University LUISS "Guido Carli"
of Rome

Ph.D Program in Economics (XXIV Cycle)

Factor Models and Dynamic Stochastic General Equilibrium models: a forecasting evaluation

Gabriele Palmegiani
a.y. 2011/2012

Advisor: Prof. Marco Lippi
Ph.D Coordinator: Prof. Pierpaolo Benigno

Abstract

This dissertation aims to put dynamic stochastic general equilibrium (DSGE) forecasts in competition with factor models (FM) forecasts considering both static and dynamic factor models as well as regular and hybrid DSGE models. The empirical study shows three main conclusions. First, DSGE models are significantly outperformed by the generalized dynamic factor model (GDFM) in forecasting output growth in both short and long run, while the diffusion index (DI) model outperforms significantly DSGE models only in the short run. Second, the most surprising result of the dissertation, we discovered that only the hybrid DSGE model outperforms significantly all other competitive models in forecasting inflation in the long run. This evidence falls out with recent papers that found just regular DSGE models able to generate significant better forecasts for inflation in the long run as well as papers where hybrid DSGE models are found to forecast poorly. Third, in most cases, the unrestricted vector autoregressive (VAR) model represents the worse forecasting model. Although our results are consistent with the prevalent literature who gives to factor models the role to forecast output variables and to DSGE models the role to forecast monetary and financial variables, this research documents that exploiting more information on many macroeconomic time series, through hybrid DSGE models, is important not only to obtain more accurate estimates, but also to get significantly better forecasts.

Keywords: Diffusion Index (DI) model, Generalized Dynamic Factor Model (GDFM), Dynamic General Equilibrium (DSGE) model, Data-Rich DSGE (drDSGE) model, Equal Predictive Ability Tests.

Contents

Chapter 1

Introduction

Recent years have seen rapid growth in the availability of economic data. Statisticians, economists and econometricians now have easy access to data on many hundreds of variables that provide the information about the state of the economy. Coinciding with this growth in available data, two main new econometric models that exploit this wider information have been proposed: the factor models (FM) and the Dynamic Stochastic General Equilibrium (DSGE) models. Factor models have been successfully applied when we have to deal with: construction of economic indicators (Altissimo et al. (2010)), business cycle analysis (Gregory et al. (1997) and Inklaar et al. (2003)), forecasting (Stock and Watson (2002a,b) and Forni et al. (2000)), monetary policy (Bernanke and Boivin (2003) and Bernanke et al. (2005)), stock market returns (Ludvigson and Ng (2007)) and interest rates (Lippi and Thornton (2004)). DSGE models have been successfully applied when we have to deal with: forecasting (Smets and Wouters (2002) and Smets and Wouters (2007)), estimation accurancy (Boivin and Giannoni (2006) and Kryshko (2009)), credit and banking (Gerali et al. (2008)), interest term of structure analysis (Amisano and Tristani (2010)) and monetary policy (Boivin and Giannoni (2008)).

Among all these applications, the recent economic global crisis has pointed out how forecasting well is central. For this reason, the main objective of this dissertation is to provide a detailed forecasting evaluation between these two econometric models taking into account of the recent developments in both factor and DSGE modelling. The novel of this research is the expanded range of forecasting models treated. Infact, our forecasting competition considers not only static factor models and regular DSGE models but also dynamic factor models, such as, the so-called Generalized Dynamic Factor Model (GDFM) of Forni et al. (2000) and hybrid DSGE models, such as, the so-called Data-Rich DSGE (drDSGE) following Boivin and Giannoni (2006) and Kryshko (2009). The dissertation is motivated by the fact that

although there are some forecasting discussions on both dynamic factor model and regular DSGE individually, there is no attempt in the literature, to carry out a strong forecasting evaluation between dynamic factor models and hybrid DSGE models. In particular, what is missing is a forecasting comparison between the GDFM and the drDSGE.

The empirical study shows three main conclusions. First, DSGE models are significantly outperformed by the GDFM in forecasting output growth in both short and long run, while the static factor model outperforms significantly DSGE models only in the short run. Second, the most surprising result of the dissertation, we discovered that only the drDSGE outperforms significantly all other competitive models in forecasting inflation in the long run. This evidence falls out with both Wang (2009) who found that a regular DSGE was able to generate significant better forecasts for inflation in the long run, and Paccagnini (2011) where hybrid models are found to forecast poorly. Therefore, the drDSGE outperforms significantly the regular DSGE in forecasting both output growth and inflation, confirming that exploiting more information on many macroeconomic time series, through the drDSGE, is important not only to obtain more accurate estimates, but also to get significant better forecasts. Third, in most cases, the unrestricted VAR is outperformed by the unconditional mean of the time series of interest, confirming that this model should not be used as benchmark model in forecasting comparisons.

This work is closely related with Wang (2009), but while we share some of the features of his study, our analysis is greatly expanded. First, we do not use the simple DSGE model of Del Negro and Schorfheide (2004) but the most elaborated DSGE model of Smets and Wouters (2007). Second, among factor models, we considered also the GDFM of Forni et al. (2000) whose forecasting performance is documented to be superior than the static factor model of Stock and Watson (2002a,b). Third, among DSGE models, we put side by side the regular DSGE model of Smets and Wouters (2007) with its representation in terms of drDSGE following Boivin and Giannoni (2006) and Kryshko (2009). Therefore, our work is also related with the fast growing literature in both factor models and DSGE models. About factor models, Forni et al. (2000) have presented and estimated their GDFM using a two-sised filter of the observations, Stock and Watson (2002a) have introduced their diffusion index model demostrating its ability to outperform autoregressions and small vector autoregressions forecasts, Stock and Watson (2002b) have shown the asymptotically efficiency of static principal components, Bai and Ng (2002) have focused on the efficient estimation of the number of

static factors under a diffusion index model, Forni et. al. (2005) have proposed a refinement of their two-sided filter into a one-sided filter to allow forecasting feasible, Forni et al. (2009) have emphasized how identification schemes in strustural VAR analysis can be adapted in their GDFM, while Stock and Watson (2010) have described in great detail dynamic factor models. About DSGE models, Smets and Wouters (2007) have extended their previous DSGE model allowing more structural shocks and more financial frictions confirming that the DSGE model has a fit comparable to that of bayesian vector autoregression (BVAR) models, Del Negro et al. (2004) and Del Negro et al. (2007) have shown that a relatively simply DSGE model employed as a prior in a VAR is able to improve the forecasting performance of the VAR relative to an unrestricted VAR or a Bayesian VAR, Rubaszek and Skrzypczynski (2008) have emphasized how DSGE model forecasts are poor in forecasting inflation and interest rates in short term, Christoffel et al. (2010) have pointed out that large bayesian VAR can outperfom DSGE forecasts, while Edge et al. (2011) have shown how their DSGE model can forecast poorly inflation and output growth.

The dissertation is organized following *Figure (1.1)*. Given a large data-set, indeed a data-set

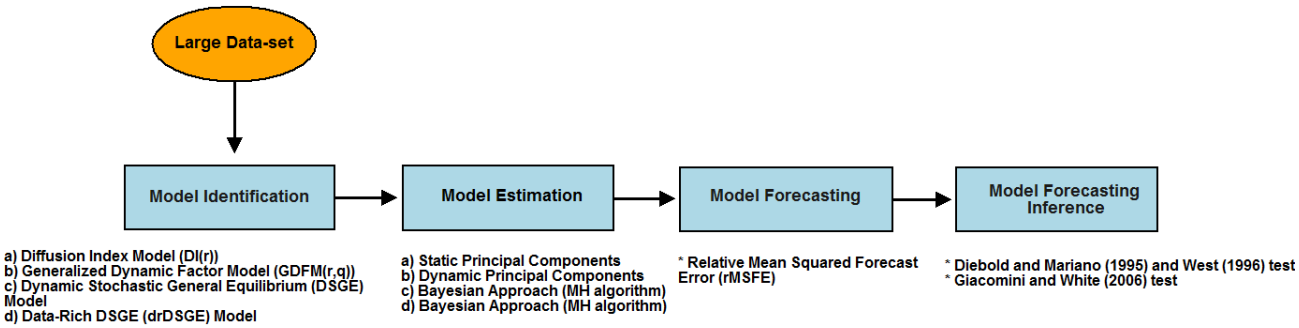

Figure 1.1: The dissertation path.

with many economic time series variables, we evaluate the forecasting performance of factor models relatively to the DSGE models passing through: **identification**, **estimation**, **forecasting** and **forecasting inference**. We open discussing the identification and estimation schemes of both factor models *(Chapter (2))* and DSGE models *(Chapter (3))*. In particular, *Chapter (2)* describes the identification and the estimation of both static and dynamic factor models with a special focus to the recent identification and estimation scheme proposed by Forni and Lippi (2001) of their so-called Generalized Dynamic Factor Model (GDFM), while *Chapter (3)* describes the estimation of the DSGE model of Smets and Wouters (2007) mo-

tivating the advantages of using its Data-Rich Environment version. Estimated the models, we move on forecasting and forecasting inference (*Chapter (4)*). The forecasting step evaluates the models forecasting performance through the *relative mean squared forecast error* ($rMSFE$) metric, defined as:

$$rMSFE(m,n)_{|h} = 1 - \frac{MSFE_{m|h}}{MSFE_{n|h}}$$

where $MSFE_{m|h}$ and $MSFE_{n|h}$ denote respectively the mean squared forecast error generated from model m at the forecasting horizion h and the mean squared forecast error generated from model n at the forecasting horizion h. The $rMSFE(m,n)_{|h}$ can be interpreted as a forecasting gain of model m relative to the model n at the forecasting horizion h when it is positive, or it can be interpreted as a forecasting loss of the model m relative to the model n at the forecasting horizion h when it is negative. In other words, the $rMSFE(m,n)_{|h}$ answers to the question: *...between model m and n, which model should be used to forecast a given time series h steps ahead?* This metric represents an appropriate tool to measure the forecasting perfomance of DSGE models as documented by Smets and Wouters (2003), Smets and Wouters (2007), Edge et al. (2010) and Edge et al. (2011).

As pointed out by Diebold and Mariano (1995) and West (1996) and Giacomini and White (2006), the main drawback of this $MSFE$ analysis is the lack of significance, indeed it is not possible to make rigorous statistical statements by simply interpreting the observed differences between $MSFEs$ because any metric has not a significance power. We need to look into model forecasting inference. We use two test of forecasting accurancy: the test of equal unconditional predictive ability of Diebold and Mariano (1995) and West (1996) (hereafter DMW test), and the test of equal conditional predictive ability of Giacomini and White (2006) (heafter GW test). Since, as shown by Clark and McCracken (2001), the unconditional test has low power in the finite sample, particularly when nested models are involved, the final results of the dissertation come from only the interpretation of the conditional test.

Chapter 2

Factor models

"All models are wrong, but some are useful"

George Box

This chapter presents the identification and the estimation schemes of the factor models used in the out-of-sample forecasting experiments. A *factor model* is an econometric model where each observed time series variable x_{it} is assumed to be linearly decomposed into two unobserved orthogonal components, the common component χ_{it} driven by a small number of common shocks u_{it}, and the idiosyncratic component ξ_{it} who accounts for the residual of that decomposition. The common component is responsable of the co-movement of the series, while the idiosyncratic component is responsable to the specific time series variation. Both the common and the idiosyncratic component are unobserved and need to be consistently estimated.

The chapter is organized as follows. In *Section (2.1)* we start considering the identification and the estimation of the Generalized Dynamic Factor Model (GDFM). In *Section (2.2)* we present the identification and the estimation of the static factor model or diffusion index (DI) model, as special case of the GDFM. In *Section (2.3)* we describe the one-sided estimation and forecasting of the GDFM. In *Section (2.4)* we face the problem of determining the number of factors, while *Section (2.5)* concludes discussing the link between static factor model and the GDFM.

2.1 The generalized dynamic factor model

The Generalized Dynamic Factor model (GDFM) is a factor model that differs from the *exact factor model*, in which the idiosyncratic components are mutually uncorrelated, because it allows the idiosyncratic shocks to be weakly serial and cross-sectional correlated. It combines

the so-called *approximate static factor model* of Chamberlain and Rothschild (1983), widely applied in financial econometrics, and the *dynamic factor model* of Geweke (1977) and Sargent and Sims (1977) for which respectively cross-sectional and serial correlation was allowed. The model is called *dynamic* since the common shocks may not impact a series simultaneously, as in the static factor model, but they can propagate with leads or lags. Then, the model is called *generalized* since the common components are derived assuming a dataset with an infinite number of series and an infinite number of observations.

To setting up the model, let's introduce the notation. Let $\mathcal{P} = (\Omega, \mathcal{I}, P)$ be a probability model and let $L_2(\mathcal{P}, \mathbb{C})$ be the linear space of all complex-valued, zero mean, square-integrable random variables defined on Ω. Let $\mathbf{x} = \{x_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ be the infinite double sequence of random variables defined on $x_{it} \in L_2(\mathcal{P}, \mathbb{C})$ and let $\mathbf{x}_{Nt} = (x_{1t}, x_{2t}, \ldots, x_{Nt})'$ be the finite N-dimensional column vector for the observation made at time t. If $\mathbf{P}$ is a complex matrix we denote $\mathbf{P}'$ as the transpose of $\mathbf{P}$ and $\mathbf{P}^*$ as the complex conjugate of $\mathbf{P}'$. With θ we denote the real interval $[-\pi, \pi]$. Then, given the subset $G \subseteq L_2(\mathcal{P}, \mathbb{C})$, we denote the closed span of G as $span(G)$ which is the minimum closed subspace of $L_2(\mathcal{P}, \mathbb{C})$ containing G. If S is a closed linear subspace of $L_2(\mathcal{P}, \mathbb{C})$ and $\mathbf{x} \subseteq L_2(\mathcal{P}, \mathbb{C})$, we denote $proj(\mathbf{x}|S)$ as the orthogonal projection of $\mathbf{x}$ on S. Therefore, we denote with $\mathbf{\Sigma}^{\mathbf{x}}(\theta)$ the spectral density matrix of the double sequence $\mathbf{x} = \{x_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ as function of the frequencies in $\theta \in [-\pi, \pi]$, while with $\mathbf{\Sigma}^{\mathbf{x}}_N(\theta)$ we denotes the spectral density matrix of the N-dimensional vector $\mathbf{x}_{Nt} = (x_{1t}, x_{2t}, \ldots, x_{Nt})'$ as function of the frequencies in $\theta \in [-\pi, \pi]$. The i-th largest eigenvalue of $\mathbf{\Sigma}^{\mathbf{x}}_N(\theta)$, is denoted by $\lambda^{\mathbf{x}}_{Ni}$, while the i-th largest eigenvalue of $\mathbf{\Sigma}^{\mathbf{x}}(\theta)$ is denoted by $\lambda^{\mathbf{x}}_i(\theta)$. We denote the spectral density matrices of the common and the idiosyncratic componenents and their eigenvalues in a similar way.

We assume that for any $N \in \mathbb{N}$ the process $\mathbf{x}_{Nt}$ is covariance stationary, that is, it has finite variance-covariance matrix: $\mathbb{E}[\mathbf{x}_{Nt}\mathbf{x}'_{N;t-k}] = \mathbf{\Gamma}^{\mathbf{x}}_{Nk}$ and spectral density $\mathbf{\Sigma}^{\mathbf{x}}_N$ with entries σ_{ij} bounded in modulus:

$$\mathbf{\Gamma}^{\mathbf{x}}_{Nk} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ik\theta}\mathbf{\Sigma}^{\mathbf{x}}_N(\theta)d\theta \qquad \mathbf{\Sigma}^{\mathbf{x}}_N = \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty} e^{ik\theta}\mathbf{\Gamma}^{\mathbf{x}}_{Nk}$$

where the spectral density matrix can be estimated applying the discrete Fourier transform to the sample covariance matrix.

Given these assumptions the model proposed by Forni and Lippi (2001) can be defined as following.

Definition 2.1.1 ***The Generalized Dynamic Factor Model:*** *Let q be a nonnegative integer. The double infinite sequence $\mathbf{x} = \{x_{it}, i \in \mathbb{N},\ t \in \mathbb{Z}\}$ is a q-dynamic factor sequence if $L_2(\mathcal{P}, \mathbb{C})$ contains an orthonormal q-dimensional white noise vector process $\mathbf{u} = \{(u_{1t}, u_{2t}, \ldots, u_{qt})'; t \in \mathbb{Z}\} = \{\mathbf{u}_t, t \in \mathbb{Z}\}$ and the double sequence $\boldsymbol{\xi} = \{\xi_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ such that:*

1. *For any $i \in \mathbb{N}$:*

$$\begin{aligned} \mathbf{x}_t &= \boldsymbol{\chi}_t + \boldsymbol{\xi}_t & (2.1) \\ \boldsymbol{\chi}_t &= \underline{\mathbf{b}}_1(L)u_{it} + \ldots + \underline{\mathbf{b}}_q(L)u_{qt} = \sum_{j=1}^{q} \underline{\mathbf{b}}_j(L)u_{jt} = \underline{\mathbf{B}}(L)\mathbf{u}_t & (2.2) \end{aligned}$$

   where $\underline{\mathbf{B}}(L) = \underline{\mathbf{b}}_1(L) + \ldots + \underline{\mathbf{b}}_q(L)$ represents the lag polynominal of order q with $\underline{\mathbf{b}}_i \in L_2^q(\theta; \mathbb{C})$ for any $i \in \mathbb{N}$ and $j = 1, 2, \ldots, q$ (or alternatively each entry $b_{ij} \in L_2(\theta, \mathbb{C})$ for any $i \in \mathbb{N}$ and $j = 1; 2; \ldots; q$).

2. *For any $i \in \mathbb{N}$, $j = 1, 2, \ldots, q$ and $k \in \mathbb{Z}$, we have $\xi_{it} \perp u_{j;t-k}$, then $\xi_{i;t} \perp \chi_{s;t-k}$ for any $i \in \mathbb{N}$, $s \in \mathbb{N}$ and $k \in \mathbb{Z}$.*

3. *$\boldsymbol{\xi}$ is idiosyncratic.*

4. *Putting $\boldsymbol{\chi} = \{\chi_{it},\ i \in \mathbb{N}, t \in \mathbb{Z}\}$, $\lambda_q^{\chi}(\theta) = \infty$ almost everywhere in θ.*

where: $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$ are referred as the vector of **common component** and the vector of the **idiosyncratic component** of $\mathbf{x}_t$, while $\mathbf{u}_t$ is referred to as the vector of common shocks.

The corrisponding model in vector form is:

$$\begin{aligned} \mathbf{x}_{N;t} &= \boldsymbol{\chi}_{Nt} + \boldsymbol{\xi}_{Nt} & (2.3) \\ &= \underline{\mathbf{B}}_N(L)\mathbf{u}_t + \boldsymbol{\xi}_{Nt} \end{aligned}$$

where $\underline{\mathbf{B}}_N(L) = (\mathbf{b}_{N1}(L); \mathbf{b}_{N2}(L); \ldots; \mathbf{b}_{Nq}(L))$ is an $(N \times q)$ matrix.

Example 2.1 ***One dynamic factor GDFM***

Let x_{it} be the time series x at the i-th cross-sectional unit with $i = 1, 2, \ldots, N$, and at time t with $t = 1, 2, \ldots, T$. Stating that x_{it} admits a generalized dynamic factor model representation with one dynamic factor, means decomposing the series as:

$$\begin{aligned} x_{it} &= \chi_{it} + \xi_{it} \\ \chi_{i;t} &= b_{1i}(L)u_{1t} + b_{2t}(L)u_{2t} + \ldots + b_{qi}(L)u_{qt} \end{aligned} \tag{2.4}$$

where: χ_{it} is the common component, and ξ_{it} is the idiosyncratic component. The common component is costructed with q unobserved common shocks or dynamic factors u_{jt} for any $j = 1, 2, \ldots, q$ that are loaded with the filters $b_{ji}(L)$ with leads and/or lags.

2.1.1 The identification of the GDFM

The GDFM model defined in *Equation (2.1)* must be identified. Identification means to find conditions on the variance-covariance of the data $\mathbf{x}_t$ for which the common $\boldsymbol{\chi}_t$ and idiosyncratic component $\boldsymbol{\xi}_t$ are identified. Following Forni et al. (2000), we need to place conditions on the spectral density matrix of the data $\mathbf{x}_t$, indeed on $\boldsymbol{\Sigma}^{\mathbf{x}}(\theta)$, under which the common and idiosyncratic components are identified as N goes to infinity.

Assumption 2.1 *Given the double sequence $\mathbf{x} = \{x_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ where $x_{it} \in L_2(\mathcal{P}, \mathbb{C})$ and given the form:*

$$\begin{aligned} \mathbf{x}_t &= \boldsymbol{\chi}_t + \boldsymbol{\xi}_t \\ &= \underline{\mathbf{B}}(L)\mathbf{u}_t + \boldsymbol{\xi}_t \end{aligned} \tag{2.5}$$

we assume that:

i) *the q-dimensional vector process $\mathbf{u} = \{(u_{1t}, u_{2t}, \ldots, u_{qt})', t \in \mathbb{Z}\}$ is an orthonormal white noise. That is, $\mathbb{E}[u_{jt}] = 0$ and $\mathbb{VAR}[u_{jt}] = \mathbb{E}[u_{jt}u'_{jt}] = 1$ for any j and t; $u_{jt} \perp u_{j;t-k}$ for any j, t, and $k \neq 0$; $u_{jt} \perp u_{s;t-k}$ for any $s \neq j$, t and k.*

ii) *$\boldsymbol{\xi} = \{\xi_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ is the double sequence such that, $\boldsymbol{\xi}_n = \{(\xi_{1t}, \xi_{2t}, \ldots, \xi_{Nt})', t \in \mathbb{Z}\}$ is a zero-mean stationary vector process for any N, and $\xi_{it} \perp u_{j;t-k}$ for any i, j, t, k;*

iii) *the filters $\underline{\mathbf{B}}(L)$ are one-sided in L and their coefficients are square summable for any*

$i \in \mathbb{N}$ *and* $j = 1, 2, \ldots, q$.

This assumption has two implications. First, it implies that the vector $\mathbf{x} = \{x_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ where $x_{it} \in L_2(\mathcal{P}, \mathbb{C})$ is stationary and zero-mean for any N. Second, it implies that the spectral density of the N-dimensional vector $\mathbf{x}_N$, indeed $\mathbf{\Sigma}_N^{\mathbf{x}}(\theta)$, can be written as the sum of the spectral density of the common component $\mathbf{\Sigma}_N^{\chi}(\theta)$ and the spectral density of the idiosyncratic component $\mathbf{\Sigma}_N^{\boldsymbol{\xi}}(\theta)$. These matrices are unobserved, then to obtain their consistent estimation we need further assumptions.

Assumption 2.2 *For any* $i \in \mathbb{N}$*, there exist a real* $c_i > 0$ *such that* $\sigma_{ii}(\theta) \leq c_i$ *for any* $\theta \in [-\pi, \pi]$.

Assumption 2.3 *The first idiosyncratic dynamic eigenvalues* $\lambda_{N1}^{\boldsymbol{\xi}}$ *is uniformly bounded. That is, there exist a real* Λ *such that* $\lambda_{N1}^{\boldsymbol{\xi}}(\theta)$ *for any* $\theta \in [-\pi; \pi]$ *and* $N \in \mathbb{N}$.

Assumption 2.4 *The first* q *common dynamic eigenvalues diverge almost everywhere in* $[-\pi, \pi]$. *That is* $\lim_{n\to\infty} \lambda_{Nj}^{\chi}(0) = \infty$ *for* $j \leq q$*, almost everywhere in* $[-\pi, \pi]$.

The Assumption 2.2 implies that all the entries $\sigma_{ij}(\theta)$ of $\Sigma_N^{\mathbf{x}}(\theta)$ are bounded in modulus, Assumption 2.3 implies that the dynamic eigenvalues of the idiosyncratic components have effects concentrated on a limited number of variables, while Assumption 2.4 implies that each common shock u_{ij} is present in infinitely many cross-sectional units with nondecreasing importance.

If the Assumptions 2.1 to 2.4 are satisfied, Forni and Lippi (2001) show that the double sequence $\mathbf{x} = \{x_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ is a **generalized dynamic factor model**, or better, is a q-generalized dynamic factor model.

