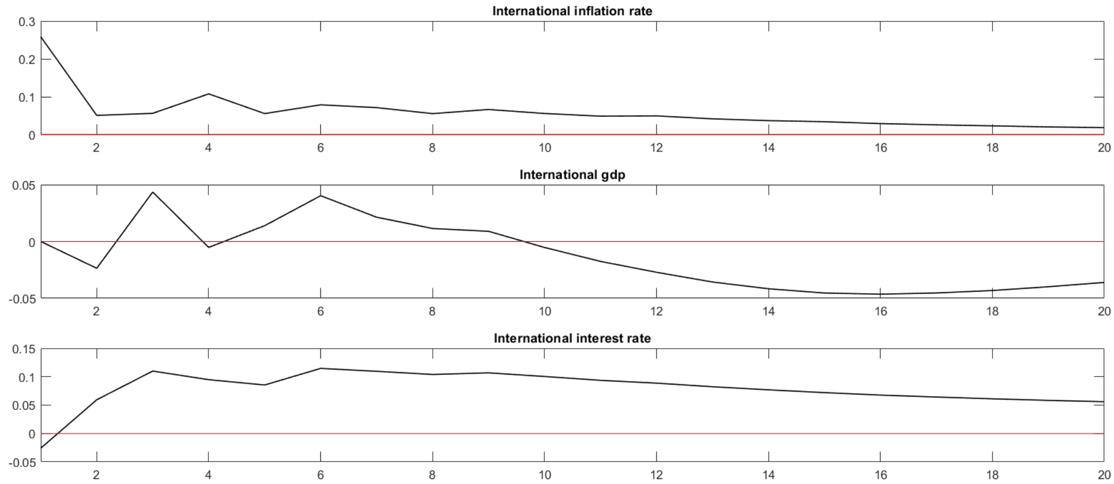

Figure 7.3: DSGE response to a shock on international prices

The curves generated from a shock on international prices, reported in figure 7.3, tell us that international economy responds in a similar manner as it does for a shock in production; this mean that we can observe a positive and long lasting response from interest rate, a more transitory idiosyncratic response and an oscillatory behaviour for what concerns gdp.

A shock to interest rate, instead, have an impact on itself that exhaust in about ten periods. More ambiguous is the behaviour of inflation and production. Figure 7.4 shows us that for what regards inflation the curve is pretty close to the steady state, however with an overall negative magnitude, while gdp tend to convey more in the long period, even thought an initial positive impact.

Those responses are in line with the theoretical expectation ad certify, with the rest of the IRFs reported in the appendix, the general good performance of the model.

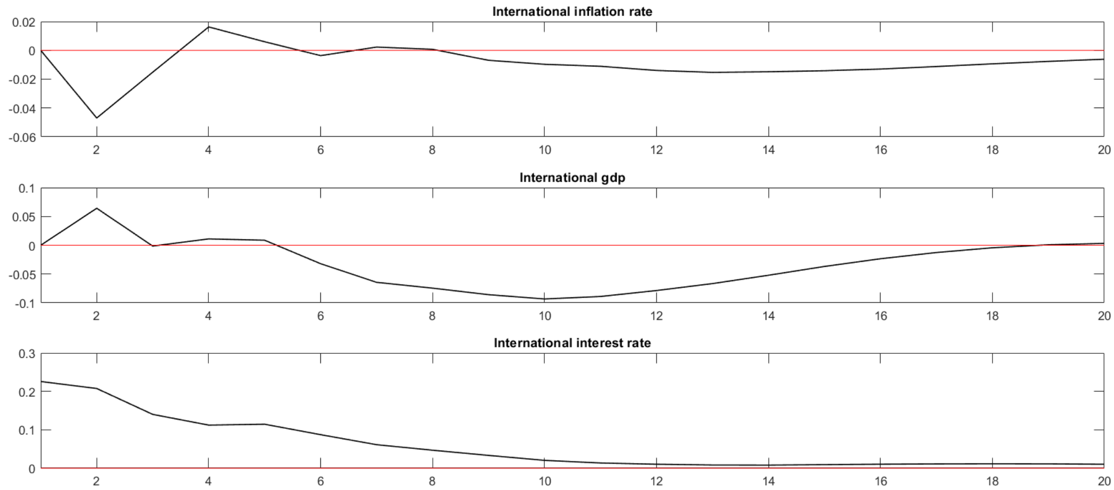

Figure 7.4: DSGE response to a shock on international interest rate

Chapter 8

Conclusions

To bring model to data is an essential step in dynamic macroeconomics in order to certify the consistency of the models. Many models of general equilibrium nowadays are designed as a small open economy, however the international sector is generally kept purely theoretical, meaning that, even though some VAR representation is imposed, no data are inserted in those series.

My task here is thus to bring model to data in this sector and, since the model was calibrated for the Italian economy, to get a more punctual estimation of the rest of the world time series by computing those variables using dynamic factor models (hence excluding Italy) instead of using a general world time series as estimated by official statistics. The model validation, with particular concerns of the foreign sector, was another target of my study.

For a meaningful validation a necessary condition is that the VAR conveys enough information to recover the shocks of interest and the related IRF. By inverse is known how for DSGE models including news or foresight shock, non-fundamentalness is endemic and a

VAR representation can be used for model validation despite it non-fundamentalness.

I showed how the block structure of DFM, essential to identify the variables and recover the dependence between blocks, allowing to build a VAR which structural representation that is fundamental thanks to the properties of dynamic factor models analysis.

Given the properties of DFM and the good performances of infinite-dimensional representation, I applied this model in order to extract the variables of interest. After having identified those time series I computed impulse response functions by imposing a Cholesky structure over the obtained VAR.

The data analysis demonstrate that the model can reproduce the general dynamics of the international sector, however the magnitude of the responses have recorded some difference nevertheless not invalidating the model but suggesting a possible different calibration of the model.

Dynamic Factor Model even showed to sometimes anticipate responses respect to the theoretical model this, even in consideration of the similar lag structure imposed, may give place of a reflection over different specification.

In general the DSGE model demonstrated to performs reasonably well in fitting IRF derived from rest of the world data, more interestingly the DFM approach - in its non-parametric infinite-dimensional representation - demonstrated to be an effective tool in this framework, opening the way for the application in other sectors of those constituting a general equilibrium model.

Bibliography

F. W. A. *Introduction to Statistical Time Series*. Wiley-Interscience; 2 edition, 1996.

M. Adolfson, S. Laséen, J. Lindé, and M. Villani. Bayesian estimation of an open economy dsge model with incomplete pass-through. *Journal of International Economics*, 72(2):481–511, 2007. URL https://EconPapers.repec.org/RePEc:eee:inecon:v:72:y:2007:i:2:p:481-511.

B. Annicchiarico, F. D. Dio, F. Felici, and L. Monteforte. IGEM: a Dynamic General Equilibrium Model for Italy. Working Papers 4, Department of the Treasury, Ministry of the Economy and of Finance, Apr. 2013. URL https://ideas.repec.org/p/itt/wpaper/wp2013-04.html.

M. Asai, M. McAleer, and J. Yu. Multivariate stochastic volatility: a review. *Journal of the Economic Review*, 25:145–175, 2006. doi: 10.1080/07474930600713564.

M. Barigozzi and M. Hallin. Generalized dynamic factor models and volatilities: Recovering the market volatility shocks. *The Econometric Journal*, 19:C33–C60, 2016. doi: 10.1111/ectj.12047.

M. Barigozzi and M. Hallin. Generalized dynamic factor models and volatilities: estimation and forecasting. *Journal of Econometrics*, 201:307–321, 2017. doi: 10.1016/j.jeconom.2017.08.010.

M. Barigozzi, M. Hallin, and S. Soccorsi. Identification of global and local shocks in international financial markets via general dynamic factor models. *Journal of Financial Econometrics*, 17:462–494, 2019. doi: https://doi.org/10.1093/jjfinec/nby006.

E. Beqiraj and M. Tancioni. Fiscal Consolidation and Sovereign Risk in the Euro-zone Periphery. Working Papers 167, University of Rome La Sapienza, Department of Public Economics, Nov. 2014. URL `https://ideas.repec.org/p/sap/wpaper/wp167.html`.

B. S. Bernanke, J. Boivin, and P. Eliasz. Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach. Working Paper 10220, National Bureau of Economic Research, January 2004. URL `http://www.nber.org/papers/w10220`.

L. Burlon, A. Notarpietro, and M. Pisani. Exchange rate pass-through into euro-area inflation. an estimated structural model. *Bank of Italy working paper*, (1192):169–194, 2018.

G. Calvo. Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12(3):383–398, 1983. URL `https://EconPapers.repec.org/RePEc:eee:moneco:v:12:y:1983:i:3:p:383-398`.

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

S. Chib, F. Nardari, and N. Shephard. Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134:341–371, 2006. doi: 10.1016/j.jeconom.2005.06.026.

L. J. Christiano, M. Eichenbaum, and C. L. Evans. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1):1–45, 2005. doi: 10.1086/426038. URL `https://doi.org/10.1086/426038`.

L. J. Christiano, M. Trabandt, and K. Walentin. Introducing financial frictions and unemployment into a small open economy model. *Journal of Economic Dynamics and Control*, 35(12):1999–2041, 2011. doi: 10.1016/j.jedc.2011.09.00. URL `https://ideas.repec.org/a/eee/dyncon/v35y2011i12p1999-2041.html`.

N. Crato and P. Paruolo. *Data-Driven Policy Impact Evaluation: How Access to Microdata is Transforming Policy Design*. Springer, 2019.

M. Creel and D. Kristensen. Abc of sv: Limited information likelihood inference in stochastic volatility jump-diffusion models. *Journal of Empirical Finance*, 31:85–108, 2015. doi: 10.1016/j.jempfin.2015.01.002.

M. Del Negro and C. Otrok. *FRB of New York Staff Report.*

E.-I. Delatola and J. E. Griffin. Bayesian nonparametric modelling of the return distribution with stochastic volatility. *Bayesian Analysis*, 6:901–926, 2011.

A. K. Dixit and J. E. Stiglitz. Monopolistic competition and optimum product diversity. *The American Economic Review*, 67(3):297–308, 1977. URL `http://www.jstor.org/stable/1831401`.

R. F. Engle and J. Marcucci. A long-run pure variance common features model for the common volatilities of the dow jones. *Journal of Econometrics*, 132:7–42, 2006. doi: 10.1016/j.jeconom.2005.01.021.

T. S. Ferguson. A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973. ISSN 00905364. URL `http://www.jstor.org/stable/2958008`.

M. Forni and M. Lippi. The Unrestricted Dynamic Factor Model: One-sided Representation Results. *Journal of Econometrics*, 163:23–28, 2010.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The Generalized Dynamic Factor Model: Identification and estimation. *The Review of Economics and Statistics*, 82:540–554, 2000.

M. Forni, M. Lippi, and L. Reichlin. Opening the black box: Structural factor models versus structural vars. *SSRN*, 50:1243–1255, 2003.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The Generalized Dynamic Factor Model: One sided estimation and forecasting. *Journal of the American Statistical Association*, 100:830–840, 2005.

M. Forni, D. Giannone, M. Lippi, and L. Reichlin. Opening the black box: Structural factor models with large cross sections. *Econometric Theory*, 25(5):1319–1347, 2009.

M. Forni, M. Hallin, M. Lippi, and P. Zaffaroni. Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *Journal of Econometrics*, 185:359–371, 2015.

M. Forni, M. Hallin, M. Lippi, and P. Zaffaroni. Dynamic factor models with infinite dimensional factor space: Asymptotic analysis. *Journal of Econometrics*, 199:74–92, 2017.