2.1.2 Recovering the Common Components in the GDFM

Defined and identified the model, we briefly review this estimation method proposed by Forni et al. (2000) to recovering consistently the common component $\boldsymbol{\chi}_t$ in *Equation (2.1)* starting from the finite N-dimensional process $\mathbf{x}_{Nt} = (x_{it}, x_{2t}, \ldots, x_{Nt})'$. The idea is to be aware that for the the spectral density matrix of the finite process $\mathbf{x}_{Nt}$, indeed $\mathbf{\Sigma}_N^{\mathbf{x}}$, there exist N vectors of complex-valued functions:

$$\mathbf{p}_{nj}(\theta) = (p_{Nj,1}(\theta)\ p_{Nj,2}(\theta)\ \ldots\ p_{Nj,N}(\theta))$$

for $j = 1, 2, \ldots, N$ such that:

a) $\mathbf{p}_{Nj}(\theta)$ is the row eigenvector of $\mathbf{\Sigma}^{\mathbf{x}}_N(\theta)$, that is:

$$\mathbf{p}_{Nj}(\theta)\mathbf{\Sigma}^{\mathbf{x}}_N = \lambda^{\mathbf{x}}_{Nj}(\theta)\mathbf{p}_{Nj}(\theta) \qquad \text{for any} \quad \theta \in [-\pi, \pi] \tag{2.6}$$

b) $|\mathbf{p}_{Nj}(\theta)|^2 = 1$ for any j and $\theta \in [-\pi, \pi]$;

c) $\mathbf{p}_{Nj}(\theta)\mathbf{p}^*_{Ns}(\theta) = 0$ for any $j \neq s$ and any $\theta \in [-\pi, \pi]$;

d) $\mathbf{p}_{Nj}(\theta)$ is θ-measurable on $[-\pi, \pi]$.

Point a) tell us simply that the $\mathbf{p}_{Nj}(\theta)$ for any $j = 1, 2, \ldots, N$ are the eigenvectors associated to the eigenvalues $\lambda^{\mathbf{x}}_{Nj}(\theta)$. These eigenvalues and eigenvectors are called *dynamic*, since they come from spectral eigenvalue decomposition *(Equation (2.6))* and not longer from the contemporaneous variance-covariance matrix decomposition. Point b) affirms that dynamic eigenvectors have unitary length, point c) states that dynamic eigenvectors are orthogonal, while point d) affirms that the dynamic eigenvectors are functions misurable on the interval $[-\pi, \pi]$.

As consequence of properties a) to d) each dynamic eigenvector $\mathbf{p}_{Nj}(\theta)$ can be expanded as Fourier Series:

$$\mathbf{p}_{Nj}(\theta) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} [\int_{-\pi}^{\pi} \mathbf{p}_{Nj}(\theta) e^{ik\theta} d\theta] \; e^{-ik\theta} \tag{2.7}$$

then applying the inverse Fourier transforme to $\mathbf{p}_{Nj}$ we can construct a square-summable, N-dimensional, bilateral filter in the time domain:

$$\underline{\mathbf{p}}_{Nj}(L) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} [\int_{-\pi}^{\pi} \mathbf{p}_{Nj}(\theta) e^{ik\theta}] \; L^k \tag{2.8}$$

where we used the underlined notation to denotes that $\underline{\mathbf{p}}_{Nj}(L)$ is the inverse Fourier transformation of $\mathbf{p}_{Nj}(\theta)$. The product of the dynamic eigenvectors times the data, indeed the scalar process:

$$\mathbf{dpc}_{jt} = \{\underline{\mathbf{p}}_{Nj}(L)\, \mathbf{x}_{Nt} \;,\; t \in \mathbb{Z}\}$$

is the so-called the j-th *dynamic principal component* of $\mathbf{x}_{Nt}$. Notice that, dynamic principal components require to pass from frequency domain to the time domain. Now, to recover the

common component from $\mathbf{x}_{Nt}$, consider the minimal closed subspace of $L_2(\Omega, I, \mathcal{P})$ containing the first q dynamic principal components:

$$\mathcal{U}_N = \overline{span}\left(\mathbf{p}_{Nj}(L)\,\mathbf{x}_{Nt} = \mathbf{dpc}_{jt}\ ,\ j = 1, 2, \ldots, q\ ,\ t \in \mathbb{Z}\right) \tag{2.9}$$

then by projecting the data on the minimal closed subspace containing the first q dynamic principal components, we get the N-dimensional common component:

$$\begin{aligned} \chi_{it,N} &= proj(x_{it}|\mathcal{U}_N) \\ &= \underline{\mathbf{K}}_{Ni}(L)\,\mathbf{x}_{Nt} \end{aligned} \tag{2.10}$$

where $\underline{\mathbf{K}}_{Ni}(L) = p^*_{N1,i}(L)\underline{\mathbf{p}}_{N1}(L) + p^*_{N2,i}(L)\underline{\mathbf{p}}_{N2}(L) \ldots + p^*_{Nq,i}(L)\underline{\mathbf{p}}_{Nq}(L)$ is the filter matrix that extracts the finite N or estimated common component $\chi_{it,N}$ from the finite N-sample data $\mathbf{x}_{Nt}$. Under the Assumption 2.1 and Assumption 2.2 this projection, indeed the estimated common component $\chi_{it,N}$, converges to χ_{it} in mean square as N goes to infinity, indeed: $\lim_{N\to\infty} \chi_{it,N} = \chi_{it}$ in mean square. This result shows that the common component χ_{it} can be recovered asymptotically from the sequence $\underline{\mathbf{K}}_{Ni}(L)\,\mathbf{x}_{Nt}$.

2.2 The static factor model

The problem with the previous estimator is that the filter $\underline{\mathbf{K}}_{Ni}(L)$ used to recover the common component from the data is a two-sided filter. A filter is *two-sided* when the observed variables are related not only with the current and past values of the factors but also with their future values. Although this leaves unaffected the estimate of the central part of the sample, the performance of the estimator deteriorates as we approach the end of the sample. This deterioration makes this method not suitable for forecasting. To outperform this forecasting problem, the literature has proposed two approaches:

- Stock and Watson (2002b) have proposed a new estimation method based on the eigenvalue decomposition of the contemporaneous variance-covariance matrix of $\mathbf{x}_{Nt}$ rather than its spectral eigenvalue decomposition;

- Forni et al. (2005) have proposed the one-sided version of their two-sided filter, which respect to Stock and Watson (2002b) retains the advantages of their dynamic approach,

described in Forni et al. (2000), allowing observed variables to be related only with current and past value of the factors.

The approach of Stock and Watson (2002b) bring us to the so-called *static factor model* or *diffusion index model* (DI model), while the approach of Forni et al. (2005) bring us to the one-sided estimation and forecasting of their generalized dynamic factor model explained in *Section (2.3)*.

The model is called *static* when the vector of factors $\mathbf{F}_t$ are loaded in $\mathbf{x}_{Nt}$ without leads and/or lags, but just contemporaneoulsy. Although the relation between $\mathbf{x}_{Nt}$ and $\mathbf{F}_t$ is static, both $\mathbf{F}_t$ and $\boldsymbol{\xi}_t$ can have a proper law of motion. For example:

$$\underbrace{\mathbf{x}_{Nt}}_{(N\times 1)} = \underbrace{\boldsymbol{\Lambda}_N}_{(N\times r)}\underbrace{\mathbf{F}_t}_{(r\times 1)} + \underbrace{\boldsymbol{\xi}_{Nt}}_{(N\times 1)} \tag{2.11}$$

$$\mathbf{A}(L)\mathbf{F}_t = \boldsymbol{\epsilon}_{Nt} \qquad \boldsymbol{\epsilon}_{Nt} \sim iid\, \mathcal{N}_N(\mathbf{0}; \mathbf{Q}_{\boldsymbol{\epsilon}}) \tag{2.12}$$

$$\boldsymbol{\Psi}(L)\boldsymbol{\xi}_{Nt} = \mathbf{v}_{Nt} \qquad \mathbf{v}_{Nt} \sim iid\, \mathcal{N}_N(\mathbf{0}; \mathbf{R}_{\mathbf{v}}) \tag{2.13}$$

is a static factor model, where: $\mathbf{A}(L) = \mathbf{I} - \mathbf{A}_1 L - \ldots - \mathbf{A}_p L^p$ is the static factors lag polynominal, $\boldsymbol{\Psi}(L) = \mathbf{I} - \boldsymbol{\Psi}_1 L - \ldots - \boldsymbol{\Psi}_s L^s$ is the idiosyncratic lag polynominal, while $\boldsymbol{\epsilon}_{Nt}$ and $\mathbf{v}_{Nt}$ are exogenous shocks of the common and idiosyncratic components respectly. We have the so-called *exact factor model*, if we assume that the matrix $\mathbf{R}_{\mathbf{v}}$ is diagonal, otherwise idiosyncratic shocks are correlated and we have the so-called *approximate factor model*. Because any VAR(p) can be rewritten as VAR(1) using the so-called *companion form*, throughout the dissertation, particular focus will be dedicated to the VAR(1) version of the previous static factor model:

$$\mathbf{x}_t = \boldsymbol{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t \tag{2.14}$$

$$\mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \boldsymbol{\epsilon}_t \qquad \boldsymbol{\epsilon}_t \sim iid\, \mathcal{N}_N(\mathbf{0}; \mathbf{Q}_{\boldsymbol{\epsilon}}) \tag{2.15}$$

$$\boldsymbol{\xi}_t = \boldsymbol{\Psi}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t \qquad \mathbf{v}_t \sim iid\, \mathcal{N}_N(\mathbf{0}; \mathbf{R}_{\mathbf{v}}) \tag{2.16}$$

where we dropped the time series index N for semplicity.

Differently from the specification by Forni and Lippi (2001), the common factors in *Equation*

(2.11) are not required to be uncorrelated in time, and they can also be correlated with the idiosyncratic components. The only condition required for identification is that: $\mathbb{VAR}[\mathbf{F}_t] = \mathbf{I}$, indeed the vector of static common factor has unit lenght. The dimension of $\mathbf{F}_t$ is always equal to $r = q(m+1)$ where q is the dimension of the vector of common shocks $\mathbf{u}_t$.

2.2.1 The estimation of the static factor model

Following Stock and Watson (2002b), let $\mathbf{\Gamma}^{\chi}_{N0}$ and $\mathbf{\Gamma}^{\xi}_{N0}$ be the variance-covariance matrices of the common component $\boldsymbol{\chi}_{Nt}$ and the idiosyncratic component $\boldsymbol{\xi}_{Nt}$ respectively. Let μ^{χ}_{Nj} and μ^{χ}_{Nj} be the largest eigenvalues, in descending orders, of $\mathbf{\Gamma}^{\chi}_{N0}$ and $\mathbf{\Gamma}^{\xi}_{N0}$ respectively.

Assumption 2.5 *We assume that:*

a) $\lim_{N\to\infty} \mu^{\chi}_{Nj} = \infty$ *for* $1 \leq j \leq r$*;*

b) *there exists a real* M*, such that* $\mu^{\xi}_{Nj} \leq M$ *for any* N*.*

Assumption 2.5 point a) establishes that, as N increases, the variance of $\mathbf{x}_{Nt}$ explained by the first r eigenvalues of the common component increases to infinity. This means that as N goes to infinity the weight of the idiosyncratic component in explaining $\mathbf{\Gamma}^{\mathbf{x}}_{N0}$ becomes smaller and smaller. Assumtpion 2.5 point b) allows that the idiosyncratic components can be correlated, but the assumption sets a limit to the amount of this correlation. As N increases, the variance of the vector $\mathbf{x}_{Nt}$ captured by the largest eigenvalue of the idiosyncratic component μ^{χ}_{Nr}, remains bounded. Then, under both point a) and point b) of the Assumption 2.5, Stock and Watson (2002b) shows that the static projection on the first r static principal components of $\mathbf{x}_{Nt}$ converge in mean square to the common component in *Equation (2.11)* for $N \to \infty$.

To recover the common component $\mathbf{\Lambda}_N\mathbf{F}_t$ in *Equation (2.11)*, we need to estimate the vector of static factors $\mathbf{F}_t$. Assume we are working on an empirical application with the finite process $\mathbf{x}_{NT} = \{x_{it}, i = 1, 2, \ldots, N, t = 1, 2, \ldots, T\}$, Stock and Watson (2002b) have proposed to estimate $\mathbf{F}_t$ as the r largest *static principal components* (SPC) starting from the estimated contemporaneous variance-covariance matrix $\hat{\mathbf{\Gamma}}^{\mathbf{x}}_0 = T^{-1}\sum_{t=1}^{T} \mathbf{x}_{NT,t}\mathbf{x}'_{NT,t}$. The first principal component is the linear combination of the observed variables that has maximum variance. It is defined as the vector: $\mathbf{spc}_{1t} = \hat{\boldsymbol{\alpha}}_{N1}\mathbf{x}_{Nt}$. The second principal component is the linear combination of the observed variables that has maximum variance after the first one and it

is uncorrelated with the first one. It is defined as the vector: $\mathbf{spc}_{2t} = \hat{\boldsymbol{\alpha}}_{N2}\mathbf{x}_{Nt}$. To recover the common component we need exactly r static principal components, then the r-th static principal component will be the vector: $\mathbf{spc}_{rt} = \hat{\boldsymbol{\alpha}}_{Nr}\mathbf{x}_{Nt}$. To estimate the number of static factors r, we have used the Alessi et al. (2007) criterion.

To derive these SPC, we need to maximize the variance explained by each principal component. Because we assumed that the data-set has zero mean, the variance of the first principal component is: $\mathbb{VAR}[\hat{\boldsymbol{\alpha}}_{N1}\mathbf{x}_{Nt}] = \mathbb{E}[\hat{\boldsymbol{\alpha}}_{N1}\mathbf{x}_{Nt}(\hat{\boldsymbol{\alpha}}_{N1}\mathbf{x}_{Nt})'] = \hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}}\hat{\boldsymbol{\alpha}}'_{N1}$. This variance can be incerased without limit unless we impose the unity lenght contraint $\hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\alpha}}'_{N1} = 1$. The problem becomes to maximize $\hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}}\hat{\boldsymbol{\alpha}}'_{N1}$ subject $\hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\alpha}}'_{N1} = 1$. The Lagrangian of constrained maximization problem is:

$$\mathcal{L} = \hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}}\hat{\boldsymbol{\alpha}}'_{N1} - \mu_1(\hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\alpha}}'_{N1} - 1) \tag{2.17}$$

where μ_1 is the Lagrange multiplier. Differentiation with respect to $\hat{\boldsymbol{\alpha}}_{N1}$ produces:

$$\frac{\partial\mathcal{L}}{\partial\hat{\boldsymbol{\alpha}}_{N1}} = 2\hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}} - 2\mu_1\hat{\boldsymbol{\alpha}}_{N1} = 0 \quad \Rightarrow \quad \hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}} = \mu_1\hat{\boldsymbol{\alpha}}_{N1} \tag{2.18}$$

Indeed, $\hat{\mu}_1$ is an eigenvalue of $\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}}$ and $\hat{\boldsymbol{\alpha}}_{N1}$ is the associated eigenvector. To decide which eigenvector with maximum variance results from the product: $\hat{\boldsymbol{\alpha}}\mathbf{x}_{Nt}$, let's multiply by $\hat{\boldsymbol{\alpha}}'_{N1}$, we obtain:

$$\overbrace{\hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}}\hat{\boldsymbol{\alpha}}'_{N1}}^{\mathbb{VAR}[\hat{\boldsymbol{\alpha}}_{N1}\mathbf{x}_{Nt}]} = \hat{\mu}_1\overbrace{\hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\alpha}}'_{N1}}^{1} \quad \Rightarrow \quad \hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}}\hat{\boldsymbol{\alpha}}'_{N1} = \hat{\mu}_1\hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\alpha}}'_{N1} = \hat{\mu}_1 \tag{2.19}$$

So, to maximize the variance $\hat{\mu}_1$ must be as large as possible. Thus, $\hat{\boldsymbol{\alpha}}_{N1}$ is the eigenvector corrisponding to the largest eigenvalue of $\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}}$ and $\mathbb{VAR}[\hat{\boldsymbol{\alpha}}_{N1}\mathbf{x}_{Nt}] = \hat{\mu}_1$ is the largest eigenvalue of $\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}}$.

When we introduce the second principal component we require the variance of the sum of the two to be maximum. The Lagrangian is:

$$\mathcal{L} = \hat{\boldsymbol{\alpha}}_{N;1}\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}}\hat{\boldsymbol{\alpha}}'_{N1} + \hat{\boldsymbol{\alpha}}_{N2}\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}}\hat{\boldsymbol{\alpha}}'_{N2} - \mu_1(\hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\alpha}}'_{N1} - 1) - \mu_2(\hat{\boldsymbol{\alpha}}_{N2}\hat{\boldsymbol{\alpha}}'_{N2} - 1) \tag{2.20}$$

where μ_2 is the Lagrange multipliers corresponding to the second principal component. Differenzation respect to $\hat{\boldsymbol{\alpha}}_{N1}$ and $\hat{\boldsymbol{\alpha}}_{N2}$ produces:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \hat{\boldsymbol{\alpha}}_{N1}} &= 2\hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}} - 2\mu_1\hat{\boldsymbol{\alpha}}_{N1} = 0 \;\Rightarrow\; \hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}} = \mu_1\hat{\boldsymbol{\alpha}}_{N1} \;\Rightarrow\; \hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}}\hat{\boldsymbol{\alpha}}'_{N1} = \hat{\mu}_1\hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\alpha}}'_{N1} = \hat{\mu}_1 \\ \frac{\partial \mathcal{L}}{\partial \hat{\boldsymbol{\alpha}}_{N2}} &= 2\hat{\boldsymbol{\alpha}}_{N2}\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}} - 2\mu_2\hat{\boldsymbol{\alpha}}_{N2} = 0 \;\Rightarrow\; \hat{\boldsymbol{\alpha}}_{N2}\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}} = \mu_1\hat{\boldsymbol{\alpha}}_{N2} \;\Rightarrow\; \hat{\boldsymbol{\alpha}}_{N2}\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}}\hat{\boldsymbol{\alpha}}'_{N2} = \hat{\mu}_1\hat{\boldsymbol{\alpha}}_{N2}\hat{\boldsymbol{\alpha}}'_{N2} = \hat{\mu}_2\end{aligned}$$

Thus, the first order conditions are maximized if we consider the first two largest eigenvalues of $\hat{\boldsymbol{\Gamma}}_{N0}^{\mathbf{x}}$. In other words, the sum of first two eigenvalues $\hat{\mu}_1 + \hat{\mu}_2$ maximizes the sum of variances given the unity lenght constraints. Therefore, since the second principal component must be orthogonal to the first one, we have:

$$\mathbb{E}[\hat{\boldsymbol{\alpha}}_{N1}\mathbf{x}_{Nt}(\hat{\boldsymbol{\alpha}}_{N2}\mathbf{x}_{Nt})'] = \hat{\boldsymbol{\alpha}}_{N1}\overbrace{\mathbb{E}[\mathbf{x}_{Nt}\mathbf{x}'_{Nt}]}^{\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}}}\hat{\boldsymbol{\alpha}}'_{N2} = \overbrace{\hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\Gamma}}_0^{\mathbf{x}}}^{\mu_1\hat{\boldsymbol{\alpha}}_{N1}}\hat{\boldsymbol{\alpha}}'_{N2} = \mu_1\overbrace{\hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\alpha}}'_{N2}}^{\hat{\boldsymbol{\alpha}}_{N1}\perp\hat{\boldsymbol{\alpha}}_{N2}} = 0$$

because eigenvectors are by definition orthogonal, indeed $\hat{\boldsymbol{\alpha}}_{N1}\hat{\boldsymbol{\alpha}}_{N2} = 0$ or $\hat{\boldsymbol{\alpha}}_{N1} \perp \hat{\boldsymbol{\alpha}}_{N2}$. Iterating this procedure r times we get all the required principal components. Computing the eigenvectors of the variance-covariance matrix of $\mathbf{x}_{Nt}$ is equivalent to solve the so-called static principal component (SPC) problem, defined as:

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_{Nj} &= argmax_{\mathbf{d}\in\mathbb{R}^N} \;\; \mathbf{d}\boldsymbol{\Gamma}_{N0}^{\mathbf{x}}\mathbf{d}' \\ \text{subject to} &\quad \mathbf{d}\mathbf{d}' = 1 \\ \text{and} &\quad \mathbf{d}\hat{\boldsymbol{\alpha}}'_{Ni} = 0 \qquad \text{for} \quad 1 \leq i < j\end{aligned} \tag{2.21}$$

for $j = 1; 2; \ldots; r$. For $r = 2$, we have shown that the solutions of this maximization problem are the eigenvectors corresponding to the r largest eigenvalues of $\boldsymbol{\Gamma}_{N0}^{\mathbf{x}}$. Then, ordering the eigenvalues $\hat{\mu}_j$ in descending order and taking the eigenvectors corresponding from the largest eigenvalue to smallest, we define:

$$\widehat{\mathbf{SPC}}_t = (\hat{\boldsymbol{\alpha}}_{N1}\mathbf{x}_{Nt} \quad \hat{\boldsymbol{\alpha}}_{N2}\mathbf{x}_{Nt} \quad \ldots \quad \hat{\boldsymbol{\alpha}}_{Nr}\mathbf{x}_{Nt})'$$

as the vector of static principal component of $\mathbf{x}_{Nt}$.

2.3 One sided estimation and forecasting of Forni et al. (2005)

The filters $\underline{\mathbf{K}}_{Ni}(L)$ in *Equation (2.10)* that extract the common components from the infinite data-set $\mathbf{x}_{Nt}$ are unknown and must be estimated. In practice, we have to deal with finite samples, then we need to extract the common components from $\mathbf{x}_{NT} = \{x_{it}, i = 1, 2, \ldots, N, t = 1, 2, \ldots, T\}$, rather than its infinite couterpart $\mathbf{x}_{Nt}$. The idea is to estimate the filters $\underline{\mathbf{K}}_{Ni}(L)$ under the assumption that $\mathbf{x}_{Nt}$ admits a **Wold representation**. Infact, if $\mathbf{x}_{Nt}$ admits a Wold representation, any periodogram smoothing or lag-window estimator $\hat{\mathbf{\Sigma}}^{\mathbf{x}}(\theta)$ is a consistent estimator of $\mathbf{\Sigma}^{\mathbf{x}}_{N}(\theta)$ for T going to infinity. Therefore, also $\hat{\mathbf{\Sigma}}^{\mathbf{x}}(\theta)$ is unknown, but it can be estimated applying the discrete Fourier transform to the sample variance-covariance matrix of $\mathbf{x}_{NT}$. Let's consider first, the estimation of $\hat{\mathbf{\Sigma}}^{\mathbf{x}}(\theta)$, then we consider the estimation of the filter $\underline{\mathbf{K}}_{Ni}(L)$ and its refinement as one sided filter.

Under the assumtpion that $\mathbf{x}_{Nt}$ admits a Wold representation:

$$\mathbf{x}_{Nt} = \sum_{k=-\infty}^{\infty} \mathbf{C}_k \mathbf{w}_{t-k} \tag{2.22}$$

where: $\{\mathbf{w}_t,\ t \in \mathbb{Z}\}$ is a second-order white noise with nonsingular covariance matrix and finite fourth-order moments, and the (ij)-th entries of the matrices $\mathbf{C}_k$ satisfies $\sum_{k=-\infty}^{\infty} |C_{i,j,k}||k|^{1/2} \leq \infty$ for all $N, i, j \in \mathbb{N}$. Given the sample variance-covariance matrix of the finite process $\mathbf{x}_{NT}$:

$$\hat{\mathbf{\Gamma}}^{\mathbf{x}}_{Nk} = (T-k)^{-1} \sum_{t=k+1}^{T} \mathbf{x}_{NT,t} \mathbf{x}'_{NT,t} \tag{2.23}$$

with $k = -M, \ldots, M$ is the lag order fixed using the so-called Bartlett lag-window $M = M(T)$. Infact, to allow estimation, the number of variance-covariance matrices has to be truncated trought the Bartlett lag-window. Estimated the sample variance-covariance matrices $\hat{\mathbf{\Gamma}}^{\mathbf{x}}_{k}$, we can estimate the spectral density matrix $\hat{\mathbf{\Sigma}}^{\mathbf{x}}(\theta)$ by applying the discrete Fourier transformation to $\hat{\mathbf{\Gamma}}^{\mathbf{x}}_{k}$. To avoid biases caused by the truncation, we need to multiply the sample variance-covariance matrices by the Bartlett weights $w_k = 1 - \frac{|k|}{M+1}$:

$$\hat{\mathbf{\Sigma}}^{\mathbf{x}}(\theta_h) = \frac{1}{2\pi} \sum_{k=-M}^{M} w_k \hat{\mathbf{\Gamma}}^{\mathbf{x}}_{Nk} e^{-ik\theta_h} \tag{2.24}$$

where: $\theta_h = 2\pi h/(2M+1)$ are the number of frequencies in which the spectral densities are estsimated, $h = 0, 1, \ldots, 2M$ are the total number of points in which the Fourier transformation is worked out, and w_k are the weights corresponding to the Bartlett lag-window of size $M = M(T)$. The estimation of the spectral density allows to decompose the variance-covariance matrices into periodic components, fruitful to explain the dynamics of the data-set. The choice of M represents the trade off between small bias (large M) and small variance (small M). Forni et al. (2000) show that fixing M as the nearest integer to the square root of the number of observations in the data-set performes well. Consistent estimates are ensured, provided that $M(T) \to \infty$ and $M(T)/T \to 0$ as $T \to \infty$.