A. E. Gelfand, S. K. Sahu, and B. P. Carlin. Efficient parametrisations for normal linear mixed models. *Biometrika*, 82(3):479–488, 09 1995. ISSN 0006-3444. doi: 10.1093/biomet/82.3.479. URL `https://doi.org/10.1093/biomet/82.3.479`.

Y. Gorodnichenko and S. Ng. Level and volatility factors in macroeconomic data. *Journal of Monetary Economics*, 91:52–68, 2017. doi: 10.1016/j.jmoneco.2017.09.004.

M. Hallin and M. Lippi. Factor Models in High-Dimensional Time Series: A Time-Domain Approach. Working Papers ECARES ECARES 2013-15, ULB – Universite Libre de Bruxelles, Mar. 2013. URL `https://ideas.repec.org/p/eca/wpaper/2013-142428.html`.

M. Hallin and R. Liška. Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102:603–617, 2007.

M. Hallin and R. Liška. Dynamic factors in the presence of blocks. *Journal of Econometrics*, 163:29–41, 2011. doi: 10.1016/j.jeconom.2010.11.004.

E. Jacquier, N. G. Polson, and P. E. Rossi. Bayesian analysis of stochastic volatility models. *Journal of Business Economic Statistics*, 20(1):69–87, 2002. ISSN 07350015. URL `http://www.jstor.org/stable/1392151`.

M. J. Jensen and J. M. Maheu. Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics*, 157:306–316, 2010. doi: 10.1016/j.jeconom.2010.01.014.

M. J. Jensen and J. M. Maheu. Estimating a semiparametric asymmetric stochastic volatility model with a dirichlet process mixture. *Journal of Econometrics*, 178:523–538, 2014. doi: 10.1016/j.jeconom.2013.08.018.

S. Kim, N. Shephard, and S. Chib. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *The Review of Economic Studies*, 65(3):361–393, 07 1998. ISSN 0034-6527. doi: 10.1111/1467-937X.00050. URL `https://doi.org/10.1111/1467-937X.00050`.

G. Koop and D. Korobilis. Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends® in Econometrics*, 3:267–358, 2010. ISSN 1551-3076. doi: 10.1561/0800000013.

S. J. Koopmanp, G. Mesters, and B. Schwaab.

M. Lippi and L. Reichlin. The Dynamic Effects of Aggregate Demand and Supply Disturbances: Comment. *American Economic Review*, 83(3):644–652, June 1993. URL https://ideas.repec.org/a/aea/aecrev/v83y1993i3p644-52.html.

A. Y. Lo. On a class of bayesian nonparametric estimates: I. density estimates. *The Annals of Statistics*, 12(1):351–357, 1984. ISSN 00905364. URL http://www.jstor.org/stable/2241054.

R. J. Mahieu and P. C. Schotman. An empirical application of stochastic volatility models. *Journal of Applied Econometrics*, 13:333–360, 1998.

Y. Meng. Bayesian analysis of a stochastic volatility model.

R. M. Neal. Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9:249–265, 2000.

C. Otrok and C. H. Whiteman. Bayesian leading indicators: Measuring and predicting economic conditions in iowa. *International Economic Review*, 34:997–1014, 1998. doi: 10.2307/2527349.

O. Papaspiliopoulos, O. R. Gareth, and M. Sköld. A general framework for the parametrization of hierarchical models. *Statistical Science*, 22(1):59–73, 2007. ISSN 0883-4237. doi: 10.1214/088342307000000014.

D. Sakaria and J. Griffin. On efficient bayesian inference for models with stochastic volatility. *Econometrics and Statistics*, 3:23–33, 2017. doi: 10.1016/j.ecosta.2016.08.002.

F. Smets and R. Wouters. An estimated stochastic dynamic general equilibrium model of the euro area. Working Paper Series 171, European Central Bank, Aug. 2002. URL `https://ideas.repec.org/p/ecb/ecbwps/2002171.html`.

F. Smets and R. Wouters. Shocks and frictions in US business cycles: A Bayesian DSGE approach. *The American Economic Review*, 97:586–606, 2007.

J. H. Stock and M. W. Watson. *New Indexes of Coincident and Leading Economic Indicators*, volume 4. Olivier Jean Blanchard and Stanley Fischer, 1989. URL `http://www.nber.org/chapters/c10968`.

J. H. Stock and M. W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179, 2002. doi: 10.1198/016214502388618960.

J. H. Stock and M. W. Watson. Implications of dynamic factor models for VAR analysis. Working Paper 11467, NBER, 2005a.

J. H. Stock and M. W. Watson. Implications of dynamic factor models for VAR analysis. Working Paper 11467, NBER, 2005b.

S. J. Taylor. Modeling stochastic volatility: A review and comparative study. *Mathematical Finance*, 4(2):183–204, 1994.

S. J. Taylor. *Asset price dynamics volatility and prediction.* 2007.

H. Uhlig. Bayesian vector autoregressions with stochastic volatility. *Econometrica*, 65:59–73, 1997. ISSN 00129682, 14680262. doi: 10.2307/2171813. URL `http://www.jstor.org/stable/2171813`.

Y. Yu and X.-L. Meng. To center or not to center: That is not the question—an ancillarity—sufficiency interweaving strategy (asis) for boosting mcmc efficiency. *Journal of Computational and Graphical Statistics*, 20(3):531–570, 2011.

Appendix A

The derivation of full conditional posterior

Non-centred parametrization

Conditional posterior for μ

$$p(\mu \mid y, h, \phi, \sigma_\omega^2) \propto p(h \mid \mu, \phi, \sigma_\omega^2)p(\mu)$$

$$\propto p(h_1 \mid \mu, h, \phi, \sigma_\omega^2)\prod_{t=1}^{n_{-i}} p(h_{t+1} \mid h_t\mu, h, \phi, \sigma_\omega^2)\mathcal{N}(\alpha_\mu, \beta_\mu)$$

$$\propto exp\Bigg\{-\frac{(h_1-\mu)^2(1-\phi^2)}{2\sigma_\omega^2} - \frac{\sum_{t=1}^{n_{-i}}(h_{t+1}-\mu-\phi(h_t-\mu))^2(1-\phi^2)}{2\sigma_\omega^2}\Bigg\}exp\Bigg\{-\frac{(\mu-\alpha_\mu)^2}{2\beta_\mu^2}\Bigg\}$$

$$\propto exp\Bigg\{-\frac{1}{2}\Bigg[\mu^2\Bigg(\frac{1-\phi^2+(n_{-i}-1)(1-\phi)^2}{2\sigma_\omega^2}+\frac{1}{\beta_\mu^2}\Bigg)-2\mu\Bigg(\frac{h_1(1-\phi^2)+(1-\phi)\sum_{t=1}^{n_{-i}}(h_{t+1}-\phi h_t)^2}{2\sigma_\omega^2}+\frac{\alpha_\mu}{\beta_\mu^2}\Bigg)\Bigg]$$

Remembering the initial assumption the above written formula simplifies to the following, still keeping in mind that $v(h) = \frac{\sigma_\omega^2}{1-\phi^2}$

$$\propto \left\{\frac{\sigma_\omega^2}{(n_{-i}-1)(1-\phi)^2+(1-\phi^2)}\right\}\left\{\frac{h_1(1-\phi^2)}{\sigma_\omega^2}+\frac{(1-\phi)}{\sigma_\omega^2}\sum_{t=1}^{n_{-i}}(h_{t+1}-\phi h_t)^2\right\}$$

Conditional posterior for σ_ω^2

$$p(\sigma_\omega^2 \mid \bar{\eta}, h, \mu, \phi) \propto p(h \mid \mu, \phi, \sigma_\omega^2)p(\sigma_\omega^2)$$

$$\propto p(h_1 \mid \mu, h, \phi, \sigma_\omega^2)\prod_{t=1}^{n_{-i}} p(h_{t+1} \mid h_t\mu, h, \phi, \sigma_\omega^2)\mathcal{IG}(\alpha_{\sigma_\omega}, \beta_{\sigma_\omega})$$

Since the Gamma distribution with shape α_{σ_ω}, scale β_{σ_ω} and has support in $(0, \infty)$. Rearranging I get:

$$\propto \left(\frac{1}{\sigma_\omega^2}\right)^{\frac{n_{-i}}{2}} exp\Bigg\{-\frac{(h_1-\mu_0)^2(1-\phi^2)}{2\sigma_\omega^2}-\frac{\sum_{t=1}^{n_{-i}}(h_{t+1}-\mu_0-\phi(h_t-\mu_0))^2}{2\sigma_\omega^2}\Bigg\}\frac{(\beta_{\sigma_\omega})^{\alpha_{\sigma_\omega}}e^{\frac{-\beta_{\sigma_\omega}}{\sigma_\omega^2}}}{\Gamma(\alpha_{\sigma_\omega})(\beta_{\sigma_\omega})^{\alpha_{\sigma_\omega}+1}}$$

where α_{σ_ω} and β_{σ_ω} are the hyperparameters, which are constant and can be specified by the researcher, hence making the terms $(\beta_{\sigma_\omega})^{\alpha_{\sigma_\omega}}$ and $\Gamma(\alpha_{\sigma_\omega})$ constant. The expression can thus be simplified:

$$\propto exp\Bigg(-\frac{1}{2}\frac{2\beta_{\sigma_\omega} + (h_1-\mu_0)^2(1-\phi^2) + \sum_{t=1}^{n_{-i}}(h_{t+1}-\mu_0-\phi(h_t-\mu_0))^2}{\sigma_\omega^2}\Bigg)\Bigg(\frac{1}{\sigma_\omega^2}\Bigg)^{(\alpha_{\sigma_\omega}+\frac{n_{-i}}{2})+1}$$

Thus the full conditional distribution for σ_ω^2 is:

$$\sigma_\omega^2 \mid h_1 \dots h_n, \phi \sim \mathcal{IG}\Bigg(2.5 + \frac{n_{-i}}{2}, \beta^*\Bigg)$$

where $\beta^* = 0.025 + \frac{(h_1-\mu_0)^2(1-\phi^2)+\sum_{t=1}^{n_{-i}}(h_{t+1}-\mu_0-\phi(h_t-\mu_0))^2}{2}$.

Conditional posterior for σ_ζ^2

$$p(\sigma_\zeta^2 \mid \bar{\eta}, h, \mu, \phi) \propto p(h \mid \mu, \phi, \sigma_e ta^2)p(\sigma_z^2 eta)$$

$$\propto p(h_1 \mid \mu, h, \phi, \sigma_\eta^2)\prod_{t=1}^{n_{-i}} p(h_{t+1} \mid h_t\mu, h, \phi, \sigma_\eta^2)\mathcal{IG}(\alpha_{\sigma_\zeta}, \beta_{\sigma_\zeta})$$

Similarly as for σ_η^2, the full conditional density is proportional to:

$$\propto p(h_1 \mid \mu, h, \phi, \sigma_\eta^2)\prod_{t=1}^{n_{-i}} p(h_{t+1} \mid h_t\mu, h, \phi, \sigma_\eta^2)\mathcal{IG}(\alpha_{\sigma_\zeta}, \beta_{\sigma_\zeta})$$