Now, we can observe that the estimated filters $\hat{\underline{\mathbf{K}}}_{Ni}(L)$ are infinite two-sided, that is:

$$\hat{\underline{\mathbf{K}}}_{Ni}(L) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} [\int_{-\pi}^{\pi} \hat{\mathbf{K}}_{Ni}(\theta) e^{ik\theta} d\theta] L^k$$

where, as we did before, we used the underlined notation to denotes that $\hat{\underline{\mathbf{K}}}_{Ni}(L)$ is the inverse Fourier transformation of $\hat{\mathbf{K}}_{Ni}(\theta)$. But, because $\mathbf{x}_{Nt}$ is not available neither for $t \leq 0$ nor $t > T$, the projection $\hat{\underline{\mathbf{K}}}_{Ni}(L)\mathbf{x}_{Nt}$ onto the space spanned by the q dynamic principal components cannot be calculated. Therefore, to allow estimation, a truncated version of the filter may be used:

$$\hat{\underline{\mathbf{K}}}_{Ni}(L) = \sum_{k=-M}^{M} \hat{\mathbf{K}}_{Ni,k} L^k$$

The method discussed above produces an estimator of the common component wich is two-sided. As discussed before, this approach has the advantage of exploring the dynamic structure of the data, but the performance of the estimated common component deteriorates as t approaches the end of the sample, indeed 1 or T. Indeed, to compute the estimator for the last observation, one needs M future observations which are not available, this problem makes forecasting not possible.

To allow forecasting, Forni et al. (2005) propose a refinement of their procedure which retains the advantages of the dynamic approach, but permits to obtain a consistent estimate of the optimal forecast of the common component of $\mathbf{x}_{Nt}$ as a one-sided filter of the observations. This method consists of two steps: in the **first step**, they follow Forni et al. (2000) get-

ting estimates of the variance-covariance matrices for the commmon and the idiosyncratic components as the inverse Fourier transform of the spectral density matrix of the common and idiosyncratic component respectively, then in the **second step**, they use these estimates to construct r contemporaneous linear combination of the observations with the smallest idiosyncratic common variance ratio. In other words, they compute the eigenvalues and the eigenvectors of the couple $(\hat{\mathbf{\Gamma}}^{\boldsymbol{\chi}}_{N0}(\theta), \hat{\mathbf{\Gamma}}^{\boldsymbol{\xi}}_{N0}(\theta))$, then, ordering the eigenvalues in descending order and taking the eigenvectors corresponding to the r largest ones, they obtain the so-called generalised principal components that allow efficient estimates and forecasts of the common component of $\mathbf{x}_{Nt}$ without the need of future values. Let's inspect these steps in a more detailed way:

first step: The first step follows Forni et al. (2000). The step is dedicated to the estimation of the variance-covariance matrices of the common $\mathbf{\Gamma}^{\boldsymbol{\chi}}_{Nk}$ and the idiosyncratic component $\mathbf{\Gamma}^{\boldsymbol{\xi}}_{Nk}$, starting from an estimator of the spectral density matrix of the data-set, indeed $\mathbf{\Sigma}^{\mathbf{x}}_{N}(\theta)$. We discussed that, under the assumption that $\mathbf{x}_{Nt}$ admits a Wold representation, any periodogram smoothing or lag window estimator $\hat{\mathbf{\Sigma}}^{\mathbf{x}}(\theta)$ of $\mathbf{x}_{NT}$ represents a consistent estimator of $\mathbf{\Sigma}^{\mathbf{x}}_{N}(\theta)$ of $\mathbf{x}_{Nt}$. Now, using the Assumption 2.1, we can decompose the estimated spectral density $\hat{\mathbf{\Sigma}}^{\mathbf{x}}(\theta)$ into the sum of the a spectral density matrix of the common and the idiosyncratic component, indeed:

$$\hat{\mathbf{\Sigma}}^{\mathbf{x}}(\theta) = \hat{\mathbf{\Sigma}}^{\chi}(\theta) + \hat{\mathbf{\Sigma}}^{\xi}(\theta)$$

where $\hat{\mathbf{\Sigma}}^{\chi}(\theta) = \sum_{j=1}^{q} \hat{\mathbf{p}}^{\mathbf{x}*}_{j}(\theta)\hat{\lambda}^{\mathbf{x}}_{j}(\theta)\hat{\mathbf{p}}^{\mathbf{x}}_{j}$ and $\hat{\mathbf{\Sigma}}^{\xi}(\theta) = \sum_{l=q+1}^{n} \hat{\mathbf{p}}^{\mathbf{x}*}_{l}(\theta)\hat{\lambda}^{\mathbf{x}}_{l}(\theta)\hat{\mathbf{p}}^{\mathbf{x}}_{l}$ are the estimated spectral density matrices of the common and idiosyncratic component respectively, while $\hat{\mathbf{p}}^{\mathbf{x}*}_{j}$ is the j-th complex conjugate eigenvector of $\hat{\mathbf{\Sigma}}^{\mathbf{x}}(\theta)$. Then, applying the inverse discrete Fourier transformation to these density matrices, the covariance matrices of the common $\boldsymbol{\chi}_t$ and idiosyncratic component $\boldsymbol{\xi}_t$ can be estimated as:

$$\hat{\mathbf{\Gamma}}^{\boldsymbol{\chi}}_{Nk}(\theta) = \int_{-\pi}^{\pi} \hat{\mathbf{\Sigma}}^{\boldsymbol{\chi}}(\theta)e^{ik\theta}d\theta \tag{2.25}$$

$$\hat{\mathbf{\Gamma}}^{\boldsymbol{\xi}}_{Nk}(\theta) = \int_{-\pi}^{\pi} \hat{\mathbf{\Sigma}}^{\boldsymbol{\xi}}(\theta)e^{ik\theta}d\theta \tag{2.26}$$

These estimated variance-covariance matrices will be used in the second step to solve the so-called generalized principal component (GPC) problem.

Second Step: The second step is dedicated to the estimation of the generalized principal components given the variance-covariance matrices estimated in the first step. More precisely, the objective is to find r independent linear combinations $\mathbf{W}_j = \hat{\mathbf{z}}_j \mathbf{x}_{Nt}$ where the weights $\mathbf{z}_j$ are defined as:

$$\begin{aligned} \hat{\mathbf{z}}_j &= argmax_{g \in \mathbb{R}^N} \ \mathbf{d}\hat{\mathbf{\Gamma}}^{\boldsymbol{\chi}}_{N0}\mathbf{d}' \\ \text{subject to} \quad & \mathbf{d}\hat{\mathbf{\Gamma}}^{\boldsymbol{\xi}}_{N0}\mathbf{d}' = 1 \\ \text{and} \quad & \mathbf{d}\hat{\mathbf{\Gamma}}^{\boldsymbol{\xi}}_{N0}\hat{\mathbf{z}}_l' = 0 \qquad \text{for } 1 \leq l < j \end{aligned} \tag{2.27}$$

for $j = 1, 2, \ldots, r$. The idea is that the information contained in the variance-covariance matrices estimated in the previous step, can be used to determine linear combinations which are more efficient than standard principal components. The improvement in efficiency is produced because the idiosyncratic variance is minimized in the first step. As we have seen before, a problem like that can be solved by compunting the eigenvalues and the eigenvectors. In this case, we need to compute the eigenvalues and the eigenvectors of the couple of matrices $(\hat{\mathbf{\Gamma}}^{\boldsymbol{\chi}}_{N0}, \hat{\mathbf{\Gamma}}^{\boldsymbol{\xi}}_{N0})$, rather than the eigenvalues and eigenvectors of the estimated contemporaneous variance-covariance matrix $\hat{\mathbf{\Gamma}}^{\mathbf{x}}_{N0}$. The aggregates that come from the couple $(\hat{\mathbf{\Gamma}}^{\boldsymbol{\chi}}_{N0}, \hat{\mathbf{\Gamma}}^{\boldsymbol{\xi}}_{N0})$ are called *generalized* to distinguish from the static aggregates that come from $\hat{\mathbf{\Gamma}}^{\mathbf{x}}_{N0}$. Practically, the generalized eigenvalues are the solutions of: $det(\hat{\mathbf{\Gamma}}^{\boldsymbol{\chi}}_{N0} - v_j \hat{\mathbf{\Gamma}}^{\boldsymbol{\xi}}_{N0}) = 0$ for $j = 1, 2, \ldots, r$, while the corresponding generalized eigenvectors are the weights $\hat{\mathbf{z}}_j$ that must satisfy:

$$\hat{\mathbf{z}}_j \hat{\mathbf{\Gamma}}^{\boldsymbol{\chi}}_{N0} = \hat{v}_j \hat{\mathbf{z}}_j \hat{\mathbf{\Gamma}}^{\boldsymbol{\xi}}_{N0} \qquad \text{for } j = 1; 2; \ldots; r \tag{2.28}$$

under the normalization conditions:

$$\hat{\mathbf{z}}_l \hat{\mathbf{\Gamma}}^{\boldsymbol{\chi}}_{N0} \hat{\mathbf{z}}_j' = 1 \qquad \text{for } l = j \tag{2.29}$$

$$\hat{\mathbf{z}}_l \hat{\mathbf{\Gamma}}^{\boldsymbol{\chi}}_{N0} \hat{\mathbf{z}}_j' = 0 \qquad \text{for } l \neq j \tag{2.30}$$

Then ordering the eigenvalues $\hat{v}_j$ in descending order and taking the eigenvectors corresponding to the r largest eigenvalues, we define $\widehat{\mathbf{gpc}}_t = (\hat{\mathbf{z}}\mathbf{x}_{Nt}\ \hat{\mathbf{z}}_2\mathbf{x}_{Nt}\ \ldots\ \hat{\mathbf{z}}_r\mathbf{x}_{Nt})'$ as the first r generalized principal component of $\mathbf{x}_{Nt}$.

2.4 Determing the number of factors: r and q

As mentioned in the introduction, the most important feature of factor models is to summarize the information contained in a large panel of variables using a small number of unobserved variables called factors. The question is that, the exact number of factors to use is not known and it must be estimated. We need to estimate both the number of static factor r and the number of dynamic factors q. In this dissertation the optimal number $\hat{r}$ of static factors is estimated using the criterion proposed by Alessi et al. (2007), whereas the optimal number $\hat{q}$ of dynamic factors is estimated using the criterion proposed by Hallin and Liška (2007). Since, in empirical applications, we have to deal with finite sequences of length T of a finite number N of variables, we describe these two criteria for a finite realization of the form $\mathbf{x}_{NT} = \{x_{it}, i = 1, 2, \ldots, N, t = 1, 2, \ldots, T\}$.

2.4.1 Determining the number of static factors

Alessi et al. (2007) have modified the criterion by Bai and Ng (2002) for determining the number of static factors in approximate factor models. They select the true number of static factors $\hat{r}$ as the number that minimizes the variance explained by the idiosyncratic component, but in order to avoid overparametrization, their minimization is subject to a penalization, indeed, they have modified the original procedure of Bai and Ng (2002) by multiplying the penalty function by a positive real number, which allows us to tune its penalizing power, by analogy with the method used by Hallin and Liška (2007) in the frequency domain. They have shown that their modified criterion is more robust in estimating the true number of static factors than the criterion of Bai and Ng (2002).

Formally, let's suppose that our data-set admits a static factor model as in *Equation (2.14)*, here reported:

$$\mathbf{x}_t = \mathbf{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t \tag{2.31}$$

where $\mathbf{x}_t$ denotes the $(N \times 1)$ vector of observations given a time t for n infinite number of time series, $\mathbf{\Lambda}$ denotes the $(N \times r)$ matrix of static factor loadings, $\mathbf{F}_t$ denotes the $(r \times 1)$ vector of static common factors, and $\boldsymbol{\xi}_t$ denotes the $(N \times 1)$ vector of idiosyncratic components. Considering just one single time series in $\mathbf{x}_t$, we can write:

$$x_{it} = \boldsymbol{\lambda}_i \mathbf{F}_t + \xi_{it} \tag{2.32}$$

where $\boldsymbol{\lambda}_i$ denotes the i-th row of the matrix of factor loadings $\mathbf{\Lambda}$. Bai and Ng (2002) have proposed an information criterion to determine the optimal number of static factors r in *Equation (2.32)* assuming to have k common static factors for the matrices $\boldsymbol{\lambda}_i$ and $\mathbf{F}_t$, denoted by $\boldsymbol{\lambda}_i^{(k)}$ and $\mathbf{F}_t^{(k)}$. Let:

$$V(k) = (N\,T)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \hat{\boldsymbol{\lambda}}_i^{(k)} \hat{\mathbf{F}}_t^{(k)})^2 \tag{2.33}$$

be the residual variance of the idiosyncratic components ξ_{it} when the matrix of factor loadings $\mathbf{\Lambda}_t^{(k)}$ and the common factors $\hat{\mathbf{F}}_t^{(k)}$ are estimated using the method of static principal components as described in *SubSection (2.2.1)*. The idea of Bai and Ng (2002) to minimize $V(k)$ in order to find the optimal number of static factors. They define the following information criterion:

$$\begin{aligned} \hat{r}_{\mathrm{IC\_bn}} &= \mathrm{argmin}_{0 \leq k < r_{max}} \; \mathrm{IC\_bn}(k) \\ \mathrm{IC\_bn}(k) &= \log(V(k)) + k\, p(N,T) \end{aligned} \tag{2.34}$$

where: $\hat{r}_{\mathrm{IC\_bn}}$ is the optimal number of static common factors, $p(N,T)$ is a penalty function which counterbalances the fit improvement due to the inclusion of additional common factors, and r_{max} is the maximun number of static factors. Notice that, when the number k of factors is increased, the variance explained by the factors increases too, then $V(k)$ decreases, so the aim of the penalty function, which is an increasing function of both n and T, is to avoid overparametrization. The information criterion $\mathrm{IC\_bn}(k)$ has to be minimised in order to determine the optimal number of static factors, its consistency is proved by Bai and Ng (2002). In empirical applications we have to fix a maximum number of static factors r_{max}, and estimate the model for all numbers of factors $k = 1; 2; \ldots; r_{max}$. As a penalty function Bai and Ng (2002) propose to use $p(N,T) = \frac{N+T}{N\,T} \log(\min\{N,T\})$.

According to what Hallin and Liška (2007) propose in a similar criterion for the number of dynamic factors q, because a penalty $p(N,T)$ leads to consistent estimation of r if and only if $c\,p(N,T)$ does, where c is an arbitrary positive real number, the idea of Alessi et al. (2007) is to modify the criterion of Bai and Ng (2002) by multiplying the penalty function $p(N,T)$ by a constant c that has no influence on the asymptotic performance of the identification method. The criterion becomes:

$$\begin{aligned} \hat{r}_{\text{IC\_a}} &= \text{argmin}_{0\leq k<r_{max}}\ \text{IC\_a}(k) \\ \text{IC\_a}(k) &= \log(V(k)) + c\,k\,p(N,T) \end{aligned} \tag{2.35}$$

where c is a constant which has the aim to tune the penalizing power of the function $p(N,T)$. Alessi et al. (2007) show that the criterion IC_a(k) corrects the tendency of IC_bn(k) to overestimate the optimal number of static factors and provide a more robust estimation of r than the original criterion IC_bn(k) proposed by Bai and Ng(2002). To select the optimal number of static common factors $\hat{r}$, Alessi et al. (2007) suggest, as Hallin and Liška (2007), an automatic procedure which basically fix the number of static factors in correspondence with the second stationary interval of the variance of the selected $\hat{r}$ for the whole region of values of the constant c.

2.4.2 The number of dynamic factors

Hallin and Liška (2007) have proposed a method for determining q in a GDFM that exploits the relation between the number of dynamic factors and the number of diverging eigenvalues of the spectral density matrix of the finite data-set $\mathbf{x}_{NT}$. The ingredientes of the information criterion are the estimated spectral density of $\mathbf{x}_{NT}$, indeed $\hat{\mathbf{\Sigma}}^{\mathbf{x}}(\theta)$, and its eigenvalues $\lambda_{NT,i}(\theta)$. The criterion proposed is:

$$\begin{aligned} \hat{q}_{\text{IC\_hl}} &= \text{argmin}_{0\leq k<q_{max}}\ \text{IC\_hl}(k) \\ \text{IC\_hl}(k) &= \log\left(\frac{1}{N}\sum_{i=k+1}^{n}\frac{1}{2M_T+1}\sum_{h=-M_T}^{M_T}\lambda_{Ni}^{T}(\theta_h)\right) + c\,k\,\overline{p}(N,T) \end{aligned} \tag{2.36}$$

where: $\theta_h = 2\pi h/T$ for $h = -M_T,\ldots,M_T$ and q_{max} is the maximum number of dynamic factors. The authors suggest using $M_T = [0.5\sqrt{T}]$ or $M_T = [0.7\sqrt{T}]$ and as penalty function

$\overline{p}(N,T) = (M_T^{-2} + M_T^{1/2} + N^{-1}\log A_T)$ with $A_T = (\min\{N, M_T^2, M_T^{-1/2}T^{1/2}\})$. Therefore, the penalty function should be large enough to avoid overestimation of $\hat{q}_{\text{IC}}$, but at the same time it should not over penalize. To select the optimal number of dynamic common factors $\hat{q}$, Hallin and Liška (2007) suggest an automatic procedure which basically fix the number of optimal dynamic factors in correspondence with the second stationary interval of the varaince of the selected $\hat{q}$ for the whole region of values of the constant c.

2.5 From the static factor model to the GDFM

In this section, we show that a static factor model can be rewritten as a GDFM under suitable assumptions. Let's start considering a particular case of the static factor model described in *Equation (2.14)*:

$$\begin{aligned} \mathbf{x}_t &= \mathbf{\Lambda F}_t + \boldsymbol{\xi}_t \\ (\mathbf{I} - \mathbf{A}L)\mathbf{F}_t &= \mathbf{B}\mathbf{u}_t = \boldsymbol{\epsilon}_t \qquad & \boldsymbol{\epsilon}_t \sim iid\, \mathcal{N}_N(\mathbf{0}; \mathbf{Q}_{\boldsymbol{\epsilon}}) \\ (\mathbf{I} - \mathbf{\Psi}L)\boldsymbol{\xi}_t &= \mathbf{v}_t \qquad & \mathbf{v}_t \sim iid\, \mathcal{N}_N(\mathbf{0}; \mathbf{R}_{\mathbf{v}}) \end{aligned} \tag{2.37}$$

where we supposed that the vector $(r \times 1)$ of factor exogenous shocks $\boldsymbol{\epsilon}_t$ depends on the $(q \times 1)$ vector of dynamic factors $\mathbf{u}_t$ trought the $(r \times q)$ matrix $\mathbf{B}$ with $q < r$. By plugging the law of motion of the static factors into the equation of the data-set, we obtain a GDFM as in Forni et al. (2000):

$$\mathbf{x}_t = \overbrace{\mathbf{\Lambda}(\mathbf{I} - \mathbf{A}L)^{-1}\mathbf{B}\mathbf{u}_t}^{\boldsymbol{\chi}_t} + \boldsymbol{\xi}_t \tag{2.38}$$

Then, we have shown that starting from a static or diffusion index model is possible to obtain a GDFM assuming that the vector of r factor exogenous shocks $\boldsymbol{\epsilon}_t$ depends on the vector of q dynamic factors $\mathbf{u}_t$. The relevant question is: *...does Model (2.38) exist?* Indeed, *...can* $\boldsymbol{\epsilon}_t$ *be expressed as the product of a matrix* $\mathbf{B}$ *times the vector of dynamic factors* $\mathbf{u}_t$*?*. The answer is given by Forni et al. (2009). They argues that *Model (2.38)* exists with a finite number of static factors r if and only if the space spanned by the q dynamic factors is finite dimensional. So, if this span is finite dimensional, the static factor model can be rewritten as a GDFM. Therefore, the advantage of *Model (2.38)* respect to *Model (2.14)* is that, by taking into account also the law of motion of the static factors, we can consider also the non-contemporaneous comovements among the observed variables. In this way the model should

be particularly useful for forecasting.

Chapter 3

DSGE models: from regular to Data-Rich Environment

"The more specific and data-rich the model, the more effective it will be"

Jean Boivin and Marc Giannoni

This chapter presents the estimation of the Dynamic Stochastic General Equilibrium (DSGE) models used in the out-of sample forecasting experiments. We work on two types of DSGE models: the DSGE model of Smets and Wouters (2007), (hereafter also referred as *regular* or *no-augmented* DSGE model), and its representation in term of the so-called Data-Rich DSGE model (drDSGE) following Boivin and Giannoni (2006) and Kryshko (2009). The motivation from rewriting and re-estimating the regular DSGE in term of drDSGE, stands on the documented gains provided by Boivin and Giannoni (2006). Although, they have shown that: first, the regular DSGE model is outperformed by the drDSGE in the estimation accurancy:

...exploiting more information (through the drDSGE) is important for accurate estimation of the model's concepts and shocks, and that it implies different conclusions about key structural parameters and the sources of economic fluctuations.

Boivin and Giannoni (2006)

second, better estimates imply better forecasts, at least for one quarter ahead for all pooled observed DSGE variables:

...more precision in estimating these variables implies then more precise forecasts of the indicators.

Boivin and Giannoni (2006)

nothing has been stated on the relative forecasting performance of the drDSGE respect to factor models, especially respect to the GDFM. It remains an open part of the empirical reseach, that we face in detailed way in *Chapter (4)* for two time series of the US economy:

the output growth and the rate of inflation.

The chapter is organized as follows. In *Section (3.1)* we open discussing why we should set aside factor models and build up DSGE models instead. In *Section (3.2)* we focus on the approach of Boivin and Giannoni (2006), while *Section (3.3)* concludes showing the DSGE model of Smets and Wouters (2007) and explaining how it has been estimated in term of Data-Rich Environment.

3.1 Why DSGE models?

As mentioned in the introduction, the main drawback of factor models is the lack of an underlying economic theory. This implies that factor models are constructed on the data rather than using a strong economic theory based on utility-maximising rational agents. This limitation makes, in principle, factor models vunerable to the so-called *Lucas critique* which argues that: it is naive to try to predict the effects of a change in economic policy entirely on the basis of relationships observed in the data because the parameters of those models were not structural, indeed not policy-invariant, and they would necessarily change whenever policy, or the rules of the game, was changed. Then, policy conclusions based on those models would therefore potentially be misleading:

> *...given that the structure of an econometric model consists of optimal decision rules of economic agents, and that optimal decision rules vary systematically with changes in the structure of series relevant to the decision maker, it follows that any change in policy will systematically alter the structure of econometric models*
>
> *Lucas (1976)*

In practice, the Lucas critique suggests that if we want to predict the effect of a policy, we should model the *deep parameters*, such as preferences, technology and resource constraints parameters, that govern the individual behavior, rather than work on the data only. This critique has been so influential that it has encouraged macroeconomists to build microfoundations in their models. In this way DSGE models have been originated.

The point is that also DSGE models have important limitations. Schorfheide (2010) evidences five main limitations or challenges. First, is the fragility of parameter estimates due to lack of identification of the parameters of the model (Canova and Sala (2009)). Second, in

a DSGE model, exogenous disturbances generate macroeconomic fluctuations and we cannot be sure whether these shocks capture aggregate uncertainty or model misspecification and the formal econometrics is weak to distinguish these two interpretations. Third, many time series show low frequency behavior which let the DSGE estimation difficult to implement. Fourth, DSGE models often appear to be misspecified in the sense that VARs are favored by statistical criteria that trade off goodness of in-sample fit against model dimensionality. Fifth, the prediction of the effects of rare policy changes often relies exclusively on extrapolation by theory which makes it difficult to provide measures of uncertainty. For example, Kocherlakota (2007) explains that while a model with the worse statistical fit delivers the better policy prediction, bad fit is not a guarantee of good policy prediction. These limitations produce DSGE model misspecification, which leads to poor estimates and forecasts.

The idea proposed by the literature to get away from both factor models and DSGE models limitations is to combine these models using an *hybrid* or *mixture* or *augmented models*. An hybrid model is an econometric model where a DSGE model is combined with a pure statistical data model (such as: a autoregressive process, a vector autoregressive process, a bayesian vector autoregressive, or a factor model), in order to cover the gap between theory and data mitigating the limitations of each model. Particular attractive is, in our view, the hybrid approach proposed by Boivin and Giannoni (2006) whose representation, estimation and forecasting is dedicated the rest of the chapter. For a survey of hybrid models see Paccagnini (2011).

3.2 The Data-Rich DSGE

This section contains both the representation and the estimation theory of the so-called Data-Rich DSGE (drDSGE) of Boivin and Giannoni (2006), while its forecasting theory is discussed together with all other forecasting models in the next chapter.