Which is a Gamma distribution with shape α_{σ_ζ} and scale β_{σ_ζ}. Once again, rearranging I get:

$$\propto \Bigg(\frac{1}{\sigma_\zeta^2}\Bigg)^{\frac{n_{-i}+k}{2}} exp\Bigg\{-\frac{\sum_{i=1}^{n_{-i}}(\bar{\eta}_i-\mu_{c,i})^2(1-\phi^2)}{2\sigma_\zeta^2} - \frac{\sum_{i=1}^{k}(\bar{\eta}_{i+1}-\mu_0-\phi(\bar{\eta}_i-\mu_0))^2}{2\sigma_\zeta^2}\Bigg\}\frac{(\beta_{\sigma_\zeta})^{\alpha_{\sigma_\zeta}} e^{\frac{-\beta_{\sigma_\zeta}}{\sigma_\zeta^2}}}{\Gamma(\alpha_{\sigma_\zeta})(\beta_{\sigma_\zeta})^{\alpha_{\sigma_\zeta}+1}}$$

Thus the conditional posterior is:

$$\sigma^2_\zeta \sim \mathcal{IG}\bigg(\frac{n_{-i}+k}{2}, \frac{1}{2}\Big[\sum_{i=1}^{n_{-i}}(\bar{\eta}_i - \mu_{c,i})^2 + \sum_{i=1}^{k}(\mu_i - \mu_0)^2\Big]\bigg)$$

Conditional posterior for ϕ

$$p(\phi \mid \bar{\eta}, h, \mu, \phi, \sigma^2_\eta) \propto p(h \mid \mu, \phi, \sigma^2_\eta)p(\phi)$$

$$\propto p(h_1 \mid \mu, h, \phi, \sigma^2_\eta)\prod_{t=1}^{n_{-i}} p(h_{t+1} \mid h_t\mu, \phi, \sigma^2_\eta)\mathcal{N}(\alpha_\phi, \beta_\phi)I_{(-1,+1)}(\phi)$$

$$\propto exp\Bigg\{-(\bar{\eta}_i - \frac{(h_1-\mu_0)(1-\phi^2)}{2\sigma^2_\eta} - \frac{\sum_{i=1}^{n_{-i}}(h_{i+1}-\mu_0-\phi(h_t-\mu_0))^2}{2\sigma^2_\eta}\Bigg\} exp\frac{-(\phi-\alpha_\phi)}{2\beta^2_\phi} I_{(-1,+1)}(\phi)$$

$$\propto exp\Bigg\{-\frac{1}{2}\Bigg[\phi^2\Bigg(\frac{(h_1-\mu_0)\sum_{i=1}^{n_{-i}}(h_t-\mu_0)}{2\sigma^2_\eta} + \frac{1}{\beta^2_\phi}\Bigg) - 2\phi\Bigg(\frac{\sum_{i=1}^{n_{-i}}(h_{t+1}-\mu_0)(h_t-\mu_0)}{\sigma^2_\eta} + \frac{\alpha_\phi}{\beta^2_\phi}\Bigg)\Bigg]\Bigg\} I_{(-1,+1)}(\phi)$$

Thus:

$$\phi^* \sim \mathcal{N}_{[-1,1]}\Bigg(\frac{\sum_{i=1}^{n_{-i}}(h_{t+1}-\mu_0)(h_t-\mu_0)}{\sum_{i=1}^{n_{-i}}(h_t-\mu_0)}, \frac{\sigma^2_\eta}{\sum_{i=1}^{n_{-i}}(h_t-\mu_0)}\Bigg)$$

Centred parametrization

The conditional posterior for centred parametrization may bee seen as an adjusted version of the non-centred one, hence I will only give details of the final result of this full conditional posterior.

As explained in the text, centred parametrization was introduced in order to improve the computation when returns show values around zero: while the non-centred parametrization was designed as a Dirichlet process mixture of Normals, the centred one is a normal distribution around a small chosen number, thus I set $\mu_i^* = \mu + \mu_0$ and $h_t^* = h_t - \mu_0$.

Full conditional posterior for μ_0

$$\mu_0 \sim \mathcal{N}\left(\frac{\sum_{i=1}^{k} \mu_i^*}{k)}, \frac{(1-\alpha)\sigma_\zeta^2}{n_2{}'}\right)$$

Full conditional posterior for σ_ω^2

$$\sigma_\omega^2 \mid h_1 \ldots h_n, \phi \sim \mathcal{IG}\left(2.5 + \frac{n_{-i}}{2}, \beta^*\right)$$

where $\beta^* = 0.025 + \frac{h_1^{*2}(1-\phi^2) + \sum_{t=1}^{n-i}(h_{t+1}^* - \phi h_t *)^2}{2}$.

Full conditional posterior for σ_ζ^2

$$\sigma_\zeta^2 \sim \mathcal{IG}\left(\frac{n_{-i}+k}{2}, \frac{1}{2}\left[\sum_{i=1}^{n_{-i}}(\bar{\eta}_i - \mu_{c,i})^2 + \sum_{i=1}^{k}(\mu_i - \mu_0)^2\right]\right)$$

Conditional posterior for ϕ

$$\phi^* \sim \mathcal{N}_{[-1,1]}\left(\frac{\sum_{i=1}^{n_{-i}} h_t^* h_{t+1}^*}{\sum_{i=1}^{n_{-i}} h_t^*}, \frac{\sigma_\eta^2}{\sum_{i=1}^{n_{-i}} h_t^*}\right)$$