First of all, let's introduce the DSGE's notation used throughout the dissertation. Let x_t be a variable at time t, let x^{ss} be the steady state value of x, indeed the value of x not affected by random shocks, let $\hat{x}_t = \log(x_t) - \log(x^{ss})$ be its log equilibrium deviation of x. Following Sims (2002), every DSGE model has around its steady state the following **lin-**

earized representation:

$$\Gamma_0 \mathbf{s}_t = \Gamma_1 \mathbf{s}_{t-1} + \Psi \mathbf{e}_t + \Pi \boldsymbol{\eta}_t \tag{3.1}$$

where: $\mathbf{s}_t$ is the vector of all DSGE endogenous variables (for example: the capital $\hat{k}_t$, or the output growth $\hat{x}_t$), $\mathbf{e}_t$ is the vector of DSGE exogenous shocks (for example: monetary shock e_t^r, preference shock e_t^c or government shock e_t^g), $\boldsymbol{\eta}_t$ is the vector of DSGE expectational error (by definition is given by: $\boldsymbol{\eta}_t = \tilde{\mathbf{s}}_t - \mathbb{E}_{t-1}[\tilde{\mathbf{s}}_t]$ where $\tilde{\mathbf{s}}_t$ is a subvector of $\mathbf{s}_t$ that contains expectational variables, satisfying $\mathbb{E}_t[\boldsymbol{\eta}_{t+1}] = 0$ for all t), while Γ_0, Γ_1 and Π are matrices of parameters. The **linearized solution** of *Equation (3.1)* delivers a VAR process for DSGE state variables:

$$\underbrace{\mathbf{y}_{nt}}_{(n\times 1)} = \underbrace{\mathbf{D}(\boldsymbol{\vartheta})}_{(n\times r)} \underbrace{\mathbf{s}_t}_{(r\times 1)} \tag{3.2}$$

$$\underbrace{\mathbf{s}_t}_{(r\times 1)} = \underbrace{\mathbf{G}(\boldsymbol{\vartheta})}_{(r\times r)} \underbrace{\mathbf{s}_{t-1}}_{(r\times 1)} + \underbrace{\mathbf{H}(\boldsymbol{\vartheta})}_{(r\times r_e)} \underbrace{\mathbf{e}_t}_{(r_e\times 1)} \qquad \mathbf{e}_t \sim \mathcal{N}(\mathbf{0}; \mathbf{Q}_{\mathbf{e}}(\boldsymbol{\vartheta})) \tag{3.3}$$

where: $\mathbf{y}_{nt}$ denotes the n-dimensional vector of DSGE observed time series, $\mathbf{s}_t$ denotes the r-dimensional vector of DSGE state variables, $\boldsymbol{\vartheta}$ denotes the vector of DSGE *deep parameters* that we wish to estimate, $\mathbf{e}_t$ denotes the r_e-dimensional vector of DSGE exogenous shocks with diagonal variance-covariance matrix $\mathbf{Q}_{\mathbf{e}}(\boldsymbol{\vartheta})$, while $\mathbf{D}(\boldsymbol{\vartheta})$, $\mathbf{G}(\boldsymbol{\vartheta})$ and $\mathbf{D}(\boldsymbol{\vartheta})$ denote matrices of parameters as a function of the deep parameters vector $\boldsymbol{\vartheta}$. As in Kryshko (2009), in order to interpret the r unobserved static factors as r state variables, we assumed that $\mathbf{s}_t$ has the same dimension of $\mathbf{F}_t$.

Handling this system for estimation and/or forecasting might generate the following drawbacks:

1. As pointed out by Ireland (2004), this system is highly stylized, indeed it can not be expected to mimic the data generating process (DGP):

> *...a method for combining the power of DSGE theory with the flexibility of VAR time-series series models, in hopes of obtaining a hybrid that shares the desirable features of both approaches to macroeconomics. The method takes as its starting point a fullyspecified DSGE model, but also admits that while this model may be powerful enough to account for and explain many key features of the US data, it remains too stylized to possibly capture all of the dynamics that can be found in the data. Hence, it augments the DSGE model so that its residuals (meaning the movements in the data that the theory cannot explain) are described by a VAR, making estimation, hypothesis testing, and*

forecasting feasible

Ireland (2004) pag.1206

To overcome this problem Ireland (2004) has proposed to sort out the linearized solution adding a measurement error, which differently from Sargent (1989) and Altug (1989) is allowed to follow an unconstrained, first order vector autoregression with no-diagonal variance-covariance matrix.

2. $\mathbf{y}_{nt}$ contains only few observed variables. Differently from factor models where all N variables in the data-set $\mathbf{x}_{Nt}$ are explored, in DSGE modelling only n obderved variables (with $N >> n$) are accounted for. For example, in their DSGE model Smets and Wouters (2007) consider only $n = 7$ observed variables, while Stock and Watson(2002a) take into account of 215 observed variables in their diffusion index model.

3. As discussed by Boivin and Giannoni (2006), in regular DSGE models it is assumed that each *theoretical concept* (such as, inflation or employment) is properly measured by a single *data indicator* in $\mathbf{x}_{Nt}$ and this choice is quite arbitrary. It means that imperfect information is not allowed, while in realty, institutions, researchers and central banks have different amounts of informations available.

To get away from these limitations the linearized DSGE solution must be augmented, otherwise information is lost and DSGE model misspecification is generated. The most powerful way, in our view, to overcome all these limitations is to combine a DSGE model with a static factor model as proposed by Boivin and Giannoni (2006).

3.2.1 The drDSGE: representation theory

The idea of Data-Rich DSGE (drDSGE) is to extract the common factor vector $\mathbf{F}_t$ from large panel of macroeconomic time series $\mathbf{x}_{Nt}$ and to match the state variable vector $\mathbf{s}_t$ of the model to the extracted common factor $\mathbf{F}_t$ (this matching generates the so-called *Data-Rich Environment*), where the law of common factors $\mathbf{F}_t$ is governed by the DSGE linearized solution. The key assumption of their approach is the separation betweeen *observed* or *data indicators* and *theoretical* or *model concepts*:

- the *data indicators* or simply *indicators* are the observed time series variables in $\mathbf{x}_{Nt}$;

- the *theoretical concepts* are time series variables in the vector $\mathbf{x}_{Nt}$ observed by econometricians or central banks, such as: employment, inflation or productivity shocks, that are assumed to be not properly measured by a single data series, but they are merely imperfect indicators of the observed time series. For example, the employment is imperfectly measured because there are discrepancies between its two main sources: one obtained from the establishment survey and the other from the population survey.

This approach allows: first, to explore a richer amount of information by combining a DSGE model with a static factor model; second, to introduce imperfect information on DSGE estimation which is particular useful to characterize the desirable monetary policy (Boivin and Giannoni(2008)); third, to interpret structurally the latent factors; fourth, to avoid the *Lucas critique*.

Let $\bar{\mathbf{s}}_t = [\mathbf{y}'_{nt} \quad \mathbf{s}'_t]'$ be the vector collecting all variables in a given DSGE model, by definition:

$$\bar{\mathbf{s}}_t \equiv \begin{bmatrix} \mathbf{y}_{nt} \\ \mathbf{s}_t \end{bmatrix} = \begin{bmatrix} \mathbf{D}(\boldsymbol{\vartheta}) \\ \mathbf{I} \end{bmatrix} \mathbf{s}_t \tag{3.4}$$

where the sign $\equiv$ means identity. Representing the vector of common factors $\mathbf{F}_t$ as a subset of the variables in $\bar{\mathbf{s}}_t$, we can define:

$$\mathbf{F}_t \equiv \mathbf{F}\bar{\mathbf{s}}_t = \mathbf{F} \begin{bmatrix} \mathbf{D}(\boldsymbol{\vartheta}) \\ \mathbf{I} \end{bmatrix} \mathbf{s}_t \tag{3.5}$$

where $\mathbf{F}$ is a matrix that generates the common factors $\mathbf{F}_t$ from the vector $\bar{\mathbf{s}}_t$ of all DSGE variables. Now, by substituting *Equation (3.5)* into *Equation (2.14)*, we obtain the **static drDSGE observation equation**:

$$\underbrace{\mathbf{x}_t}_{(N\times 1)} = \underbrace{\boldsymbol{\Lambda}}_{(N\times r)} \underbrace{\mathbf{F}_t}_{(r\times 1)} + \underbrace{\boldsymbol{\xi}_t}_{(N\times 1)} \quad \Rightarrow \quad \underbrace{\mathbf{x}_t}_{(N\times 1)} = \underbrace{\boldsymbol{\Lambda}(\boldsymbol{\vartheta})}_{(N\times r)} \underbrace{\mathbf{s}_t}_{(r\times 1)} + \underbrace{\boldsymbol{\xi}_t}_{(N\times 1)} \tag{3.6}$$

Then, the **drDSGE state space representation** is:

$$\mathbf{x}_t = \boldsymbol{\Lambda}(\boldsymbol{\vartheta})\mathbf{s}_t + \boldsymbol{\xi}_t \tag{3.7}$$

$$\mathbf{s}_t = \mathbf{G}(\boldsymbol{\vartheta})\mathbf{s}_{t-1} + \mathbf{H}(\boldsymbol{\vartheta})\mathbf{e}_t \qquad \mathbf{e}_t \sim \mathcal{N}(\mathbf{0}; \mathbf{Q}_{\mathbf{e}}(\boldsymbol{\vartheta})) \tag{3.8}$$

where $\boldsymbol{\xi}_t$ can be interpreted as serially correlated measurement errors. Adding their law of motion, as we did in *Equation (2.16)*, we obtain the **drDSGE static representation**:

$$\mathbf{x}_t = \boldsymbol{\Lambda}(\boldsymbol{\vartheta})\mathbf{s}_t + \boldsymbol{\xi}_t \tag{3.9}$$

$$\mathbf{s}_t = \mathbf{G}(\boldsymbol{\vartheta})\mathbf{s}_{t-1} + \mathbf{H}(\boldsymbol{\vartheta})\mathbf{e}_t \qquad \mathbf{e}_t \sim \mathcal{N}(\mathbf{0}; \mathbf{Q}_{\mathbf{e}}(\boldsymbol{\vartheta})) \tag{3.10}$$

$$\boldsymbol{\xi}_t = \boldsymbol{\Psi}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t \qquad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}; \mathbf{R}_{\mathbf{v}}) \tag{3.11}$$

where $\boldsymbol{\Lambda}(\boldsymbol{\vartheta})\mathbf{s}_t$ can be interpreted as the static DSGE common component of $\mathbf{x}_t$ since the state variables $\mathbf{s}_t$ are loaded in $\mathbf{x}_t$ just in a contemporaneoulsy way. Because $\mathbf{s}_t$ contains the vector of structural shocks $\mathbf{u}_t$, such as, the technical progress a_t and the vector of errors in data indicators $\boldsymbol{\zeta}_t$, such as, the gdp measurement error ϵ_t^y, we may assume that these shocks have effect in the present and in the past. Then the associated **dynamic drDSGE representation** becomes:

$$\mathbf{x}_t = \underline{\mathbf{B}}(L)\begin{bmatrix}\mathbf{u}_t \\ \boldsymbol{\zeta}_t\end{bmatrix} + \boldsymbol{\xi}_t = \underline{\mathbf{B}}(L)\underline{\mathbf{s}}_t + \boldsymbol{\xi}_t \tag{3.12}$$

$$\boldsymbol{\xi}_t = \boldsymbol{\Psi}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t \qquad \text{where: } \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}; \mathbf{R}_{\mathbf{v}}(\boldsymbol{\vartheta})) \tag{3.13}$$

where $\underline{\mathbf{B}}(L)$ are one-sided filters in the lag operator L as we defined in *Equation (2.2)*, and $\underline{\mathbf{s}}_t = [\mathbf{u}_t \ \boldsymbol{\zeta}_t]'$ can be interpreted as the dynamic (primitive) factors associated to the state variables or static factors $\mathbf{s}_t$. This representation is not used by Boivin and Giannoni (2006) and it remains an open part of the empirical research.

3.2.2 Regular DSGE versus drDSGE

In the drDSGE representation, the key role is played by the matrix $\boldsymbol{\Lambda}(\boldsymbol{\vartheta})$ in *Equation (3.9)*. In a *regular DSGE* model, the number of observed variables n contained in $\mathbf{y}_{nt}$ is usually kept small (most often equal to the number of structural shocks) and theoretical concepts are often assumed to be perfectly measured by a single data indicator in $\mathbf{x}_{Nt}$. So, that there exists a **one-to-one relation** between theoretical concepts and the data indicators. It implies that matrix $\boldsymbol{\Lambda}(\boldsymbol{\vartheta})$ is a $(r \times r)$ identity matrix, where r is the number of state variables.

On the other hand, in a drDSGE model there are **many-to-many relations** between $\mathbf{x}_{Nt}$ and $\mathbf{s}_t$, so that matrix $\boldsymbol{\Lambda}(\boldsymbol{\vartheta})$ becomes $(N \times r)$ with $(N >> r)$. It permits to brige the gap

between data indicators in $\mathbf{x}_{Nt}$ and theoretical concepts in $\bar{\mathbf{s}}_t$. Therefore, the data indicators in $\mathbf{x}_{Nt}$ are partitioned into two groups of variables:

- the *core series* $\mathbf{x}_t^F \in \mathbf{x}_{Nt}$ which correspond to only one model concept in $\bar{\mathbf{s}}_t$;

- the *no-core series* $\mathbf{x}_t^S \in \mathbf{x}_{Nt}$ which are not directly relation with one specific model concept in $\bar{\mathbf{s}}_t$ but are related with more than one model concept.

In other words, the *core series* are time series in $\mathbf{x}_{Nt}$ that cannot be expressed as a linear combination of model concepts $\bar{\mathbf{s}}_t$, while the *no-core series* are time series in $\mathbf{x}_{Nt}$ that can be expressed as a linear combination of more than one model concept in $\bar{\mathbf{s}}_t$. The idea is to separate key DSGE observed variables from no-key DSGE variables. For example, the *core series* might have been various measures of real output (such as: the real GDP or the industrial production), of inflation (such as: the CPI inflation or the PCE deflator inflation) or the nominal interest rate, instead the *no-core* series might include exchange rates, real exports and imports, stock returns and similar data indicators not related directly to any model concept in $\mathbf{x}_{Nt}$.

The drDSGE measurement equation becomes:

$$\underbrace{\begin{bmatrix} \mathbf{x}_t^F \\ - - - \\ \mathbf{x}_t^S \end{bmatrix}}_{\mathbf{x}_t \ (N\times 1)} = \underbrace{\begin{bmatrix} \boldsymbol{\Lambda}(\boldsymbol{\vartheta})^F \\ - - - \\ \boldsymbol{\Lambda}(\boldsymbol{\vartheta})^S \end{bmatrix}}_{\boldsymbol{\Lambda}(\boldsymbol{\vartheta}) \ (N\times r)} \mathbf{s}_t + \underbrace{\begin{bmatrix} \boldsymbol{\xi}_t^F \\ - - - \\ \boldsymbol{\xi}_t^S \end{bmatrix}}_{\boldsymbol{\xi}_t \ (N\times 1)} \tag{3.14}$$

where the matrix $\boldsymbol{\Lambda}(\boldsymbol{\vartheta})^F$ contains just one non-zero element for each row, while the matrix $\boldsymbol{\Lambda}(\boldsymbol{\vartheta})^S$ contains more than one non-zero element for each row and measurement errors $\boldsymbol{\xi}_t$ may be serially correlated, but uncorrelated across different data indicators:

$$\boldsymbol{\xi}_t = \boldsymbol{\Psi}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t \qquad\qquad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}; \mathbf{R}_{\mathbf{v}}) \tag{3.15}$$

indeed, the matrices $\mathbf{\Psi}$ and $\mathbf{R_v}$ are diagonal. For example, the measurement equation assumes the following typical form:

$$\underbrace{\begin{bmatrix} \text{Output Growth \# 1} \\ \text{Output Growth \# 2} \\ \vdots \\ \text{Output Growth \# } \tilde{n}_x \\ \text{Inflation \# 1} \\ \text{Inflation \# 2} \\ \vdots \\ \text{Inflation \# } \tilde{n}_\pi \\ \vdots \\ - - - - - \\ \text{Exchange Rate \# 1} \\ \text{Exchange Rate \# 2} \\ \vdots \\ \text{Exchange Rate \# } \tilde{n}_{er} \\ \vdots \end{bmatrix}}_{\mathbf{x}_t \ (N\times 1)} = \underbrace{\begin{bmatrix} 1 & 0 & \dots & 0 \\ \gamma_{x,2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{x,\tilde{n}} & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & \gamma_{\pi,2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \gamma_{\pi,\tilde{n}} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ - - - & - - - & - - - & - - - \\ \gamma_{1,1} & \gamma_{1,2} & \cdots & \gamma_{1,\tilde{n}} \\ \gamma_{2,1} & \gamma_{2,2} & \cdots & \gamma_{2,\tilde{n}} \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{\tilde{n}_{er},1} & \gamma_{\tilde{n}_{er},2} & \cdots & \gamma_{\tilde{n}_{er},\tilde{n}} \\ \vdots & \vdots & \dots & \vdots \end{bmatrix}}_{\mathbf{\Lambda}(\boldsymbol{\vartheta}) \ (N\times r)} \underbrace{\begin{bmatrix} x_t \\ \pi_t \\ \vdots \end{bmatrix}}_{\mathbf{s}_t \ (r\times 1)} + \underbrace{\begin{bmatrix} \xi^x_{1,t} \\ \xi^x_{2,t} \\ \vdots \\ \xi^x_{\tilde{n},t} \\ \vdots \\ \xi^\pi_{1,t} \\ \xi^\pi_{2,t} \\ \vdots \\ \xi^\pi_{\tilde{n},t} \\ \vdots \\ - - - \\ \xi^{er}_{1,t} \\ \xi^{er}_{2,t} \\ \vdots \\ \xi^{er}_{\tilde{n},t} \\ \vdots \end{bmatrix}}_{\boldsymbol{\xi}_t \ (N\times 1)} \tag{3.16}$$

where we assumed that each model concept can be measured by $\tilde{n}$ time series variables, $\gamma_{i,\tilde{n}}$ denotes a generic parameter of the matrix $\mathbf{\Lambda}(\boldsymbol{\vartheta})$ for the $\tilde{n}$-th time series for the same i-th model concept. As we can see, *core series* admits just one no-null parameter in each row, it means that they are directly related with only one model concept in $\mathbf{s}_t$. Therefore, to tune the magnitude of each model concept, the value of $\gamma_{i;\tilde{n}}$ of just one variable is unity as the first row. Meanwhile, *non-core series* are assumed linear combinations of all DSGE state variables $\mathbf{s}_t$ trought the parameters $[\gamma_{i;1} \ \dots \ \gamma_{i;\tilde{n}}]$. So, the **state space representation** of the drDSGE

is:

$$\underbrace{\begin{bmatrix} \mathbf{x}_t^F \\ - - - \\ \mathbf{x}_t^S \end{bmatrix}}_{\mathbf{x}_t \ (N\times 1)} = \underbrace{\begin{bmatrix} \boldsymbol{\Lambda}(\boldsymbol{\vartheta})^F \\ - - - \\ \boldsymbol{\Lambda}(\boldsymbol{\vartheta})^S \end{bmatrix}}_{\boldsymbol{\Lambda}(\boldsymbol{\vartheta}) \ (N\times r)} \underbrace{\mathbf{s}_t}_{(r\times 1)} + \underbrace{\begin{bmatrix} \boldsymbol{\xi}_t^F \\ - - - \\ \boldsymbol{\xi}_t^S \end{bmatrix}}_{\boldsymbol{\xi}_t \ (N\times 1)} \tag{3.17}$$

$$\underbrace{\mathbf{s}_t}_{(r\times 1)} = \underbrace{\mathbf{G}(\boldsymbol{\vartheta})}_{(r\times r)} \underbrace{\mathbf{s}_{t-1}}_{(r\times 1)} + \underbrace{\mathbf{H}(\boldsymbol{\vartheta})}_{r\times r_{\mathbf{e}})} \underbrace{\mathbf{e}_t}_{(N_{\mathbf{e}} \text{ x } 1)} \qquad \mathbf{e}_t \sim \mathcal{N}(\mathbf{0}; \mathbf{Q}_{\mathbf{e}}(\vartheta)) \tag{3.18}$$

$$\underbrace{\boldsymbol{\xi}_t}_{(N\times 1)} = \underbrace{\boldsymbol{\Psi}}_{(N\times N)} \underbrace{\boldsymbol{\xi}_{t-1}}_{(N\times 1)} + \underbrace{\mathbf{v}_t}_{(N\times 1)} \qquad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}; \mathbf{R}_{\mathbf{v}}) \tag{3.19}$$

where we assumed that the matrices $\mathbf{Q}_{\mathbf{e}}(\vartheta)$, $\mathbf{R}_{\mathbf{v}}$ and $\boldsymbol{\Psi}$ are diagonal. The essential feature of the drDSGE is that the panel dimension of data set N is much higher than the number of DSGE model states r (with: $N >> r$). This representation is quite similar to the static factor model representation (from *Equation (2.14)* to *Equation (2.16)*). The differences are: first, the law of motion of the unobserved factors is now governed by a DSGE model solution; second, the some factor loadings are restricted by the economic meaning of the DSGE model concepts.

3.2.3 The drDSGE estimation step

Following Boivin and Giannoni (2006), the state space representation (from *Equation (3.17)* to *Equation (3.19)*) represents the starting point to estimate the drDSGE. This system is estimated using Bayesian methods under Markov Chain Monte Carlo (MCMC) algorithm. For convenience, we divide parameters of the model into two types: the first type are the *deep parameters* in vector $\boldsymbol{\vartheta}$, and the second type are the parameters collected by the state space representation of the model as $\boldsymbol{\Xi} = \{\boldsymbol{\Lambda}(\boldsymbol{\vartheta}), \boldsymbol{\Psi}, \mathbf{R}_{\mathbf{v}}\}$. We denote with $\mathbf{x}^T = \{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$ and $\mathbf{s}^T = \{\mathbf{s}_1, \ldots, \mathbf{s}_T\}$ the data and the states up to time $t = T$ respectively. Because of the normality of the structural shocks $\mathbf{e}_t$ and the measurement error innovations $\mathbf{v}_t$, the system from *Equation (3.17)* to *Equation (3.19)* is a **linear Gaussian state space model** and the likelihood function of data $p(\mathbf{x}^T|\boldsymbol{\vartheta}, \boldsymbol{\Xi})$ can be evaluated using the Kalman filter.

Generally speaking a **Bayesian estimation** for an unknown parameter vector $\boldsymbol{\vartheta}$ is implemented based on following procedure:

step 1a: Set the prior distribution $p(\boldsymbol{\vartheta})$, which is the distribution of $\boldsymbol{\vartheta}$ that the researcher

have in mind before observing the data.

step 2a: Convert the prior distribution to the posterior distribution $p(\boldsymbol{\vartheta}|\mathbf{x}^T)$, which is the distribution of $\boldsymbol{\vartheta}$ conditional on the data $\mathbf{x}^T$, using the Bayes theorem:

$$p(\boldsymbol{\vartheta}|\mathbf{x}^T) = \frac{p(\mathbf{x}^T|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta})}{\int p(\mathbf{x}^T|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta})\ d\boldsymbol{\vartheta}} \tag{3.20}$$

where $p(\mathbf{x}^T|\boldsymbol{\vartheta})$ denotes the likelihood function of the data given the *deep parameters* vector .

In our case, the aim is to estimate the couple $(\boldsymbol{\vartheta}, \boldsymbol{\Xi})$, rather than just one single unknown vector. The posterior distribution of the couple is:

$$p(\boldsymbol{\vartheta}, \boldsymbol{\Xi}|\mathbf{x}^T) = \frac{p(\mathbf{x}^T|\boldsymbol{\vartheta}, \boldsymbol{\Xi})p(\boldsymbol{\vartheta}, \boldsymbol{\Xi})}{\int p(\mathbf{x}^T|\boldsymbol{\vartheta}, \boldsymbol{\Xi})p(\boldsymbol{\vartheta}, \boldsymbol{\Xi})\ d\boldsymbol{\vartheta}\ d\boldsymbol{\Xi}} \tag{3.21}$$

where $p(\boldsymbol{\vartheta}, \boldsymbol{\Xi})$ denotes its prior distribution, while $p(\boldsymbol{\vartheta}, \boldsymbol{\Xi}|\mathbf{x}^T)$ denotes its likelihood function. In order to generate draws from the posterior distribution $p(\boldsymbol{\vartheta}, \boldsymbol{\Xi}|\mathbf{x}^T)$, since it is not directly tractable, we divide it into the following four conditional posterior distributions:

$$p(\boldsymbol{\Xi}|\boldsymbol{\vartheta}, \mathbf{x}^T) \quad p(\mathbf{s}^T|\boldsymbol{\Xi}, \boldsymbol{\vartheta}; \mathbf{x}^T) \quad p(\boldsymbol{\Xi}|\mathbf{s}^T, \boldsymbol{\vartheta}; \mathbf{x}^T) \quad p(\boldsymbol{\vartheta}|\boldsymbol{\Xi}, \mathbf{x}^T) \tag{3.22}$$

and we adopt the **Metropolis-within-Gibbs** algorithm, wherethe Gibbs sampler generats draws from joint posterior distribution $p(\boldsymbol{\vartheta}, \boldsymbol{\Xi}|\mathbf{x}^T)$ by repeating iteratively generation of draws from conditional posterior distributions $p(\boldsymbol{\Xi}|\boldsymbol{\vartheta}, \mathbf{x}^T)$ and $p(\boldsymbol{\vartheta}|\boldsymbol{\Xi}, \mathbf{x}^T)$.