Appendix B

Calibrated parameters

Table B.1: Parameters calibrated

Parameter	**Description**	**Value**
μ	Money growth rate	1.01
β	Discount factor	0.996
δ	Depreciation rate	0.025
α	Capital share in production	0.333
$_L$	Constant in labour disutility function	7.5
σ_L	Labor supply elasticity	1
λ_w	Wage markup	1.05
ω_i	Imported investment share	0.52
ω_c	Imported consumption share	0.0285
$\rho_{\bar{\pi}}$	Inflation target persistence	0.975
gr	G/Y ratio	0.2037
τ_y	Labor income tax	0.1771
τ_c	Value added tax	0.1249
$_q$	Cash to money ratio	0.3776
ρ_{τ_k}	Persistence parameter	0.9
ρ_{τ_w}	Persistence parameter	0.9
η_c	Elasticity of substitution	1.5
ν	share of wage in advance	1
ξ_w	Calvo parameter	0.69
ξ_d	Calvo parameter	0.891
ξ_{mc}	Calvo parameter	0.444
ξ_{mi}	Calvo parameter	0.271
ξ_x	Calvo parameter	0.612
ξ_e	Calvo parameter	0.787
κ_w	Indexation parameter	0.497
κ_d	Indexation parameter	0.217
κ_{mc}	Indexation parameter	0.5
κ_{mi}	Indexation parameter	0.5
κ_x	Indexation parameter	0.5
λ_d	Mark-up parameter	1.2
λ_{mc}	Mark-up parameter	1.2
λ_{mi}	Mark-up parameter	1.2
b	Habit parameter	0.7
η_i	Elasticity of substitution	1.5
η_f	Elasticity of substitution	1.5
μ_z	Technology growth	1.005
τ_k	Capital income tax	0.135
τ_w	Labour pay-roll tax	0.197

Appendix C

Impulse response functions

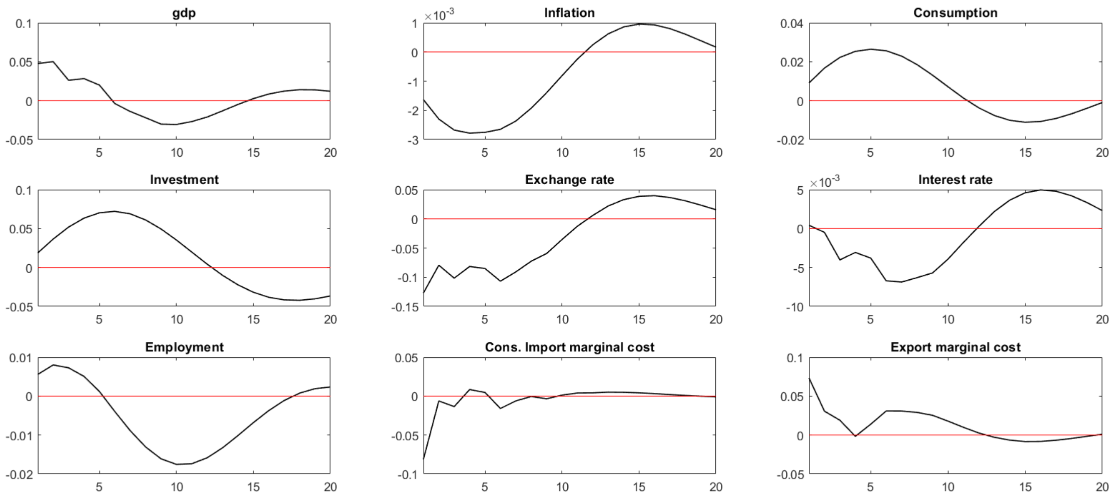

Figure C.1: DSGE response to a foreign gdp shock

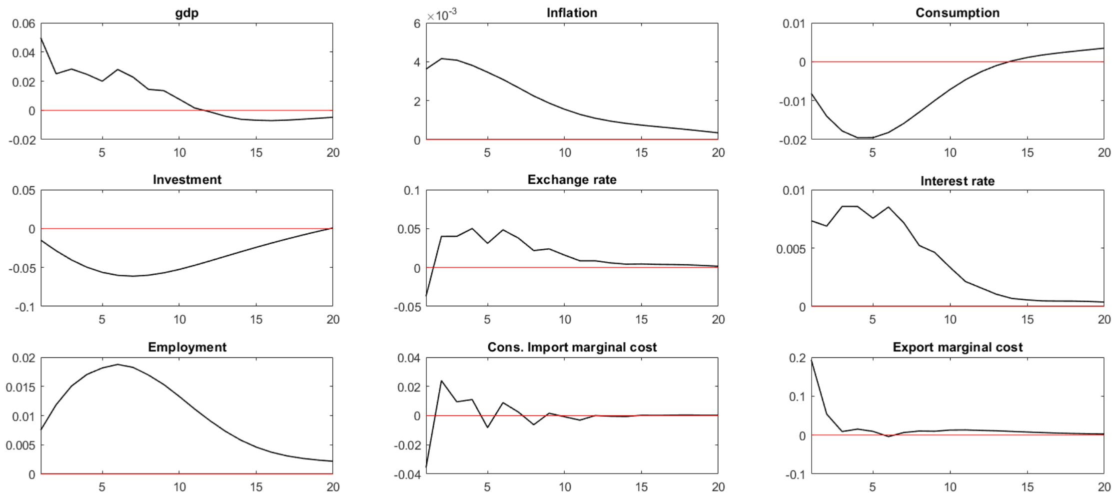

Figure C.2: DSGE response to a foreign inflation shock

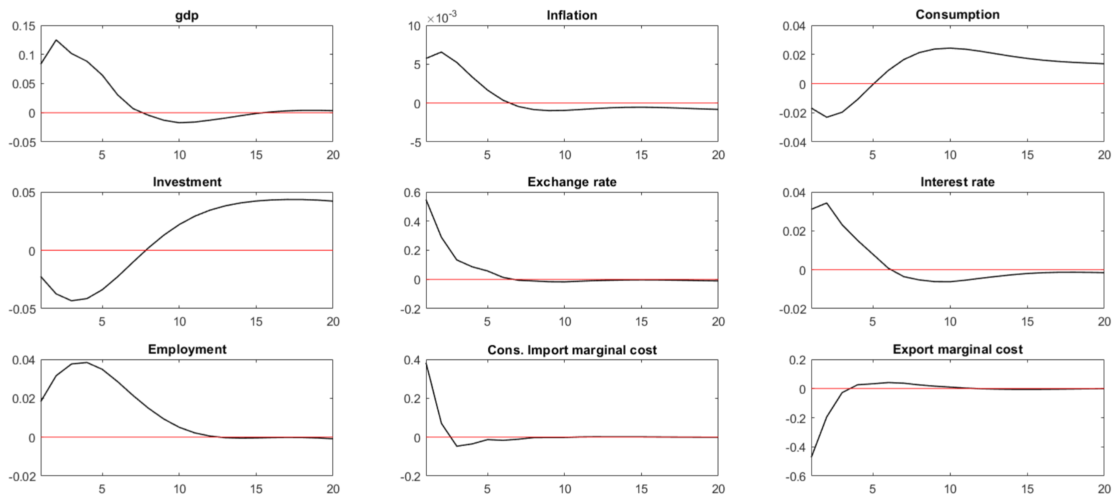

Figure C.3: DSGE response to a foreign interest rate shock

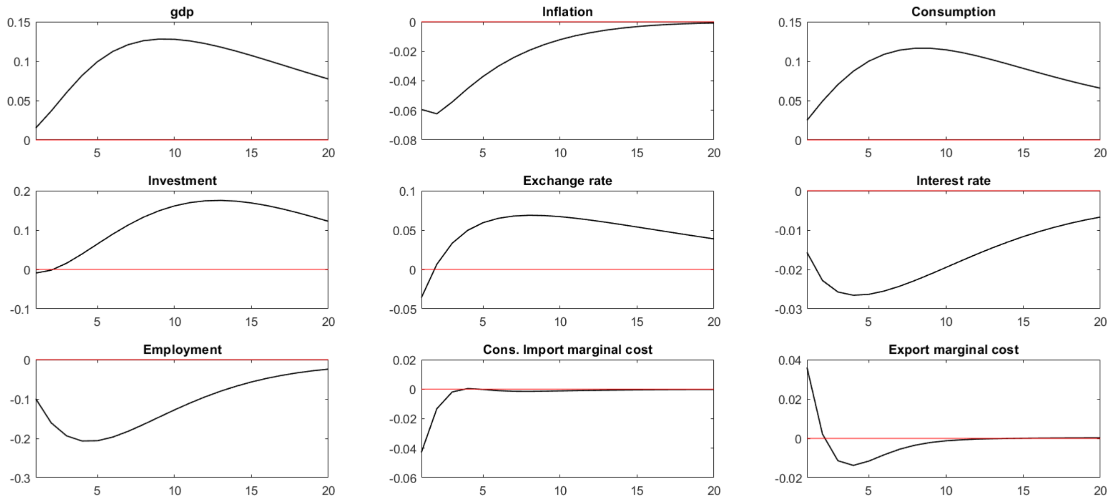

Figure C.4: DSGE response to a transitory technology shock

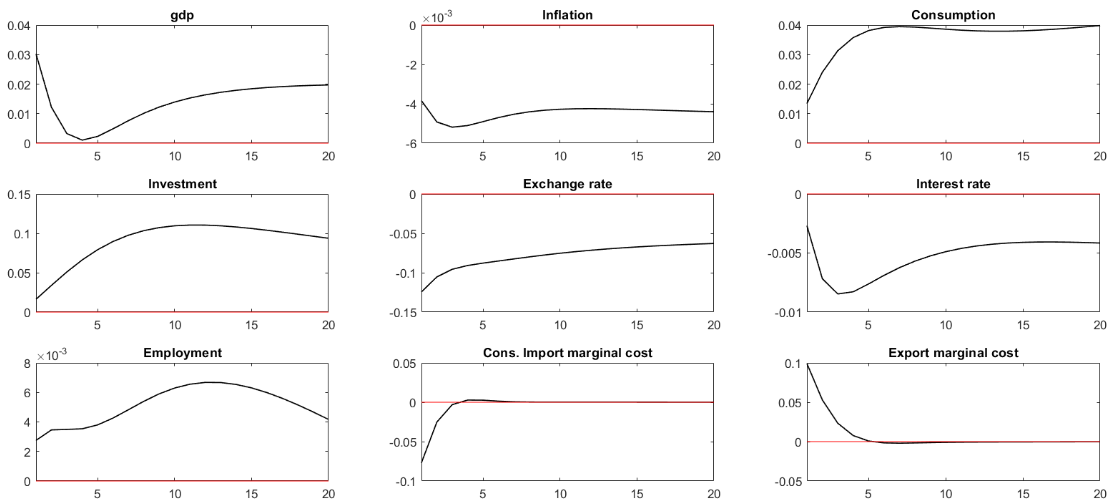