To be precise, the main steps of **Metropolis-within-Gibbs** algorithm used in drDSGE estimation are:

step 1b: Specify initial values of parameters $\boldsymbol{\vartheta}^{(0)}$ and $\boldsymbol{\Xi}^{(0)}$. And set the iteration index g at $g = 1$.

step 2b: Solve the DSGE model numerically at $\boldsymbol{\vartheta}^{(g-1)}$ based on Sims (2002)' method and obtain $\mathbf{G}(\boldsymbol{\vartheta}^{(g;1)})$, $\mathbf{H}(\boldsymbol{\vartheta})$, and $\mathbf{Q}(\boldsymbol{\vartheta})$ in *Equation (3.18)*.

step 3b: Draw $\boldsymbol{\Xi}^{(g)}$ from $p(\boldsymbol{\Xi}|\boldsymbol{\vartheta}^{(g-1)}, \mathbf{x}^T)$.

(3.1b) Generate unobserved state variables $\mathbf{s}_t^{(g)}$ from $p(\mathbf{s}^T|\boldsymbol{\Xi}^{(g-1)}, \boldsymbol{\vartheta}, \mathbf{x}^T)$ using simulation

smoother by DeJong and Shephard (1995).

(3.2b) Generate parameters $\boldsymbol{\Xi}^{(g)}$ from $p(\boldsymbol{\Xi}|\mathbf{s}^T, \boldsymbol{\vartheta}, \mathbf{x}^T)$, using the sampled draw $\mathbf{s}^{T(g)}$.

step 4b: Draw deep parameters $\boldsymbol{\vartheta}^{(g)}$ from $p(\boldsymbol{\vartheta}|\boldsymbol{\Xi}^{(g)}, \mathbf{x}^T)$ using Metropolis step:

(4.1b) Sample from proposal density $p(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^{(g-1)})$ and, using the sampled draw $p(\boldsymbol{\vartheta}^{(proposal)})$, calculate the acceptance probability ap as follows:

$$ap = \left[\frac{p(\boldsymbol{\vartheta}^{(proposal)}|\boldsymbol{\Xi}^{(g)}, \mathbf{x}^T)\ p(\boldsymbol{\vartheta}^{(g-1)}|\boldsymbol{\vartheta}^{(proposal)})}{p(\boldsymbol{\vartheta}^{(g-1)}|\boldsymbol{\Xi}^{(g)}, \mathbf{x}^T)\ p(\boldsymbol{\vartheta}^{(proposal)}|\boldsymbol{\vartheta}^{(g-1)})}\ ;\ 1 \right]$$

(4.2b) Accept $\boldsymbol{\vartheta}^{(proposal)}$ with probability ap and reject it with probability $1 - ap$. Set $\boldsymbol{\vartheta}^{(g)} = \boldsymbol{\vartheta}^{(proposal)}$ when accepted and $\boldsymbol{\vartheta}^{(g)} = \boldsymbol{\vartheta}^{(g-1)}$ when rejected.

step 5b: Set the iteration index $g = g + 1$, return to Step 2 up to $g = G$ where G is the number of MCMC iterations.

Step 4b of this algorithm plays an essential role. Infact, it is important to make the acceptance probability ap as close to one as possible especially around the mode of the posterior density $p(\boldsymbol{\vartheta}|\boldsymbol{\Xi}, \mathbf{x}^T)$ because the same values are sampled consecutively if ap is low. To achieve this purpose, we should choose the proposal density $p(\boldsymbol{\vartheta}^{(proposal)}|\boldsymbol{\vartheta}^{(g-1)})$ that mimics the posterior density $p(\boldsymbol{\vartheta}|\boldsymbol{\Xi}, \mathbf{x}^T)$ especially around its mode. This is why we firstly run *regular DSGE* model estimation and compute the posterior mode of the DSGE model parameters to obtain initial value $\boldsymbol{\vartheta}^{(0)}$ of Step 1. Then, we generate smoothed state variables $\mathbf{s}_t^{(0)}$ using $\boldsymbol{\vartheta}^{(0)}$ and obtain initial value $\boldsymbol{\Xi}_t^{(0)}$ from OLS regressions of $\mathbf{x}_t$ on $\mathbf{s}_t^{(0)}$. The previous literature sugggest to use the so-called *random-walk MH algorithm* (see An and Schorfheide (2007)) as Metropolis step in Step 4b, where the proposal density $\boldsymbol{\vartheta}^{(proposal)}$ is sampled from the random-walk model:

$$\boldsymbol{\vartheta}^{(proposal)} = \boldsymbol{\vartheta}^{(g-1)} + \boldsymbol{\tau}_t \qquad\qquad \boldsymbol{\tau}_t \ \sim i.i.d\ \mathcal{N}(\mathbf{0}; c\mathsf{H})$$

where H is the Hessian matrix of the logarithm of the posterior distribution, indeed, $-l_p''^{-1}(\hat{\boldsymbol{\vartheta}})$ where $l_p(\boldsymbol{\vartheta}) = \ln(p(\boldsymbol{\vartheta}|\boldsymbol{\Xi}, \mathbf{x}^T))$, while c is a scalar called the *adjustment coefficient*, whose choice will be explained below.

The merit of using this *random-walk proposal* is that $p(\boldsymbol{\vartheta}^{(g-1)}|\boldsymbol{\vartheta}^{(proposal)}) = p(\boldsymbol{\vartheta}^{(proposal)}|\boldsymbol{\vartheta}^{(g-1)})$,

so that the acceptance probability ap collapses to:

$$ap = \min \left[\frac{f(\boldsymbol{\vartheta}^{(proposal)}}{f(\boldsymbol{\vartheta}^{(g-1)})} \; ; \; 1 \right]$$

which does not depend on the proposal density $p(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^{(g-1)})$. We must, however, be careful for $p(\boldsymbol{\vartheta}^{(proposal)})$ not to deviate from $p(\boldsymbol{\vartheta}^{(g-1)}$ so much because the acceptance probability ap may be low when those deviate far from each other. This may be achieved by making c low, but $p(\boldsymbol{\vartheta}^{(proposal)}$ may be sampled only from the narrow range if c is too low. In random walk sampler, the optimal acceptance rate ap according to Roberts et al. (1997) and Neal and Roberts (2008) is around 25%, ranging from 0.23 for large dimensions to 0.45 for univariate case. Following the previous literature, we simply use this random-walk MH algorithm with $\mathsf{H} = -l_p''^{-1}(\hat{\boldsymbol{\vartheta}})$.

For the prior densities, we follow the general approach used for DSGE modelling. We assume that the exogenous shocks $\mathbf{e}_t$ such as technology shock, preference shocks or monetary shocks are persistent for their past shocks and their law of motions follow an AR(1) process, such that: $u_t = \rho u_{t-1} + \varsigma_t$ where the error term ς_t is *i.i.d.* Since the coefficient ρ must be between zero and one to satisfy the stationary property, their prior densities are assumed to follow *beta distributions*, while the variances of the error term ς_t are setted up on *inverted gamma distributions*. For the other parameters we assumed *normal distributions*.

3.3 The DSGE model of Smets and Wouters (2007)

This section presents the DSGE model of Smets and Wouters (2007) providing the log-linearized equilibrium conditions *(SubSection 3.3.1)*, and shows how it can be casted in the Data-Rich Environment *(SubSection (3.3.2))* as described in the previous section.

The DSGE model of Smets and Wouters (2007) is a medium-scale New Keynesian model with price and wage rigidities, capital accumulation, investment adjustment cost, and habit formation. Although their model represents the workhorse of the applied DSGE research, it has the following main drawbacks: first, it assumes few observable variables (7 observed variables); second, it assumes that each theoretical concept is measured correctly by a single

arbitrary time series, third, it assumes homogeneous agents; forth, assumes a closed economy; fifth, there is not financial intermediation (indeed the capital is directly rent by the household). Here, we focus on a way to get away from the first two drawbacks.

3.3.1 The equilibrium conditions

The DSGE model of Smets and Wouters (2007) is built up using the following equilibrium conditions:

(1) The resource contraint:

$$\hat{y}_t = c_y \hat{c}_t + i_y \hat{i}_t + z_y \hat{z}_t + u_t^g \tag{3.23}$$

where $\hat{y}_t$ denotes the output, $\hat{c}_t$ denotes the consumption, $\hat{i}_t$ denotes the investment, $\hat{z}_t$ denotes the capital utilization rate, u_t^g denotes the exogenous government spending shock, $c_y = 1 - g_y - i_y$ denotes the steady state share of consumption (scalar), g_y denotes the steady state exogenous spending/output ratio (scalar), and i_y denotes the steady state investment/output ratio (scalar). This equilibrium condition states simply that output must be asborbed by consumption, investment, capital utilization costs and exogenous government spending. Differently from Smets and Wouters(2007) we assume that exogenous spending follows just a first-order autoregressive process with an $i.i.d$-Normal error term without be affected by the productivity shock as follows: $u_t^g = \rho_g u_{t-1} + e_t^g$

(2) Consumption Euler Equation:

$$\hat{c}_t = c_1 \hat{c}_{t-1} + (1 - c_1)\mathbb{E}_t[\hat{c}_{t+1}] + c_2(\hat{l}_t - \mathbb{E}_t[\hat{l}_{t+1}]) - c_3(\hat{r}_t - \mathbb{E}_t[\hat{\pi}_{t+1}] + u_t^c) \tag{3.24}$$

where $c_1 = \frac{h}{1+h}$ where $h = \frac{\lambda}{\gamma}$, $c_2 = \frac{(\sigma_c - 1)(W_*^h L_*/C_*)}{(1+h)}$, $c_3 = \frac{1-h}{(1+h)\sigma_c}$ are scalars, h is a measure of the habit persistence that depends positively on λ which is the strength of the first-order external habit formation and negatively on γ which is the steady state growth rate of the economy while σ_c is the inverse of the intertemporal elasticity of substitution. They assume that the external habit stock $\mathtt{H}_t$ is proportional to aggregate past consumption: $\mathtt{H}_t = h\ c_{t-1}$. When $h = 0$, this equation reduces to the traditional forward-looking consumption equation. As in Smets and Wouters (2007), the exogenous

shock u_t^c is assumed to be a AR(1) process with coefficient ρ_c, indeed: $u_t^c = \rho_c u_{t-1}^c + e_t^c$.

(3) Investment Euler Equation:

$$\hat{i}_t = i_1 \hat{i}_{t-1} + (1 - i_1)\mathbb{E}_t[\hat{i}_{t+1}] + i_2 \hat{q}_t + u_t^i \tag{3.25}$$

where $i_1 = \frac{1}{1+\beta\gamma^{(1-\sigma_c)}}$ and $i_2 = \frac{1}{1+\beta\gamma^{(1-\sigma_c)}\gamma^2\varphi}$ are scalars, φ is the steady-state elasticity of the capital adjustment cost function and β is the discount factor applied by households, u_t^i is the investment shock, $\hat{q}_t$ is the real value of the existing capital stock. As in Smets and Wouters (2007), the exogenous shock u_t^i is assumed to be a AR(1) process with coefficient ρ_i, indeed: $u_t^i = \rho_i u_{t-1}^i + e_t^i$.

(4) Value of Capital Euler Equation :

$$\hat{q}_t = q_1 \mathbb{E}_t[\hat{q}_{t+1}] + (1 - q_1)\mathbb{E}_t[\hat{r}_{t+1}^k] - (\hat{r}_t - \hat{\pi}_{t+1} + u_t^c) \tag{3.26}$$

where $q_1 = \beta\gamma^{\sigma_c}(1 - \delta)$ is a scalar where δ is the appreciation rate for capital. The equation states that the current value of the capital stock (or the shadow price of capital or Tobin's q) $\hat{q}_t$ depends positively on its expected future value and the expected real rental rate on capital $\mathbb{E}_t[\hat{r}_{t+1}^k]$ and negatively on the previous real interest rate and the risk premium disturbance.

(5) Aggregate production function

$$\hat{y}_t = \phi_p(\alpha \hat{k}_t^s + (1 - \alpha)\hat{l}_t + u_t^a) \tag{3.27}$$

where ϕ_p is one plus the share of fixed costs in production, reflecting the presence of fixed costs in production, $\hat{k}_t^s$ is the output produced using capital, $\hat{l}_t$ are the hours worked, and u_t^a is the productivity shock. As in Smets and Wouters (2007), we assume that the productivity shock admits a AR(1) process: $u_t^a = \rho_a u_{t-1}^a + e_t^a$.

(6) Capital used in production

$$\hat{k}_t^s = \hat{k}_{t-1} + \hat{z}_t \tag{3.28}$$

where $\hat{z}_t$ is the capital utilization rate and $\hat{k}_{t-1}$ is the capital used the previous period. The equation states that newly installed capital becomes only effective with a one lag.

(7) Degree capital utilization

$$\hat{z}_t = z_1 \hat{r}_t^k \tag{3.29}$$

where $z_1 = \frac{1-\psi}{\psi}$ where ψ is a parameter that takes values between 0 and 1 and measures the elasticity of the capital utilization adjustment cost. When $\psi = 1$, it is extremely costly to change the utilisation of capital and as a result the utilisation of capital remains constant. In contrast, when $\psi = 0$, the marginal cost of changing the utilisation of capital is constant and as a result in equilibrium the rental rate on capital is constant.

(8) Installed Capital

$$\hat{k}_t = k_1 \hat{k}_{t-1} + (1 - k_1)\hat{i}_t + k_2 u_t^i \tag{3.30}$$

where $k_1 = (1-\delta)/\gamma$ and $k_2 = (1-(1-\delta)/\gamma)(1-\beta\gamma^{(1-\sigma_c)})\gamma^2\psi$ are scalars. The equilibrium equation states that the accumulation of installed capital $\hat{k}_t$ is not only a function of the flow of investment but also of the relative efficiency of these investment expenditures as captured by the investment disturbance u_t^i.

(9) Mark-up equation:

$$\hat{\mu}_t^p = \widehat{mpl}_t - \hat{w}_t = \alpha(\hat{k}_t^s - \hat{l}_t) - \hat{w}_t + u_t^a \tag{3.31}$$

where $\hat{\mu}_t^p$ denotes the price mark-up by firms, $\widehat{mpl}_t$ denotes the marginal product of labour, $\hat{w}_t$ denotes the real wage, u_t^a denotes the productivity shock, $\hat{l}_t$ denotes the hours worked, and $\hat{k}_t^s$ denotes the capital installed. Being the Smets and Wouters (2007) model constructed under monopolist competition, this equilibrium equation states simply that cost minimisation by firms implies the price mark-up $\hat{\mu}_t^p$ over the marginal cost.

(10) Inflation:

$$\hat{\pi}_t = \pi_1 \hat{\pi}_{t-1} + \pi_2 \mathbb{E}_t[\hat{\pi}_{t+1}] - \pi_3 \hat{\mu}_t^p + u_t^\pi \tag{3.32}$$

where $\pi_1 = \frac{\iota_p}{1+\beta\gamma^{1-\sigma_c}\iota_p}$, $\pi_2 = \frac{\beta\gamma^{1-\sigma_c}}{1+\beta\gamma^{1-\sigma_c}\iota_p}$ and $\pi_3 = \frac{1}{1+\beta\gamma^{1-\sigma_c}\iota_p}\ \frac{(1-\beta\gamma^{1-\sigma_c}\xi_p)(1-\xi_p)}{\xi_p((\phi_p-1)\xi_p+1)}$ are scalars where ξ_p is the degree of price stickiness. The equation represents the New Keynesian Phillips Curve with price adjustment a la Calvo. It states that inflation $\hat{\pi}_t$ depends positively on past and expected future inflation, negatively on the current price mark-up $\hat{\mu}_t^p$ and positively on a inflation disturbance (or as called by Smets and Wouters (2007) price mark-up disturbance) u_t^π. Differently from Smets and Wouters (2007), we

assume simply that the price mark-up disturbance is assumed to follow AR(1) process: $u_t^\pi = \rho_\pi u_t^\pi + e_t^\pi$, where e_t^π is an $i.i.d$-Normal price mark-up shock. When the degree of indexation to past inflation is zero $\iota_p = 0$, this equation reverts to a standard purely forward-looking Phillips curve $\pi_1 = 0$.

(10) Rental Rate of Capital:

$$\hat{r}_t^k = \hat{w}_t - (\hat{k}_t - \hat{l}_t) \tag{3.33}$$

this equation states that the rental rate of capital $\hat{r}_t^k$ is negatively related to the capital-labour ratio $\hat{k}_t - \hat{l}_t$ and positively to the real wage $\hat{w}_t$.

(12) Wage mark-up:

$$\hat{\mu}_t^w = \hat{w}_t - \widehat{mrs}_t = \hat{w}_t - (\sigma_l \hat{l}_t + \frac{1}{1-\lambda}(\hat{c}_t - \lambda \hat{c}_{t-1})) \tag{3.34}$$

where $\widehat{mrs}_t$ is the current marginal rate of substitution, σ_l is the elasticity of labour supply with respect to the real wage. Being the model constructed under monopolistic competition, this equation states simply that the wage mark-up will be equal to the difference between the real wage and the marginal rate of substitution between working and consuming.

(13) Real Wage:

$$\hat{w}_t = w_1 \hat{w}_{t-1} + (1 - w_1)(\mathbb{E}[\hat{w}_{t-1}] + \mathbb{E}[\hat{\pi}_{t+1}]) - w_2 \hat{\pi}_t + w_3 \hat{\pi}_{t-1} - w_4 \mu_t^w + u_t^w \tag{3.35}$$

where $w_1 = \frac{1}{1+\beta\gamma 1-\sigma_c}$, $w_2 = \frac{1+\beta\gamma 1-\sigma_c\iota_w}{1+\beta\gamma 1-\sigma_c}$, $w_4 = \frac{1}{1+\beta\gamma 1-\sigma_c}\ \frac{(1-\beta\gamma 1-\sigma_c\xi_w)(1-\xi_w)}{\xi_w((\phi_w-1)\xi_w+1)}$ and $w_3 = \frac{\iota_w}{1+\beta\gamma 1-\sigma_c}$ are scalars where ξ_w and ι_w are defined analogously to their counterparts in the price setting conditions. This equation states that the real wage $\hat{w}_t$ is a function of expected and past real wages, expected current and past inflation, the wage mark-up μ_t^w and a wage-markup disturbance u_t^w. Differently from Smets and Wouters (2007), we assumed that the wage-markup disturbance follows: $u_t^w = \rho_w u_t^w + e_t^w$ where e_t^w is an $i.i.d$-Normal term.

(14) Monetary Policy Reaction function:

$$\hat{r}_t = \rho \hat{r}_{t-1} + (1-\rho)\{r_\pi \hat{\pi}_t + r_y(\hat{y}_t - \hat{y}_t^p)\} + r_{\Delta y}|(\hat{y}_t - \hat{y}_t^p) - (\hat{y}_{t-1} - \hat{y}_{t-1}^p)| + u_t^r \tag{3.36}$$

where $\hat{r}_t$ is the interest rate and u_t^r is the monetary policy shock. The equation state that the monetary authorities follow a generalised Taylor rule by gradually adjusting the policy-controlled interest rate $\hat{r}_t$ in response to inflation and the output gap, defined as the difference between actual $\hat{y}_t$ and potential output $\hat{y}_t^p$. As in Smets and Wouters (2007), we assume that the monetary policy shock follows a first-order autoregressive process with an *i.i.d*-Normal error term: $u_t^r = \rho_r u_{t-1}^r + e_t^r$.

These equilibrium conditions determine:

☞ 7 exogenous shocks or 7 exogenous variables;

(15) Consumption Shock: $u_t^c = \rho_c u_{t-1}^c + e_t^c$

(16) Investment Shock: $u_t^i = \rho_i u_{t-1}^i + e_t^i$

(17) Productivity Shock: $u_t^a = \rho_a u_{t-1}^a + e_t^a$

(18) Govenment Shock: $u_t^g = \rho_g u_{t-1}^g + e_t^g$

(19) Mark-up Shock (or Inflation Shock): $u_t^\pi = \rho_\pi u_{t-1}^\pi + e_t^\pi$

(20) Wage Shock: $u_t^w = \rho_w u_{t-1}^w + e_t^w$

(21) Monetary Policy Shock: $u_t^r = \rho_r u_{t-1}^r + e_t^r$

☞ 7 forecasts errors;

(22) Consumption Forecast Error: $\hat{c}_t = \mathbb{E}_{t-1}[\hat{c}_t] + \eta_t^c$

(23) Investment Forecast Error: $\hat{i}_t = \mathbb{E}_{t-1}[\hat{i}_t] + \eta_t^i$

(24) Value of Capital Forecast Error: $\hat{q}_t = \mathbb{E}_{t-1}[\hat{q}_{t-1}] + \eta_t^q$

(25) Wage Forecast Error: $\hat{w}_t = \mathbb{E}_{t-1}[\hat{w}_{t-1}] + \eta_t^w$

(26) Mark-up Forecast Error (or Inflation Forecast Error): $\hat{\pi}_t = \mathbb{E}_{t-1}[\hat{\pi}_{t-1}] + \eta_t^\pi$

(27) Cost of Capital Forecast Error: $\hat{r}_t^k = \mathbb{E}_{t-1}[\hat{r}_{t-1}^k] + \eta_t^{r^k}$

(28) Labour demand Forecast Error: $\hat{l}_t = \mathbb{E}_{t-1}[\hat{l}_{t-1}] + \eta_t^l$

☞ 7 observed variables: output $\hat{y}_t$, consumption $\hat{c}_t$, investment $\hat{i}_t$, hours worked $\hat{l}_t$, inflation $\hat{\pi}_t$, real wage $\hat{w}_t$ and the real interest rate $\hat{r}_t$. Briefly: $\mathbf{y}_t = [\hat{y}_t \ \hat{c}_t \ \dots \ \hat{r}_t]'$

☞ 40 all endogenous variables:

These conditions from **(1)** to **(28)** form a system of rational linear expectational difference equations, that can be jointly solved using the Sims (2002) method. The *Equation (3.1)*

becomes:

$$\Gamma_0 \underbrace{\begin{bmatrix} \hat{y}_t \\ \hat{\pi}_t \\ \hat{w}_t \\ \hat{k}_t \\ \hat{q}_t \\ \hat{i}_t \\ \hat{c}_t \\ \hat{r}_t \\ \hat{r}_t^k \\ \hat{l}_t \\ \vdots \\ \mathbb{E}[\pi_{t+1}] \\ \vdots \\ u_t^c \\ u_t^i \\ \vdots \\ e_t^a \\ e_t^g \\ \vdots \end{bmatrix}}_{\mathbf{s}_t \ (40 \ x \ 1)} = \Gamma_1 \underbrace{\begin{bmatrix} \hat{y}_{t-1} \\ \hat{\pi}_{t-1} \\ \hat{w}_{t-1} \\ \hat{k}_{t-1} \\ \hat{q}_{t-1} \\ \hat{i}_{t-1} \\ \hat{c}_{t-1} \\ \hat{r}_{t-1} \\ \hat{r}_{t-1}^k \\ \hat{l}_{t-1} \\ \vdots \\ \mathbb{E}[\pi_t] \\ \vdots \\ u_{t-1}^c \\ u_{t-1}^i \\ \vdots \\ e_{t-1}^a \\ e_{t-1}^g \\ \vdots \end{bmatrix}}_{\mathbf{s}_{t-1} \ (40 \ x \ 1)} + \Psi \underbrace{\begin{bmatrix} e_t^c \\ e_t^i \\ e_t^w \\ e_t^a \\ e_t^\pi \\ e_t^g \\ e_t^r \end{bmatrix}}_{\mathbf{e}_t \ (7 \ x \ 1)} + \Pi \underbrace{\begin{bmatrix} \eta_t^\pi \\ \eta_t^w \\ \eta_t^q \\ \eta_t^i \\ \eta_t^l \\ \eta_t^c \\ \eta_t^{r^k} \end{bmatrix}}_{\eta_t \ (7 \ x \ 1)} \tag{3.37}$$

where the coefficient matrices Γ_0, Γ_1, Ψ and Π are opportunely setted. The vector of deep parameters is:

$$\boldsymbol{\vartheta} = [\varphi \ \sigma_c \ h \ \xi_w \ \sigma_L \ \xi_p \ \iota_w \ \iota_p \ \psi \ \phi \ r_\pi \ \rho \ r_y \ r_{\Delta y} \ \bar{\pi} \ \beta \ \bar{l} \ \bar{\gamma} \ \alpha]'$$

it has been estimated using Bayesian estimation under the Metropolis-Hastings algorithm (**Step 1a** and **Step 2a**). The forecasts generated from the DSGE model are h-step ahead forecasts of quarter-to-quarter output growth and inflation.

3.3.2 The data-rich form

The drDSGE observation equation *(Equation (3.16))*, can be obtained just by adding observable time series variables to the vector $\mathbf{y}_{nt}$ as *core series* and/or *no-core series*. In this

dissertaion, we used just **Case C** of Boivin and Giannoni (2006), where 21 time series are added as *core series* and 7 are added as *no-core series*.