Figure C.5: DSGE response to a permanent technology shock

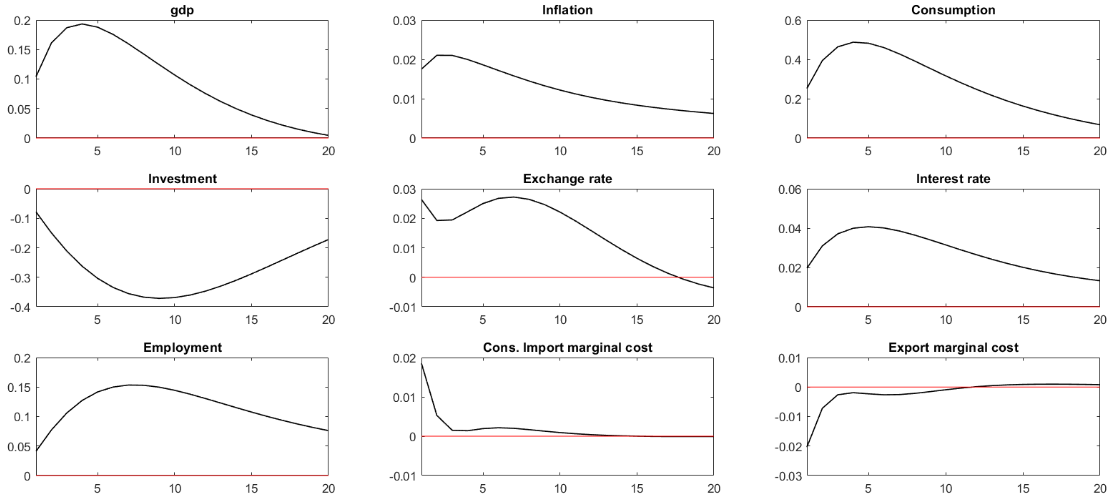

Figure C.6: DSGE response to a consumption preference shock

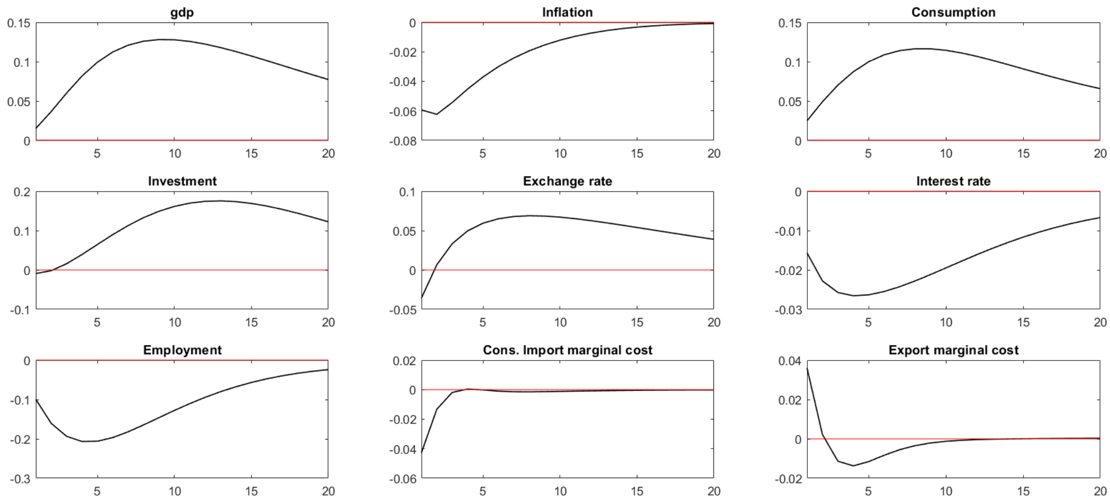

Figure C.7: DSGE response to a consumption tax shock

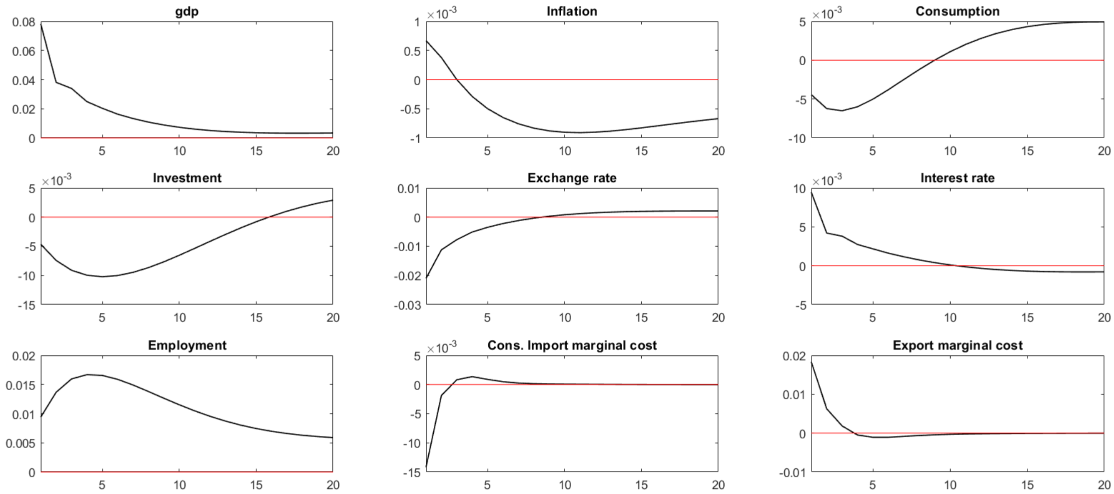

Figure C.8: DSGE response to a government spending shock

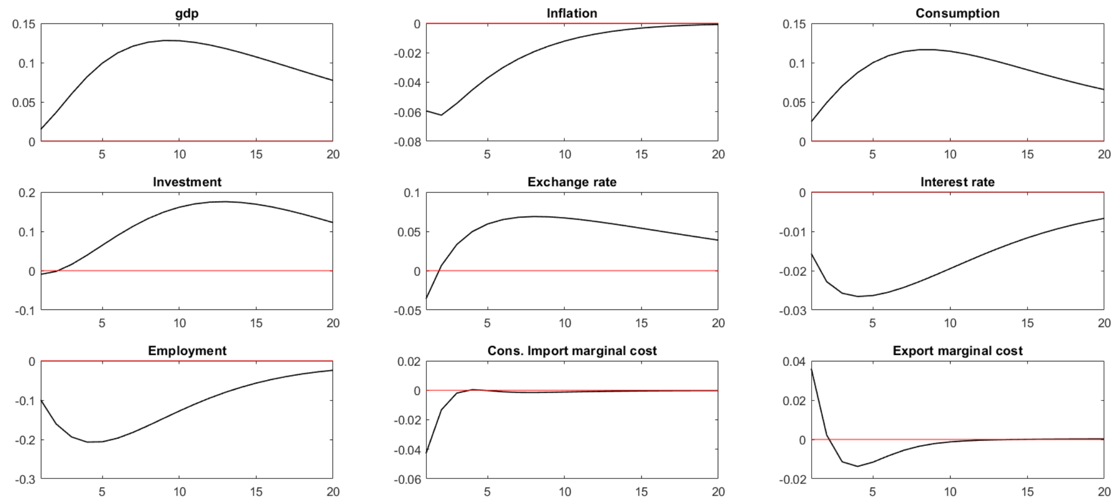

Figure C.9: DSGE response to a interest rate shock