The drDSGE has been estimated starting from the state space representation (*Equation (3.17)* to *Equation (3.19)*) using the Metropolis-within-Gibbs algorithm (from **Step 1b** to **5b**). As the regular DSGE, the forecasts generated from the drDSGE model are h-step ahead forecasts of quarter-to-quarter output growth and inflation.

Chapter 4

The forecasting results

"The best thing about being a statistician is that you get to play in everyone's backyard"

John Tukey

This chapter presents the out-of-sample forecasting experiments of the models discussed in the previous chapters using real data on 83 quarterly U.S time series variables similar to those variables used by Stock and Watson (2002a) and reported in Appendix A. We focus on two key U.S macroeconomic time series variables: the output growth and the inflation. The empirical study has shown three main conclusions: first, DSGE models are outperformed significantly by the GDFM in forecasting the output growth in both short and long run, while the static factor model outperforms significantly DSGE models only in the short run. Second, the most surprising result of the thesis, we discovered that the drDSGE outperforms significantly all other competitive models in forecasting inflation in the long run, while the regular DSGE does not. Therefore, the drDSGE outperforms significantly the regular DSGE in forecasting both output growth and inflation. Third, in most cases, the unrestricted VAR represents the worse forecasting model. The implications of these results are discussed troughtout the chapter.

The chapter is organized as follows. In *Section (4.1)* we open presenting the forecasting experiments and the metric used. In *Section (4.2)* we discuss the forecasting models in competition. In *Section (4.3)* we describe the tests of equal predictive ability, while *Section (4.4)* concludes providing the final results.

4.1 The forecasting experiments

The out-of-sample forecasting experiments are organized as follows. As described in the introduction, we use rolling regressions with sample size fixed at $R = 80$ to forecast up to $h = 12$

quarters ahead the output growth and inflation. The models in competition are: the unconditional mean of the time series, the autoregressive process (hereafter AR(p)), the vector autoregressive process (hearafter VAR(p)), the static factor model or diffusion index (hereafter DI(r)) model of Stock and Watson (2002a), the generalized dynamic factor model (hereafer GDFM(r,q)) of Forni et al. (2000) and Forni et al. (2005), the regular DSGE of Smets and Wouters (2007) and its Data-Rich Environment form following Boivin and Giannoni (2006). The orders p, r and q has been estimated using different ways. The autoregressive order p has been estimated using the Bayesian Information Criterion (BIC), the number of static factor model r has been estimated using the procedure of Alessi et al. (2007), while the number of dynamic factors q has been estimated using the procedure of Hallin and Liška (2007). The first estimation sample starts from 1959:1 and ends in 1978:4 so that the first forecasting date is 1979:1. Earlier observations are used to compute the initial growth rates. After all models have been estimated, the first set of out-of-sample forecasts is computed. Then, sample range shifts one-step forward to 1959:2-1979:1 in order to compute the second set of forecasts. All models are fully re-estimated for each rolling sample with estimation procedures described in the previous chapter. The estimation is performed $S = 96$ times to obtain a series of forecasts for each forecast horizon and each model. The last sample is 1973:1-2001:4 and the last forecasting date is 2004:4.

The metric used to evalute the relative forecasting performance is the relative mean squared forecast error ($rMSFE$), defined as:

$$rMSFE(m,n)_{|h} = 1 - \frac{MSFE_{m|h}}{MSFE_{n|h}} \tag{4.1}$$

where $MSFE_{m|h}$ and $MSFE_{n|h}$ denote respectively the mean squared forecast error generated from model m at the forecasting horizion h and the mean squared forecast error generated from model n at the forecasting horizon h. The metric interpretation is: fixed h, if $rMSFE(m,n)_{|h} > 0$ means a forecasting gain of model m relatively to model n, simmetrically if $rMSFE(m,n)_{|h} < 0$ means a forecasting loss of the model m relatively to model n. The $MSFEs$ have been constructed in the following way. Let $\mathbf{x}_{NT}$ be the finite dataset of N stationary time series up to time T used in the empirical out-of-sample forecasting experiment where $T = R + s - 1$ is the end of each rolling sample s of size $R = 80$. Let y_t be the time

series variable in $\mathbf{x}_{NT}$ that we wish to forecast h steps ahead, and let S be the number of replications. The mean square forecasting error ($MSFE$) of stochastic process y_t repect to the i-th forecasting model has been worked out as:

$$MSFE_{i|h} = S^{-1} \sum_{i=1}^{S} (y_{T+h} - \hat{y}^{i,s}_{T+h|T})^2 \tag{4.2}$$

where y_{T+h} denotes the observed stochastic process y_t at time $T+h$, and $\hat{y}^{i,s}_{T+h|T}$ denotes its unknown point forecast predictor using the i-th forecasting model for the s-th rolling sample.

4.2 Forecasting models

The point forecast predictor $\hat{y}^{i,s}_{T+h|T}$ in *Equation (4.2)* has been worked out using the following forecasting models: the unconditional mean, the autoregressive model, the vector autoregressive model, the static factor model, the generalized dynamic factor model, the DSGE of Smets and Wouters (2007) and its representation in term of Data-Rich Environment.
The use of the unconditional mean of the series of interest as forecasting model is quite straightforward. In this case the point forecast predictor coincides with the unconditional mean of the series for any forecasting horizon. For this reason, we prefer to skip this case considering the other forecasting models directly.

4.2.1 Forecasting with the AR model

Let y_t be our observed stationary time series at time t. The most simple way to forecast a time series is assuming that it follows an autoregressive process. If y_t admits an **autoregressive process** of order p (hereafter AR(p)), we have:

$$y_T = \underline{\alpha} + \delta(L) y_T + \epsilon_T \tag{4.3}$$

where y_T denotes the time series of interest at the end of the estimation sample, $\underline{\alpha}$ denotes the constant, $\delta(L) = 1 - \delta_1 L - \ldots - \delta_p L^p$ denotes the autoregressive polynominal of order p in the lag operator L fixed using the Bayesian Information Criterion (BIC) that loads the past history of y_T, while ε_T is the stochastic error term. The **autoregressive forecasting model** becomes:

$$y^{AR}_{T+h} = \underline{\alpha} + \delta_h(L) y_T + \epsilon_{T+h} \qquad h = 1, \ldots, 12 \tag{4.4}$$

where $\delta_h(L) = 1 - \delta_1 L^{-h} - \ldots - \delta_p L^{p-h}$ denotes the autoregressive polynominal $\delta(L)$ shifted h-steps ahead, while ε_{T+h} denotes the stochastic error term shifted h-steps ahead. The AR forecasts have been generated by estimating the previous equation by OLS for each forecasting horizon. What we get is:

$$\hat{y}^{AR}_{T+h|T} = \underline{\hat{\alpha}} + \hat{\delta}_h(L) y_T \qquad\qquad h = 1, \ldots, 12$$

where $\hat{y}^{AR}_{T+h|T}$ is the desidered point forecast predictor used in *Equation (4.2)*

4.2.2 Forecasting with the VAR model

Let $\mathbf{y}_{nt}$ be the n-dimensional vector of observed stationary time series variables. If $\mathbf{y}_{nt}$ admits a **vector autoregressive process** of order p (hereafter VAR(p)), we have:

$$\mathbf{y}_{nT} = \sum_{j=1}^{p} \mathbf{A}_j \mathbf{y}_{n,T-j} + \boldsymbol{\epsilon}_T \qquad\qquad \boldsymbol{\epsilon}_T \sim iid\, \mathcal{N}(\mathbf{0}; \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}) \tag{4.5}$$

where y_{nT} is our observed time series varaibles at the end of the estimation sample, $\mathbf{A}_j$ are $(n \times n)$ matrices of parameters and $\boldsymbol{\epsilon}_T$ is the n-dimensional white noise process at the end of the estimation sample. Being our time series of interest into the set of observed time series variables, indeed $y_t \in \mathbf{y}_{nt}$, the **VAR forecasting model** is:

$$y^{VAR}_{T+h} = \underline{\alpha} + \delta_h(L) y_T + \boldsymbol{\gamma}'_h(L) \tilde{\mathbf{y}}_T + \epsilon_{T+h} \qquad\qquad h = 1, \ldots, 12 \tag{4.6}$$

where $\tilde{\mathbf{y}}_T$ denotes the vector of other observed time series variables in $\mathbf{y}_{nt}$ and $\boldsymbol{\gamma}'_h(L) = 1 - \gamma_1 L^{-h} - \ldots - \gamma_p L^{p-h}$ denotes the autoregressive polynominal shifted h steps ahead that loads the past hystory of $\tilde{\mathbf{y}}_T$. The VAR forecasts have been generated by estimating the previous equation by OLS for each forecasting horizon. What we obtain is:

$$\hat{y}^{VAR}_{T+h|T} = \underline{\hat{\alpha}} + \hat{\delta}_h(L) y_T + \hat{\boldsymbol{\gamma}}'_h(L) \tilde{\mathbf{y}}_T \qquad\qquad h = 1, \ldots, 12$$

where $\hat{y}^{VAR}_{T+h|T}$ is the desidered point forecast predictor used in *Equation (4.2)*.

4.2.3 Forecasting with the Diffusion Index Model

Let $\mathbf{x}_{Nt} = (x_{1t}, x_{2t}, \ldots, x_{Nt})'$ be the N-dimensional vector (with N large) of all observed stationary time series variables in our data-set. Under the so-called **Diffusion Index (DI) model** or **static factor model** of Stock and Watson (2002a,b), $\mathbf{x}_{Nt}$ can be decomposed as:

$$\begin{aligned} \underbrace{\mathbf{x}_{NT}}_{(N\times 1)} &= \underbrace{\boldsymbol{\Lambda}_N}_{(N\times r)}\underbrace{\mathbf{F}_{NT}}_{(r\times 1)} + \underbrace{\boldsymbol{\xi}_{NT}}_{(N\times 1)} = \boldsymbol{\chi}_{NT} + \boldsymbol{\xi}_{NT} \\ \mathbf{F}_{NT} &= \mathbf{A}\mathbf{F}_{N,T-1} + \boldsymbol{\epsilon}_{NT} \qquad\qquad \boldsymbol{\epsilon}_{Nt} \sim iid\, \mathcal{N}(\mathbf{0}; \mathbf{Q}_{\boldsymbol{\epsilon}}) \\ \boldsymbol{\xi}_{NT} &= \boldsymbol{\Psi}\boldsymbol{\xi}_{N,T-1} + \mathbf{v}_{NT} \qquad\qquad \mathbf{v}_T \sim iid\, \mathcal{N}(\mathbf{0}; \mathbf{R}_{\mathbf{v}}) \end{aligned}$$

where $\mathbf{F}_{NT}$ denotes the vector of r static common factors, $\boldsymbol{\Lambda}_N$ denotes the matrix of factor loadings, $\boldsymbol{\chi}_{NT}$ denotes the vector of common components and $\boldsymbol{\xi}_{NT}$ denotes the vector of idiosyncratic components. We assumed diagonal variance-covariance matrices $\mathbf{Q}_{\boldsymbol{\epsilon}}$ and $\mathbf{R}_{\mathbf{v}}$, indeed we are working with an *exact Diffusion Index model*. The **DI forecasting model** can be written as:

$$y^{DI}_{T+h} = \underline{\alpha} + \underline{\boldsymbol{\beta}}'\hat{\mathbf{F}}_{NT} + \delta_h(L)y_T + \epsilon_{T+h} \qquad\qquad h = 1, \ldots, 12$$

where $\hat{\mathbf{F}}_{NT}$ are the estimated *static principal components* factors. The DI forecasts have been generated by estimating the previous equation using OLS for each forecasting horizon:

$$\hat{y}^{DI}_{T+h|T} = \hat{\underline{\alpha}} + \hat{\underline{\boldsymbol{\beta}}}'\hat{\mathbf{F}}_{NT} + \hat{\delta}_h(L)y_T \qquad\qquad h = 1, \ldots, 12$$

where $\hat{y}^{DI}_{T+h|T}$ is the desidered point forecast predictor used in *Equation (4.2)*.

4.2.4 Forecasting with the GDFM

Following Forni et al. (2000), if $\mathbf{x}_{NT}$ admits a generalized dynamic factor model (GDFM), the **measurement equation** takes the following form:

$$\mathbf{x}_{NT} = \boldsymbol{\Lambda}_N(L)\mathbf{f}_{NT} + \boldsymbol{\xi}_{NT} \tag{4.7}$$

where $\mathbf{f}_{NT}$ denotes the vector of q dynamic factors and $\mathbf{\Lambda}_N(L)$ denotes the matrix of dynamic factor loadings. The **GDFM forecasting model** can be written as:

$$y_{T+h}^{GDFM} = \underline{\alpha} + \underline{\boldsymbol{\beta}}'(L)\hat{\mathbf{f}}_{NT} + \delta_h(L)y_T + \varepsilon_{T+h} \qquad h = 1, \ldots, 12 \tag{4.8}$$

where $\hat{\mathbf{f}}_{NT}$ are the estimated *dynamic principal components* factors. The GDFM forecasts have been generated by estimating the previous equation by OLS for each forecasting horizon:

$$\hat{y}_{T+h|T}^{GDFM} = \underline{\hat{\alpha}} + \underline{\hat{\boldsymbol{\beta}}}'(L)\hat{\mathbf{f}}_{NT} + \hat{\delta}_h(L)y_T \qquad h = 1, \ldots, 12$$

where $\hat{y}_{T+h|T}^{GDFM}$ is the desidered point forecast predictor used in *Equation (4.2)*

4.2.5 Forecasting with the regular DSGE

The forecasts from the Smets and Wouters (2007) DSGE model, explained in *Section (3.3)*, have been generated using the state space representation given in *Equation (3.2)* and *Equation (3.3)* with a measurement error. The point forecast predictors has been formed by iterating on the last estimate of the unobserved state using the state equation *Equation (3.3)* and then backing out the corresponding value for the observable using the measurement equation *Equation (3.2)*. We do this using Bayesian estimation under the Metropolis-Hastings algorithm as described from **Step 1a** to **Step 2a** of *SubSection (3.2.3)*. The mean of the posterior forecast distributions is taken as the point forecast of the relevant variable. The Brooks and Gelman (1998) test has shown that all Markov chains for each estimation sample have converged nicely.

4.2.6 Forecasting with the drDSGE

The forecasts from the drDSGE model, have been generated using the state space representation formed by *Equation (3.17)*, *Equation (3.18)* and *Equation (3.19)*. As in the regular DSGE case, the point forecast predictors has been formed by iterating on the last estimate of the unobserved state using the state equation *Equation (3.18)* and then backing out the corresponding value for the observable using the measurement equation *Equation (3.17)*. We do this using Bayesian estimation under the Metropolis-within-Gibbs algorithm as described from **Step 1b** to **Step 5b** of *SubSection (3.2.3)*. The mean of the posterior forecast distributions is taken as the point forecast of the relevant variable. The Brooks and Gelman (1998)

test has shown that all Markov chains for each estimation sample have converged nicely.

4.3 Tests of equal predictive ability

To asses the significance power of the observed differences in the $MSFE$ between alternative forecasting models, we used two tests of equal predictive ability: the **unconditional predictive ability test** of Diebold and Mariano (1995) and West (1996) (hereafter: DMW test) and the **conditional predictive ability test** of Giacomini and White (2006) (hereafter: GW test).

4.3.1 Test of equal unconditional predictive ability

Let $\epsilon^1_{T+h|T}$ and $\epsilon^2_{T+h|T}$ be two forecast error series from two alternative forecasting models. The null hypothesis of equal unconditional predictive ability is:

$$H_0: \quad \mathbb{E}[L(\epsilon^1_{T+h|T}) - L(\epsilon^2_{T+h|T})] = \mathbb{E}[\Delta L_{T+h|T}] = 0 \tag{4.9}$$

where $L(\cdot)$ denotes a given loss function and $\Delta L_{T+h|T}$ denotes the loss differential series. Given a large sample from a standard normal $\mathcal{N}(0;1)$, the correspondent test statistics of equal unconditional predictive ability is:

$$S_{DMW} \ : \quad \frac{\bar{\Delta} L_\tau}{\sqrt{\frac{2\pi \hat{f}_\Delta(0)}{K}}} \tag{4.10}$$

where $\bar{\Delta} L_\tau$ denotes the sample mean of the loss differential $\Delta L_{T+h|T}$, indeed, $\bar{\Delta} L_\tau = \tau^{-1} \sum_{\tau=T}^{T_1-h} \Delta L_{\tau+h|T}$, K denotes the total number of predictions, while $\hat{f}_\Delta(0)$ denotes the estimate of the spectral density of the loss differential at frequency zero given by:

$$f_\Delta(0) = 2\pi^{-1} \sum_{\varsigma=-\infty}^{\infty} \gamma_\Delta(\varsigma)$$

where $\gamma_\Delta(\varsigma)$ denotes the autocovariances of the loss differential at displacement ς. This spectral density can be consistenly estimated using the heteroskedasticity and autocorrelation consistent (HAC) estimator proposed by Newey and West (1987):

$$\hat{f}_\Delta(0) = 2\pi^{-1}[\gamma_\Delta(0) + 2\sum_{\varsigma=1}^{C}(1 - \frac{\varsigma}{C})\hat{\gamma}_\Delta(\varsigma)]$$

where $\hat{\gamma}_{\Delta}(\varsigma)$ denotes the sample estimates of the autocovariances and C denotes the truncation lag. In this dissertation, we use a quadratic loss function, the truncation $C = h - 1$ and the critical values suggested by Diebold and Mariano (1995). The test works in this way: if the observed statistic *Equation (4.10)* is smaller than the critical value, we can reject the null hyphotesis *Equation (4.9)* stating that the difference between forecast error series is significanlty different from zero, then the observed difference in $MSFE$ has a significant power.

4.3.2 Test of equal conditional predictive ability

The DMW test has three main drawbacks. First, it answers basically to the question of which forecast was more accurate on average, it does not answer to the question of whether one can predict which forecast will be more accurate at a future date. So, it does not take into account of the information set at a given time, then it does not allow to ask whether additional information can help to identify which forecast is more appropriate for that date. Second, as pointed out by Clark and McCracken (2001), it has low power in finite samples. Third, it cannot accommodate Bayesian, no-parametric and semi-parametric method estimations. To overcome these drawbacks the literature has suggested conditional tests instead. At the moment the most used and suggested conditional predictive ability test is the test of Giacomini and White (2006). Differently from DMW test, the GW test has three main advantages: first, it can be applied to Bayesian, no-parametric and semi-parametric method estimations; second, under a rolling regression scheme the limiting distribution of the test statistic is still standard normal; third, it has higher power in finite samples.

The null hypothesis of equal conditional predictive ability is:

$$H_0 : \quad \mathbb{E}[L(\epsilon^1_{T+h|T}) - L(\epsilon^2_{T+h|T})|\mathcal{I}_T] = \mathbb{E}[\Delta L_{T+h|T}|\mathcal{I}_T] = 0 \tag{4.11}$$

where $L(\cdot)$ denotes a given loss function, $\Delta L_{T+h|T}$ denotes the loss differential series and $\mathcal{I}_T$ denotes the information set matured at time T, indeed, the end of the s-th rolling sample. The test statistic of Giacomini and White (2006) can be viewed as the test statistic of Diebold and Mariano (1995) and West (1996) given in *Equation (4.10)*, because under the rolling regression scheme, Giacomini and White (2006) show that the limiting distribution of the test

statistic is still standard normal. In this dissertation, we use a quadratic loss function and the critical values suggested by Giacomini and White (2006). The test works in the same way as the unconditional one.

4.4 Empirical results

This section presents the results of this dissertation. We start providing the relative mean square forecasting error ($rMSFE$) analysis in *SubSection (4.4.1)*, then to asses its significance we apply in *Subsection (4.4.2)* the tests of equal predictive ability previously explained. Since the GW test is superior than the DMW test, the final results of the dissertation come from only the interpretation of the conditional test.

4.4.1 The mean square forecasting error analysis

We set up the $rMSFE$ analysis calculating *Equation (4.1)* between the following combinations of forecasting models:

***Figure (4.1)*: Models versus the unconditional mean** The *Figure (4.1)* plots the $rMSFE$ of forecasting models respect to the unconditional mean of the series of interest. In the upper graph, we have: $rMSFE(m, \bar{y}) = 1 - \frac{MSFE(m)}{MSFE(\bar{y})}$, while in the lower graph we have: $rMSFE(m, \bar{\pi}) = 1 - \frac{MSFE(m)}{MSFE(\bar{\pi})}$, where m={AR(p^*),VAR(p^*),DI(r^*),DSGE,drDSGE} are the models in competition, $\bar{y}$ is the unconditional mean of the output growth, and $\bar{\pi}$ is the unconditional mean of inflation. The observed values are reported in *Table (4.1)*, where the better $rMSFEs$ for any forecasting horizon h are denoted in bold. These values depend critically upon the choice of: the number of lags p, the number of static factors r, and the number of dynamic factors q. The order p has been estimated using the Bayesian Information Criterion (BIC), the number of static factors r has been estimated using the Alessi et al. (2007) criterion as discussed in *SubSection (2.4.1)*, while the number of dynamic factors q has been estimated using the Hallin and Liška (2007) as discussed in *SubSection (2.4.2)*. To determine r and q, we used *Figure (4.2)* where the graph on the left suggests the optimal number of static factor r^*, while the graph on the right suggests the optimal number of dynamic factor q^*. Practically, to determine r, we need to look for the first zero variance interval for c (the dotted blu line on the left graph), corresponding to a stable value of $\hat{r}_{IC} < r_{max}$. This interval is located in

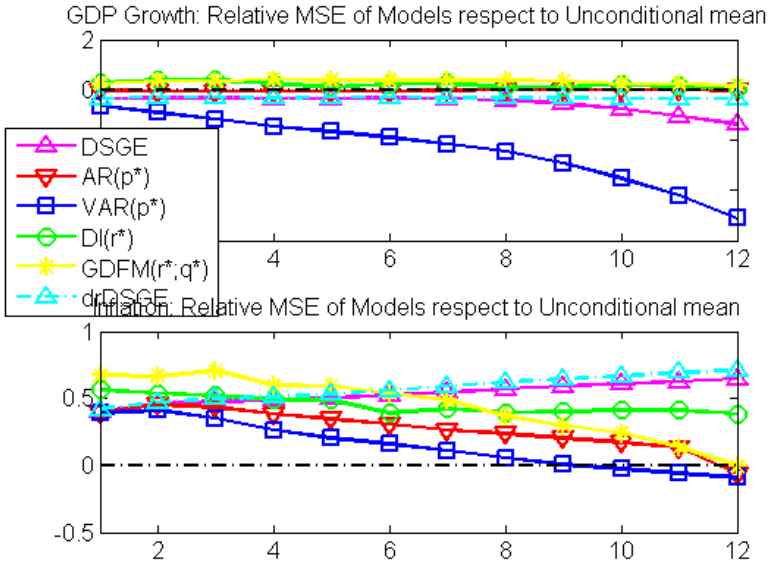

Figure 4.1: The figure plots the $rMSFEs$ of forecasting models relatively to the time series unconditional mean. The corresponding values are reported in *Table (4.1)*.

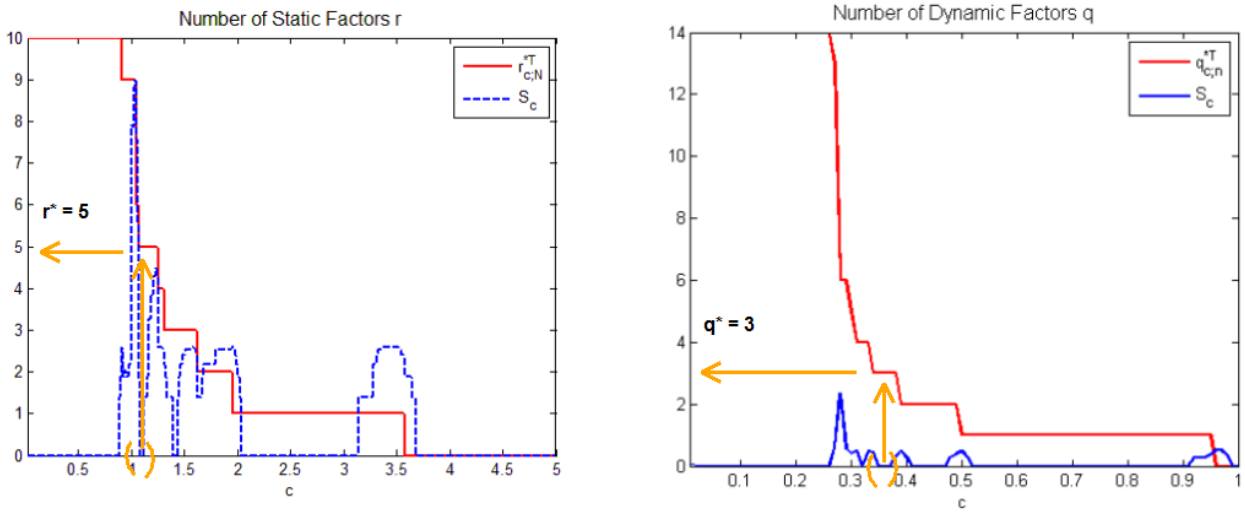

Figure 4.2: The figure plots the Alessi et al. (2007) criteria [on the left] and the Hallin and Liška (2007) criteria [on the right] used to determine respectively the number of static factor r and the number od dynamic factors q.

$rMSFE$ **of models versus the unconditional mean**						
Output Growth						
	AR(p^*)	DI(r^*)	VAR(p^*)	DSGE	GDFM(r^*,q^*)	drDSGE
$h=1$	-0.0320	**0.3037**	-0.6299	-0.3653	0.2226	-0.3553
$h=2$	-0.0496	**0.3694**	-0.9077	-0.3223	0.3634	-0.3011
$h=3$	-0.0718	**0.3718**	-1.1813	-0.3192	0.3594	-0.3002
$h=4$	-0.0494	0.2700	-1.4518	-0.3386	**0.3984**	-0.3156
$h=5$	-0.0580	0.0801	-1.6694	-0.3409	**0.3854**	-0.2811
$h=6$	-0.0469	0.2318	-1.8954	-0.3166	**0.3843**	-0.3011
$h=7$	-0.0234	0.2512	-2.1720	-0.3473	**0.3863**	-0.3173
$h=8$	-0.0034	0.0762	-2.4567	-0.4047	**0.3897**	-0.3247
$h=9$	-0.0061	0.1225	-2.9116	-0.5428	**0.3472**	-0.3328
$h=10$	0.0015	0.2055	-3.5283	-0.7684	**0.2474**	-0.3384
$h=11$	-0.0028	0.1399	-4.2084	-1.0366	**0.2258**	-0.3401
$h=12$	-0.0158	0.0304	-5.1195	-1.3826	**0.1580**	-0.3446
Inflation						
$h=1$	0.3940	0.5637	0.3876	0.4020	**0.6738**	0.4195
$h=2$	0.4558	0.5388	0.4094	0.4564	**0.6653**	0.4694
$h=3$	0.4350	0.5225	0.3514	0.4763	**0.7058**	0.4998
$h=4$	0.3819	0.4906	0.2620	0.4854	**0.5956**	0.5094
$h=5$	0.3448	0.4899	0.2015	0.5001	**0.5907**	0.5321
$h=6$	0.3068	0.3882	0.1590	0.5255	0.5401	**0.5615**
$h=7$	0.2659	0.4245	0.1077	0.5443	0.4861	**0.5943**
$h=8$	0.2360	0.3953	0.0533	0.5699	0.3673	**0.6196**
$h=9$	0.2008	0.3998	0.0082	0.5939	0.2992	**0.6459**
$h=10$	0.1725	0.4098	-0.0287	0.6108	0.2427	**0.6608**
$h=11$	0.1334	0.4104	-0.0618	0.6258	0.1377	**0.6958**
$h=12$	-0.0623	0.3863	-0.0913	0.6456	-0.0042	**0.7096**

Table 4.1: The entries in the table are the $rMSFEs$ of alternative forecasting models relatively to the time series unconditional mean. A positive entry indicates model informative forecasts. A negative entry indicates noninformative model forecasts. The entries in bold indicate the most informative model forecasts for any forecasting horizon h. For example, for output growth at one step ahead, the most informative forecasts are produced by the DI model with r^*=3 static factors, while the AR(p^*), the VAR(p^*), the DSGE and the drDSGE yield noninformative forecasts that are outperformed by the unconditional mean of the series.

correspondence of $r^* = 5$ static factors. So in *Figure (4.1)*, we have estimated the DI model with $r^* = 5$ static factors. The same reasoning is valid to estimate q. Then we need to look for the first zero variance interval for c (the blu line on the right graph), corresponding to a stable value of $\hat{q}_{IC} < q_{max}$. This interval is located in correspondence of $q^* = 3$ dynamic factors. So, we have estimated the GDFM with $r^* = 5$ static factors and $q^* = 3$ dynamic factors.

In terms of forecasting output growth, we found that factor models yield lower $rMSFEs$ respect to the all other competitive models in both short and long run. In particular, the DI(r^*) model produces lower $MSFEs$ in the short run (up to 3 quarters ahead), while the GDFM(r^*,q^*) yields lower $MSFEs$ in the long run (from 4 quarter up to 12 quarters ahead). Therefore, the AR(p^*), the VAR(p^*), the DSGE, and the drDSGE do not provide informative forecasts (only the AR(p^*) model has a positive $rMSFE$ at $h = 10$), meaning that the unconditional mean should be used instead.

With regard of inflation, we found that the GDFM(r^*,q^*) yields lower $MSFEs$ in the short run (up to 5 quarters ahead), while the drDSGE produces lower $MSFEs$ in the long run (from 6 quarter up to 12 quarters ahead). Therefore, we discovered that the DI(r^*) is able to produce lower $MSFEs$ than the GDFM(r^*,q^*) in the long run (from 8 quarter ahead up to 12 quarters ahead). This results is against the prelevant literature who gives to the GDFM better accurancy in forecasting time series variables than DI especially in the long run (Forni et al. (2000) and Forni et al. (2005)).

***Figure (4.3)*: DI(r) versus AR(p^*)** The *Figure (4.3)* plots the $rMSFE$ of diffusion index model with altenative number of static factors, respect to the autoregressive model with the optimal lag p^* fixed using the BIC. In the upper graph, we have: $rMSFE(DI(r), AR(p^*))|_{y_t} = 1-\frac{MSFE(DI(r))}{MSFE(AR(p^*))}|_{y_t}$ for $r = BIC, 1, 2, \ldots, 7$, while in the lower graph we have: $rMSFE(DI(r), AR(p^*))|_{\pi_t} = 1-\frac{MSFE(DI(r))}{MSFE(AR(p^*))}|_{\pi_t}$ for $r = BIC, 1, 2, \ldots, 7$. The observed values are reported in *Table (4.2)*, where the better $rMSFEs$ for any forecasting horizon h are denoted in bold.

For both output growth and inflation we see that only few static factors are needed to outperform the AR(p^*) model. Infact, we need just 2 factors to outperforme the AR(p^*) model for any forecasting horizon. It confirms the findings of Stock and Watson (2002b) where their DI model was found superior in $MSFE$ than an autoregressive process.

With regard to output growth, there are considerable forecasting gains when we pass

$rMSFE$ **of** $DI(r)$ **with** $r = BIC, 1, 2, \ldots, 7$ **versus** $AR(p^*)$								
Output Growth								
	DI(BIC)	DI(1)	DI(2)	DI(3)	DI(4)	DI(5)	DI(6)	DI(7)
$h = 1$	0.3253	0.1936	**0.3313**	0.2950	0.2588	0.2773	0.3258	0.3181
$h = 2$	0.3992	0.0895	0.4150	0.3906	0.3869	**0.4171**	0.4131	0.4093
$h = 3$	0.4138	0.0774	0.3727	0.3685	0.4108	**0.4232**	0.3623	0.3785
$h = 4$	0.3043	0.0462	0.2391	0.3312	**0.3901**	0.3538	0.3223	0.3436
$h = 5$	0.1305	0.0043	0.1185	0.2928	0.3188	**0.3874**	0.3349	0.3527
$h = 6$	0.2662	0.0016	0.2499	0.3430	**0.3654**	0.3576	0.3191	0.3171
$h = 7$	0.2684	-0.0934	0.2534	0.3289	**0.3748**	0.2443	0.2804	0.3095
$h = 8$	0.0793	-0.1250	0.2714	0.2843	**0.3395**	0.0351	0.1949	0.2662
$h = 9$	0.1278	-0.0303	0.2565	0.2461	**0.3267**	0.0493	0.1002	0.2052
$h = 10$	0.2043	-0.0390	0.2162	0.2182	**0.2732**	0.1217	0.1468	0.1341
$h = 11$	0.1423	-0.0319	0.2467	0.2486	**0.3122**	0.1593	0.1325	0.1761
$h = 12$	0.0454	-0.0948	0.2354	0.2422	**0.3077**	0.0820	0.0599	0.1173
Inflation								
$h = 1$	0.2801	0.0342	0.3882	0.2882	0.2736	0.3903	0.3339	**0.4036**
$h = 2$	0.1525	-0.0400	0.2362	0.1276	0.0659	0.2611	0.3557	**0.3691**
$h = 3$	0.1549	-0.0304	0.2723	0.2169	0.1616	0.1729	0.3634	**0.4275**
$h = 4$	0.1758	0.0463	0.2602	0.1270	0.2031	0.0936	0.3006	**0.4248**
$h = 5$	0.2215	0.1014	0.1706	0.1015	0.2194	0.1008	0.2826	**0.3728**
$h = 6$	0.1174	0.1647	0.2509	0.2991	0.2522	0.1686	0.4068	**0.4226**
$h = 7$	0.2160	0.2336	0.1077	0.1929	0.2917	0.1485	0.3864	**0.4372**
$h = 8$	0.2085	0.2723	0.2289	0.0921	0.2356	0.2148	0.4084	**0.4577**
$h = 9$	0.2491	0.3044	0.0892	0.0129	0.2073	0.1974	**0.4005**	0.3624
$h = 10$	0.2868	0.3375	0.0526	-0.0504	0.2020	0.1767	**0.3795**	0.3711
$h = 11$	0.3197	**0.3602**	0.1173	-0.0077	0.2374	0.1362	0.2382	0.3576
$h = 12$	0.4223	**0.4321**	0.2312	0.1315	0.3115	0.1961	0.3279	0.3883

Table 4.2: The entries in the table are the $rMSFEs$ of diffusion index (DI(r)) models with an alternative number of static factors $r = BIC, 1, 2, \ldots, 7$ relatively to the autoregressive model (AR(p)) with the lag p fixed using the BIC. A positive entry indicates DI informative forecasts, while a negative entry indicates DI noninformative forecasts. The entries in bold indicate the most informative forecasts for any forecasting horizon h. For example, for output growth at one step ahead, the most informative forecasts are produced by the DI model with two static factors.

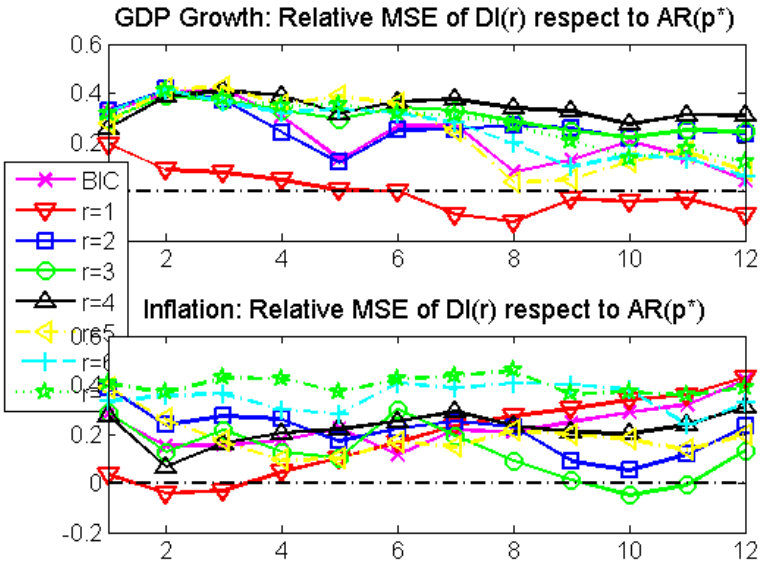

Figure 4.3: The figure plots the $rMSFEs$ of diffusion index models with an alternative number of static factors $r = BIC, 1, 2, \ldots, 7$ relatively to the autoregressive model (AR) with the lag p fixed using the BIC. The corresponding values are reported in *Table (4.2)*.

from 2 to 4 factors especially when the forecasting horizon increases. At 4 quarters ahead, the DI(4) yields 15.1% higher $rMSFE$ than DI(2), at 6 quarters ahead, the DI(4) yields 11.55% higher $rMSFE$ than DI(2), while at 12 quarters ahead the DI(4) yields 7.23% higher $rMSFE$ than DI(2).

With regard to inflation, there are considerable gains when we consider a larger number of factors, 6 or 7, at least for the short and medium run. At 1 quarter ahead, the DI(7) yields 36.94% higher $rMSFE$ than DI(1), and 6.97% higher $rMSFE$ than DI(6). At 6 quarters ahead, the DI(7) yields 25.79% higher $rMSFE$ than DI(1), and 1.58% higher $rMSFE$ than DI(6). But at 12 quarter ahead, the DI(1) yields 4.38% higher $rMSFE$ than DI(7), and 10.42% higher $rMSFE$ than DI(6).

***Figure (4.4)*: regular DSGE versus VAR(p)** The *Figure (4.4)* plots the $rMSFE$ of DSGE model respect to the VAR(p) model with alternative number of lags p. In the upper graph, we have: $rMSFE(DSGE; VAR(p))|_{y_t} = 1 - \frac{MSFE(DSGE)}{MSFE(VAR(p))}|_{y_t}$ for $p = BIC, 1, 2, \ldots, 5$, while in the lower graph we have: $rMSFE(DSGE; VAR(p))|_{\pi_t} = 1 - \frac{MSFE(DSGE)}{MSFE(VAR(p))}|_{\pi_t}$ for $p = BIC, 1, 2, \ldots, 5$. The observed values are reported in *Table (4.3)*, where we have denoted in bold the cases where forecasting using a VAR is superior than forecasting with the DSGE, and in italic the values of $rMSFEs$ for which the underlying $VAR(p)$ loses less respect to the DSGE.

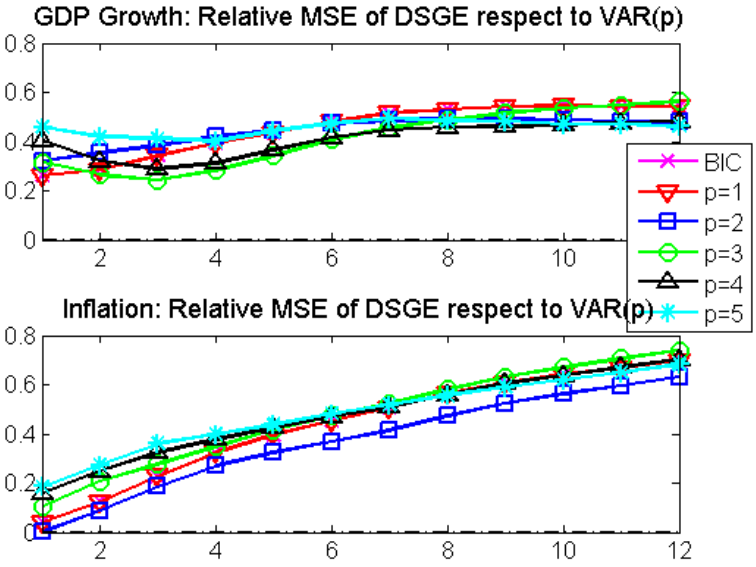

Figure 4.4: The rMSFE of $DSGE$ versus the $VAR(p)$.

$rMSFE$ **of** $DSGE$ **versus** $VAR(p)$ **with** $p = BIC, 1, 2, \ldots, 5$						
Output Growth						
	VAR(BIC)	VAR(1)	VAR(2)	VAR(3)	VAR(4)	VAR(5)
$h = 1$	0.2602	*0.2602*	0.3210	0.3158	0.4055	0.4592
$h = 2$	0.2840	0.2840	0.3525	*0.2656*	0.3219	0.4187
$h = 3$	0.3430	0.3430	0.3849	*0.2430*	0.2895	0.4129
$h = 4$	0.3923	0.3923	0.4191	*0.2833*	0.3126	0.4003
$h = 5$	0.4357	0.4357	0.4434	*0.3419*	0.3643	0.4433
$h = 6$	0.4833	0.4833	0.4725	*0.4088*	0.4157	0.4728
$h = 7$	0.5141	0.5141	0.4851	0.4557	*0.4461*	0.4951
$h = 8$	0.5310	0.5310	0.4902	0.4887	*0.4593*	0.4843
$h = 9$	0.5423	0.5423	0.4924	0.5147	*0.4622*	0.4806
$h = 10$	0.5451	0.5451	0.4889	0.5347	*0.4686*	0.4701
$h = 11$	0.5426	0.5426	0.4815	0.5487	0.4755	*0.4701*
$h = 12$	0.5431	0.5431	0.4798	0.5632	0.4823	*0.4623*
Inflation						
$h = 1$	0.0399	0.0399	*0.0013*	0.1074	0.1603	0.1856
$h = 2$	0.1227	0.1227	*0.0886*	0.2087	0.2496	0.2726
$h = 3$	0.2271	0.2271	*0.1857*	0.2765	0.3246	0.3607
$h = 4$	0.3230	0.3230	*0.2705*	0.3477	0.3770	0.3992
$h = 5$	0.3947	0.3947	*0.3251*	0.4161	0.4250	0.4379
$h = 6$	0.4529	0.4529	*0.3688*	0.4745	0.4697	0.4806
$h = 7$	0.5048	0.5048	*0.4147*	0.5241	0.5103	0.5182
$h = 8$	0.5614	0.5614	*0.4726*	0.5816	0.5611	0.5568
$h = 9$	0.6068	0.6068	*0.5232*	0.6311	0.6039	0.5907
$h = 10$	0.6395	0.6395	*0.5623*	0.6706	0.6384	0.6192
$h = 11$	0.6673	0.6673	*0.5961*	0.7047	0.6699	0.6489
$h = 12$	0.6961	0.6961	*0.6311*	0.7391	0.7031	0.6827

Table 4.3: The entries in the table are the $rMSFEs$ of the dynamic stochastic general equilibrium (DSGE) model of Smets and Wouters (2007) relatively to the vector autoregressive model (VAR(p)) with an alternative number of lags $p = BIC, 1, 2, \ldots, 5$. A positive entry indicates DSGE informative forecasts, while a negative entry indicates a noninformative DSGE forecasts. The entries in italic indicate the VAR model that loses less respect to the regular DSGE.

For both time series, the table shows that there are no cases where the the VAR(p) model is able to produce lower $MSFE$ than the the DSGE of Smets and Wouters (2007). This result is in line with the findings of Del Negro and Schorfheide (2004), where a VAR(4) is used as the benchmark. Here, we find that the DSGE model is able to outperform not only the VAR(4) but all the VAR models considerared. For the output growth, the VAR(1) is the VAR model that loses less respect to the DSGE at 1 quarter ahead, while in the long run is the VAR(5) the model that minimize the loses respect to the regular DSGE. The results of inflation are quite similar. The only difference is that now is the VAR(2) the model that loses less respect to the regular DSGE.

***Figure (4.5)*: DI(r) versus GDFM(r^*,q^*)** The *Figure (4.5)* plots the $rMSFE$ of DI model with alternative number of static factors r respect to the GDFM($r^*;q^*$). In the upper graph, we have: $rMSFE(DI(r), GDFM(r^*,q^*))|_{y_t} = 1 - \frac{MSFE(DI(r))}{MSFE(GDFM(r^*;q^*))}|_{y_t}$ for $r = BIC, 1, 2, \ldots, 7$, while in the lower graph we have: $rMSFE(DI(r), GDFM(r^*,q^*))|_{\pi_t} = 1 - \frac{MSFE(DI(r))}{MSFE(GDFM(r^*;q^*))}|_{\pi_t}$ for $r = BIC, 1, 2, \ldots, 7$. The observed values are reported in

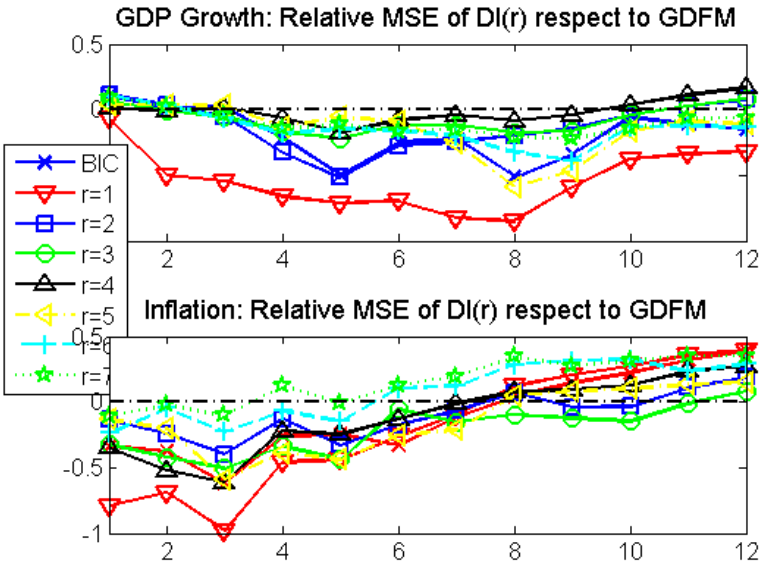

Figure 4.5: The figure plots the $rMSFEs$ of the diffusion index model (DI(r)) with an alternative number of static factors $r = BIC, 1, 2, \ldots, 7$ relatively to the generalized dynamic factor model (GDFM(p, q)) with the number of static factors r fixed using Alessi et al. (2007) criterion and the number of dynamic factors q fixed using the Hallin and Liška (2007) criterion. The corresponding values are reported in *Table (4.4)*.

Table (4.4), where we have denoted in bold the cases where forecasting using a $DI(r)$ is superior than forecasting with the GDFM(r^*,p^*), and in italic the values of $rMSFEs$ for which the underlying $DI(r)$ loses less respect to the GDFM(r^*,p^*).

$rMSFE$ **of** $DI(r)$ **versus** $GDFM(r^*;q^*)$ **with** $r = BIC, 1, 2, \ldots, 7$								
Output Growth								
	DI(BIC)	DI(1)	DI(2)	DI(3)	DI(4)	DI(5)	DI(6)	DI(7)
$h = 1$	0.1044	-0.0705	**0.1123**	0.0641	0.0160	0.0406	0.1050	0.0947
$h = 2$	0.0095	-0.5011	0.0355	-0.0046	-0.0108	**0.0390**	0.0325	0.0261
$h = 3$	0.0193	-0.5435	-0.0495	-0.0565	0.0142	**0.0350**	-0.0668	-0.0398
$h = 4$	-0.2134	-0.6637	-0.3273	-0.1665	*-0.0639*	-0.1272	-0.1821	-0.1450
$h = 5$	-0.4967	-0.7139	-0.5174	-0.2173	-0.1727	*-0.0545*	-0.1449	-0.1142
$h = 6$	-0.2477	-0.6976	0.5925	-0.1172	*-0.0790*	-0.0923	-0.1577	-0.1612
$h = 7$	-0.2200	-0.8233	-0.2449	-0.1190	*-0.0425*	-0.2601	-0.1999	-0.1515
$h = 8$	-0.5137	-0.8496	-0.1979	-0.1767	*-0.0859*	-0.5864	-0.3237	-0.2065
$h = 9$	-0.3442	-0.5880	-0.1460	-0.1620	*-0.0377*	-0.4653	-0.3867	-0.2250
$h = 10$	-0.0557	-0.3785	-0.0400	-0.0373	**0.0357**	-0.1652	-0.1320	-0.1489
$h = 11$	-0.1110	-0.3366	0.0242	0.0268	**0.1092**	-0.0890	-0.1236	-0.0672
$h = 12$	-0.1516	-0.3208	0.0775	0.0858	**0.1648**	-0.1075	-0.1342	-0.0649
Inflation								
$h = 1$	-0.3373	-0.7942	-0.1365	-0.3223	-0.3495	-0.1327	-0.2374	*-0.1080*
$h = 2$	-0.3779	-0.6909	-0.2419	-0.4184	-0.5187	-0.2014	-0.0475	*-0.0258*
$h = 3$	-0.6231	-0.9789	-0.3976	-0.5038	-0.6102	-0.5884	-0.2226	*-0.0995*
$h = 4$	-0.2596	-0.4576	-0.1307	-0.3342	-0.2179	-0.3853	-0.0689	**0.1209**
$h = 5$	-0.2462	-0.4384	-0.3278	-0.4383	-0.2495	-0.4395	-0.1484	*-0.0041*
$h = 6$	-0.3303	-0.2590	-0.1718	-0.0564	-0.1272	-0.2532	0.1059	**0.1297**
$h = 7$	-0.1198	-0.0946	-0.0699	-0.1528	-0.0117	-0.2162	0.1236	**0.1961**
$h = 8$	0.0443	0.1214	0.0689	-0.0963	0.0771	0.0519	0.2857	**0.3452**
$h = 9$	0.1436	0.2067	-0.0387	-0.1257	0.0960	0.0847	0.3164	**0.2729**
$h = 10$	0.2207	0.2761	-0.0352	-0.1478	0.1280	0.1004	**0.3220**	0.3129
$h = 11$	0.3162	**0.3569**	0.1129	-0.0128	0.2336	0.1318	0.2343	0.3544
$h = 12$	0.3889	**0.3992**	0.1867	0.0812	0.2716	0.1495	0.2890	0.3529

Table 4.4: The entries in the table are the $rMSFEs$ of the diffusion index model (DI(r)) with an alternative number of static factors $r = BIC, 1, 2, \ldots, 7$ relatively to the generalized dynamic factor model (GDFM(r, q)) with the number of static factors r fixed using Alessi et al. (2007) criterion and the number of dynamic factors q fixed using the Hallin and Liška (2007) criterion. We found $r^* = 5$ e $q^* = 3$. A positive entry indicates DI informative forecasts, while a negative entry indicates noninformative DI forecasts. The entries in italic indicate the DI model that loses less respect to the GDFM. The entries in bold indicate the most informative forecasts for any forecasting horizion h. For example, for inflation at one step ahead, there are no cases in which a DI yields informative forecasts and the DI(7) is the model that loses less, while at four step ahead the DI(7) is able to produce informative forecasts.

Most of the values contained in the table are negative, meaning that there are few occasions in which the DI model yield a lower $MSFE$. Regard to output growth, there are few cases where the GDFM is outperformed by the DI, while for inflation these cases are increased. For the output growth, the DI model with 5 and 4 factors tend to outperform the GDFM(r^*, q^*) in both short term (up to 2 quarters ahead) and long term (from 10 to 12 quarters ahead). For inflation, there are no cases where the DI model is able to produce informative forecasrs in the short run, while in the medium run and in the long run the DI(7) and the DI(1) are able to produce lower $MSFEs$ respectively.

***Figure (4.6)*: DSGE models versus GDFM(r^*,q^*)** The *Figure (4.6)* plots the $rMSFE$ of DSGE models respect to the GDFM$(r^*; q^*)$. In the upper graph, we have: $rMSFE(z, GDFM(r^*; q^*))|_{y_t}$ $1-\frac{MSFE(z)}{MSFE(GDFM(r^*,q^*))}|_{y_t}$ for $z = \{DSGE, drDSGE\}$, while in the lower graph we have: $rMSFE(z; GDFM(r^*; q^*))|_{\pi_t} = 1-\frac{MSFE(z)}{MSFE(GDFM(r^*,q^*))}|_{\pi_t}$ where $z = \{DSGE, drDSGE\}$. The observed values are reported in *Table (4.5)*, where we have denoted in bold the cases

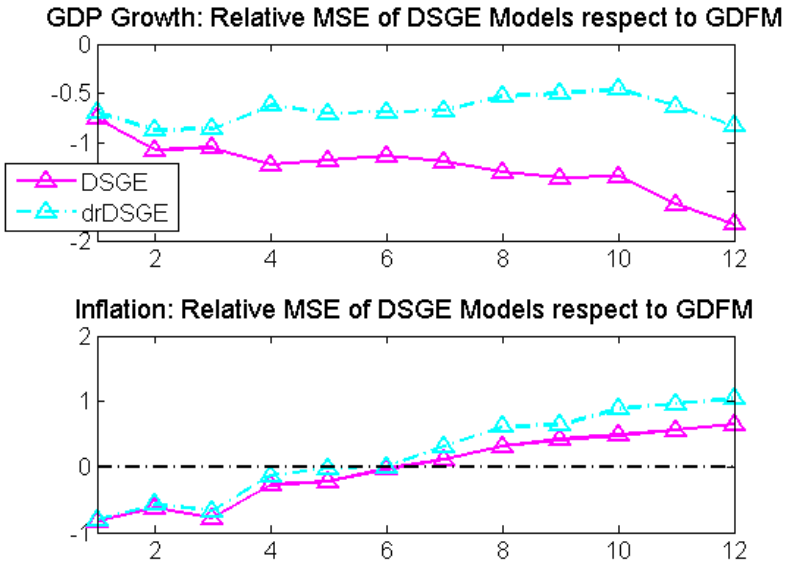

Figure 4.6: The figure plots the $rMSFEs$ of the dynamic stochastic general equilibrium models relatively to the generalized dynamic factor model (GDFM(r, q)) with $r^* = 5$ e $q^* = 3$. The corresponding values are reported in *Table (4.5)*.

where a DSGE model is able to outperform the GDFM(r^*,q^*) in terms of $rMSFE$. About the output growth, the GDFM yields lower $MSFEs$ than the DSGE models for any forecasting horizon. It confirms the results of *Table (4.1)* where the DSGE models was found to generate higher $rMSFEs$ than the GDFM.

$rMSFE$ **of DSGE models versus GDFM(r^*,q^*)**				
	Output Growth		**Inflation**	
	DSGE	drDSGE	DSGE	drDSGE
$h=1$	-0.7573	-0.6963	-0.8331	-0.8129
$h=2$	-1.0791	-0.8770	-0.6239	-0.5754
$h=3$	-1.0595	-0.8643	-0.7801	-0.6705
$h=4$	-1.2252	-0.6249	-0.2724	-0.1309
$h=5$	-1.1810	-0.7016	-0.2213	-0.0215
$h=6$	-1.1334	-0.6885	-0.0328	-0.0019
$h=7$	-1.1926	-0.6752	0.1142	**0.3131**
$h=8$	-1.3087	-0.5316	0.3287	**0.6204**
$h=9$	-1.3632	-0.5034	0.4203	**0.6396**
$h=10$	-1.3477	-0.4597	0.4849	**0.8861**
$h=11$	-1.6358	-0.5606	0.5658	**0.9660**
$h=12$	-1.8296	-0.7987	0.6465	**1.0471**

Table 4.5: The entries in the table are the $rMSFEs$ of the dynamic stochastic general equilibrium models relatively to the generalized dynamic factor model (GDFM(r,q)) with $r^*=5$ e $q^*=3$. A positive entry indicates DSGE informative forecasts, while a negative entry indicates noninformative DSGE forecasts. The entries in bold indicate the most informative DSGE forecasts. For example, for output growth there are no cases in which DSGE models yield informative forecasts, while for inflation at eight step ahead both DSGE and drDSGE produce informative forecasts but the drDSGE forecasts are more informative.

About the output growth, the GDFM yields lower $MSFEs$ than the DSGE models for any forecasting horizon. It confirms the results of *Table (4.1)* where the DSGE models was found to generate higher $rMSFEs$ than the GDFM. Differently, when we have to forecast inflation, we find that DSGE models are able to produce lower $MSFEs$ than the GDFM only in the long run (from 7 to 12 quarters ahead).

The interesting result is the MSFE performance gap between the DSGE and the drDSGE. This gap, as shown in *Figure (4.6)* increases when the forecasting horizon is increased as a pair of open scissors. Regarding the output growth, at 1 quarter ahead, the drDSGE loses 6.01% less (in absolute value) than the DSGE, at 6 quarters ahead, the drDSGE loses 44.49% less than the DSGE, and at 12 quarters ahead, the drDSGE loses 103.09% less than the DSGE. The same situation happens for inflation. At 1 quarter ahead, the drDSGE loses 2.02% less (in absolute value) than the regular DSGE, at 6 quarters ahead, the drDSGE loses 3.09% less than the regular DSGE, and at 12 quarters ahead, the drDSGE loses 40.06% less than the regular DSGE. This result is in line with the findings of Boivin and Giannoni (2006), who show that more accurate estimates implies better forecasts at least one step ahead.

Concluding, the $rMSFE$ analysis has pointed out that output growth is not forecasted informatively by DSGE models, while factor models yield lower and informative $MSFEs$ for

any forecasting horizon. Simmetrically, for the inflation, DSGE models tend to produce lower $MSFEs$ than factor models especially in the long run. We could take these results as definitive, but since the $MSFE$ analysis has not significance power, we have to work on forecasting inference implementing predictive ability tests. Among these tests, we interpreted the conditional predictive ability test of Giacomini and White (2006).

4.4.2 Equal predictive ability results

The *Table (4.6)* reports the test statistic of the unconditional equal predictive ability test, while the *Table (4.7)* reports the test statistic of the conditional equal predictive ability test. These statistics have the following interpretation: *plus signs* indicate that the forecating model in rows have *lower* mean squared forecasting errors than the corresponding forecating model in columns, then the model in row outperforms significantly the model in column. Simmetrically, `negative signs` indicate that the forecating model in rows have `higher` mean squared forecast errors than the corresponding forecasting model in columns, then the model in column outperforms significantly the model in column. We denoted in bold entries that are significant at 5% level, while we denoted in underlined bold entries that are significant at 1% level. Critical levels of test statistics are fixed as suggested by Diebold and Mariano (1995) and West (1996) and Giacomini and White(2006) respectively. The interpretation of the GW test is considered as definitive.

Unconditional Predictive Ability Test: Looking at output growth forecasts, we find that there are not occasions where DSGE models show significant differences between competing models. The GDFM outperforms significantly DSGE models in both short and long run, while the DI model outperforms significantly only the regular DSGE at 1 quarter ahead and at 12 quarters ahead. Therefore, the drDSGE model produces significantly better forecasts than the regular DSGE in both short and long run.

Looking at inflation forecasts, the GDFM is able to outperform significantly DSGE models only 1 quarter ahead, while at 8 and 12 quarters ahead the drDSGE dominates significatively all other competing models. Surprising also the regular DSGE is found significant better than the GDFM at 12 quarters ahead. This result will not be confirmed by the GW test. Therefore, interestingly the DI outperforms significanlty the GDFM at 12 quarters ahead. We could take these results as definitive, but since the DMW test has low power in finite samples

Test of equal unconditional predictive ability (DMW test)

Output Growth

If $h = 1$:

	Mean	AR(p*)	DI(r*)	VAR(p*)	DSGE	GDFM(r*;q*)
AR(p*)	-0.7918	0	0	0	0	0
DI(p*)	1.7463	**1.9841**	0	0	0	0
VAR(p*)	-1.4522	-1.2933	-1.8290	0	0	0
DSGE	-1.4676	-1.2195	**-2.2254**	0.9709	0	0
GDFM(r*;q*)	1.2930	1.5014	-0.7404	1.8507	**2.3574**	0
drDSGE	-1.5135	-1.2156	-1.8882	1.2017	**2.1941**	**-2.1765**

If $h = 4$:

AR(p*)	-0.5556	0	0	0	0	0
DI(p*)	1.2366	1.4135	0	0	0	0
VAR(p*)	-1.6753	-1.5254	-1.7965	0	0	0
DSGE	-1.0382	-0.7364	-1.6713	1.5747	0	0
GDFM(r*;q*)	1.8508	1.9591	1.6143	1.8781	**2.0129**	0
drDSGE	-0.6513	-1.3356	-1.8882	1.6017	**2.1451**	**-1.9872**

If $h = 8$:

AR(p*)	-0.0512	0	0	0	0	0
DI(p*)	0.3839	0.3398	0	0	0	0
VAR(p*)	-1.6083	-1.5807	-1.7394	0	0	0
DSGE	-0.9811	-0.8915	-1.6256	1.6460	0	0
GDFM(r*;q*)	1.8207	1.8121	1.5060	1.7693	1.8888	0
drDSGE	-0.8755	-0.6091	-0.2698	2.1278	1.9126	**-2.3176**

If $h = 12$:

AR(p*)	-0.1642	0	0	0	0	0
DI(p*)	0.1482	0.2473	0	0	0	0
VAR(p*)	-1.9174	-1.8859	-1.9223	0	0	0
DSGE	**-2.3896**	**-2.2422**	**-2.4234**	1.6847	0	0
GDFM(r*;q*)	0.8046	0.9602	1.2034	1.9475	**2.6667**	0
drDSGE	-1.0103	-0.7691	-0.2698	**2.1278**	**2.4166**	**-2.5455**

Inflation

If $h = 1$:

	Mean	AR(p*)	DI(r*)	VAR(p*)	DSGE	GDFM(r*;q*)
AR(p*)	**2.6920**	0	0	0	0	0
DI(p*)	**2.6646**	1.3368	0	0	0	0
VAR(p*)	1.8728	-0.0862	-1.1656	0	0	0
DSGE	**2.4421**	0.0836	-1.1691	0.1227	0	0
GDFM(r*;q*)	**3.3779**	**2.0804**	1.4511	1.6775	**1.9935**	0
drDSGE	**2.5034**	1.0870	-0.9985	0.1459	**1.9765**	**-1.9661**

If $h = 4$:

AR(p*)	**2.0203**	0	0	0	0	0
DI(p*)	1.9513	0.7181	0	0	0	0
VAR(p*)	0.7002	-0.5918	-0.8532	0	0	0
DSGE	**2.5035**	0.6524	-0.0292	0.7443	0	0
GDFM(r*;q*)	**2.7669**	1.2880	0.9917	1.0974	0.7235	0
drDSGE	**2.5382**	**2.5476**	-0.9813	**2.5111**	**2.0198**	-1.5322

If $h = 8$:

AR(p*)	1.1756	0	0	0	0	0
DI(p*)	1.8409	1.2537	0	0	0	0
VAR(p*)	0.1158	-0.6590	-1.0122	0	0	0
DSGE	**3.1769**	**1.9926**	0.9056	1.3551	0	0
GDFM(r*;q*)	1.5614	0.7413	-0.1877	0.8808	-0.9626	0
drDSGE	**3.2154**	**2.5985**	**1.9798**	**3.3387**	**2.0223**	**2.7655**

If $h = 12$:

AR(p*)	-0.2215	0	0	0	0	0
DI(p*)	1.7187	**2.0849**	0	0	0	0
VAR(p*)	-0.1774	-0.1059	-1.2946	0	0	0
DSGE	**3.8593**	**3.0489**	1.4927	1.7108	0	0
GDFM(r*;q*)	-0.0143	0.2202	**-2.1764**	0.2105	**-2.1933**	0
drDSGE	**5.1879**	**4.8805**	**1.9801**	**3.7598**	**3.2433**	**4.0131**

Table 4.6: This table contains the results of pairwise tests of equal unconditional predictive accuracy of alternative forecasting models using a quadratic loss function. The entries in the table are the test-statistic of equal unconditional predictive ability for the methods in the corresponding row and column. A positive (negative) entry indicates that the model in row is able to produce a significant lower (higher) mean squared forecast error than the corresponding model in column. The entries in bold indicate test-statistics that are significant at 5% level. The entries in underlined bold indicate test-statistics that are significant at 1% level. For example, for output at twelve step ahead, the GDFM forecasts outperforms significantly at 1% level the drDSGE forecasts.

(Clark and McCracken (2001) and Clark and McCracken (2010)), we need to interpret the conditional predictive ability test instead.

Test of equal conditional predictive ability (GW test)

Output Growth

If $h = 1$:

	Mean	AR(p^*)	DI(r^*)	VAR(p^*)	DSGE	GDFM(r*,q*)
AR(p^*)	**8.2512**	0	0	0	0	0
DI(p^*)	5.1260	5.0767	0	0	0	0
VAR(p^*)	3.7613	3.2106	**-7.5801**	0	0	0
DSGE	2.1322	1.5354	**-8.2113**	5.2770	0	0
GDFM(r*,q*)	1.6440	1.8884	0.6445	**8.6058**	**9.2396**	0
drDSGE	2.1523	1.6877	**-8.1921**	**8.5967**	**8.1265**	**-7.9321**

If $h = 4$:

AR(p^*)	0.5407	0	0	0	0	0
DI(p^*)	2.3150	2.8183	0	0	0	0
VAR(p^*)	3.2082	2.5118	-4.0286	0	0	0
DSGE	2.9597	2.2769	-2.6651	3.4087	0	0
GDFM(r*,q*)	3.2138	3.3985	1.9478	4.2912	3.0909	0
drDSGE	3.2021	2.5847	-2.4217	4.2517	**6.0154**	-2.9321

If $h = 8$:

AR(p^*)	2.6710	0	0	0	0	0
DI(p^*)	1.3636	1.3212	0	0	0	0
VAR(p^*)	3.6685	4.1270	-3.6573	0	0	0
DSGE	3.1501	3.4670	-3.8994	2.1355	0	0
GDFM(r*,q*)	0.6531	0.4712	2.1853	2.3374	3.7045	0
drDSGE	5.2544	4.8821	-5.8621	**6.9063**	**6.4113**	-4.9321

If $h = 12$:

AR(p^*)	5.2311	0	0	0	0	0
DI(p^*)	3.9721	1.7715	0	0	0	0
VAR(p^*)	**8.9385**	**9.3304**	**-8.3023**	0	0	0
DSGE	**6.0661**	5.9044	-5.7588	**7.4603**	0	0
GDFM(r*,q*)	0.3319	1.4539	1.2591	**8.3659**	**6.2057**	0
drDSGE	**6.4923**	5.9872	**-6.5821**	**6.0567**	**8.3445**	**-6.9355**

Inflation

If $h = 1$:

	Mean	AR(p^*)	DI(r^*)	VAR(p^*)	DSGE	GDFM(r*,q*)
AR(p^*)	**15.7459**	0	0	0	0	0
DI(p^*)	**9.5378**	0.7863	0	0	0	0
VAR(p^*)	**12.0002**	5.8369	0.7711	0	0	0
DSGE	**13.8996**	4.3562	0.5602	5.8350	0	0
GDFM(r*,q*)	**21.6968**	4.6374	3.0911	4.9811	4.0713	0
drDSGE	**12.2570**	**6.0129**	0.6032	4.8701	5.8033	-3.9034

If $h = 4$:

AR(p^*)	2.9308	0	0	0	0	0
DI(p^*)	**8.0435**	1.4455	0	0	0	0
VAR(p^*)	4.8213	1.7167	1.0193	0	0	0
DSGE	4.8167	2.4049	0.9501	1.5464	0	0
GDFM(r*,q*)	**20.9087**	2.0743	1.1859	1.2578	3.6244	0
drDSGE	**6.2033**	**7.4279**	4.5611	**6.6037**	**6.5534**	5.9851

If $h = 8$:

AR(p^*)	**15.3335**	0	0	0	0	0
DI(p^*)	**13.5552**	**7.1261**	0	0	0	0
VAR(p^*)	3.5745	0.2973	2.3745	0	0	0
DSGE	**7.4830**	3.5709	2.7996	1.2484	0	0
GDFM(r*,q*)	**23.4710**	4.4193	1.3812	2.3919	2.1791	0
drDSGE	**9.7838**	**8.1772**	**5.9918**	**11.5446**	**7.9253**	**8.7679**

If $h = 12$:

AR(p^*)	**10.0696**	0	0	0	0	0
DI(p^*)	**11.7056**	**7.8748**	0	0	0	0
VAR(p^*)	**15.1905**	**11.9503**	2.1421	0	0	0
DSGE	**11.5110**	5.6169	0.8479	1.2841	0	0
GDFM(r*,q*)	**8.0811**	1.0432	1.8838	1.8486	4.2154	0
drDSGE	**18.8901**	**9.0332**	**7.0105**	**20.3044**	**8.2211**	**11.6031**

Table 4.7: This table contains the results of pairwise tests of equal conditional predictive accuracy of alternative forecasting models using a quadratic loss function. The entries in the table are the test-statistic of equal conditional predictive ability for the methods in the corresponding row and column. A positive (negative) entry indicates that the model in row is able to produce a significant lower (higher) mean squared forecast error than the corresponding model in column. The entries in bold indicate test-statistics that are significant at 5% level. The entries in underlined bold indicate test-statistics that are significant at 1% level. For example, for inflation at one step ahead, the drDSGE forecasts outperforms significantly the AR(p^*) forecasts.

Conditional Predictive Ability Test: Regarding the output growth, the test reveals on one hand that the GDFM is able to generate significantly better forecasts than DSGE models in both short and long run. On the other side, the DI model is able to outperforms significantly the DSGE models only in the short run. Therefore, is confirmed the superiority of the drDSGE in outperforming significantly the regular DSGE in the short, medium and long run.

Regarding the inflation, we discovered the most important result of the dissertation: only the drDSGE outperforms significantly all other competitive models in forecasting inflation in the long run. In other words, in the long run significant forecasts can be obtained only by combining a DSGE model with a static factor model. It means that exploiting more information on many macroeconomic time series, through the drDSGE, is important not only to obtain more accurate estimates, but also to get significant better forecasts.

Conclusion

We conducted several out-of-sample forecasting experiments to assess the forecasting power of factor models relatively to DSGE models. We found three main conclusions. First, DSGE models are significantly outperformed by the GDFM in forecasting output growth in both short and long run, while the static factor model outperforms significantly DSGE models only in the short run. Second, the most surprising result of the dissertation, we discovered that only the drDSGE outperforms significantly all other competitive models in forecasting inflation in the long run. This evidence falls out with both Wang (2009) who found that the regular DSGE of Del Negro and Schorfheide (2004) was able to generate significant better forecasts for inflation in the long run, and Paccagnini (2011) where hybrid models are found to forecast poorly. Therefore, the drDSGE outperforms significantly the regular DSGE in forecasting both output growth and inflation, confirming that exploiting more information on many macroeconomic time series, through the drDSGE, is important not only to obtain more accurate estimates, but also to get significant better forecasts. Third, in most cases, the unrestricted VAR represents the worse forecasting model, suggesting that this model should not be used as benchmark model in forecasting comparisons.

Given the wide variety of DSGE models in the literature, this dissertation should not be understood as a final research into the relative predictive ability of DSGE models relatively to factor models, but it should encourage further research in this topic. Our results raise several issues for future research. In our view four issues are preminent. First, we have shown that forecasting results vary according to the type of DSGE considered, then future research should consider a wider range of DSGE models with alternative structural restrictions. Second, being the drDSGE a static model, it would be useful to generalize its representation allowing state variables to be loaded with leads and lags. It might raise further forecasting gains. Third, we have estimated factor models assuming linearity but linearity is often not prevalent in the data-set. Then, it would be useful to introduce nonlinear dynamic factor models.

Fourth, throughout the dissertation we assumed weakly stationarity time series. Although data-set differentiation and standardization achieve stationary in most cases, this is a strong assumption that should be relaxed.

Appendix A:

the data-set used

This appendix gives an overview of the dataset used to construct the factors. The data are presented in the following ordering: series number, series mnemonic, series description and transformation code. The transformation codes are 1 = no transformation, 2 = first difference, 3 = first difference of logs, 4 = second difference of logs. All price series are obtained from Moody's Economy and all other series are obtained from `Datastream`. The series mnemonics and descriptions are taken directly from the associated sources. The interest rate spreads are calculated using the average federal funds rate obtained from Moody's Economy. The abbreviations appearing in the series descriptions are sa/sadj = seasonally adjusted, cura = current prices, seasonally adjusted, vola = volumn index, seasonally adjusted.

Table 8: The data-set used

	Mnemonic	Description	Transformation
Prices			
1	cpiuaa_us	cpi: urban consumer apparel, (1982-84=100, sa)	4
2	cpiuac_us	cpi: urban consumer commodities, (1982-84=100, sa)	4
3	cpiuad_us	cpi: urban consumer durables, (1982-84=100, sa)	4
4	cpiuam_us	cpi: urban consumer medical care, (1982-84=100, sa)	4
5	cpiuas_us	cpi: urban consumer services, (1982-84=100, sa)	4
6	cpiuat_us	cpi: urban consumer transportation, (1982-84=100, sa)	4
7	cpiul1_us	cpi: urban consumer all items less food, (1982-84=100, sa)	4
8	cpiul2_us	cpi: urban consumer all items less shelter, (1982-84=100, sa)	4
9	cpiul5_us	cpi: urban consumer all items less medical care, (1982-84=100, sa)	4
10	ppisp1000_us	ppi: stage of processing crude materials, (index 1982=100, sa)	4
11	ppisp2000_us	ppi: stage of processing intermediate materials, (index 1982=100, sa)	4
12	ppisp3000_us	ppi: stage of processing finished goods, (index 1982=100, sa)	4
13	ppisp3100_us	ppi: stage of processing finished consumer goods, (index 1982=100, sa)	4
Consumption			
14	uscdtan_b	pce durables, new autos (ar) cura	3
15	uscondurb	personal consumption expenditures durables (ar) cura	3
16	usconndrb	personal consumption expenditures nondurables (ar) cura	3
17	usconsrvb	personal consumption expenditures services (ar) cura	3
18	usperconb	personal consumption expenditures (ar) cura	3
Employment			
19	usem21_o	employed mining vola	3
20	usem23_o	employed construction vola	3
21	usem42_o	employed wholesale trade vola	3

... continued on next page

Table 8 *... continued from previous page*

	Mnemonic	Description	Transformation
22	usem81_o	employed otherservices vola	3
23	usemig_o	employed government vola	3
24	usemimd_o	employed durable goods vola	3
25	usemip_o	employed totalprivate vola	3
26	usemir_o	employed retail trade vola	3
27	usemit_o	employed trade, transportation, utilities vola	3
28	usempallo	employed nonfarm industries total (payroll survey) vola	3
29	usempg_o	employed goodsproducing vola	3
30	usempmano	employed manufacturing vola	3
31	usemps_o	employed serviceproviding vola	3
32	usemptoto	total civilian employment vola	3
33	ushlpwadq	help wanted proportion of labor markets w/rising wantad vola	1
34	usun_totq	unemployment rate sadj	2
35	usundurne	average durationof unemployment (weeks) vola	1
36	usunw14_q	unemployed distribution 5 to 14 weeks sadj	1
37	usunw15_q	unemployed distribution 15 weeks over sadj	1
38	usunw26_q	unemployed distribution 15 to 26 weeks over sadj	1
39	usunw5_q	unemployed distribution less than 5 weeks sadj	1
40	usvactoto	index of help wanted advertising vola	3
Housing			
41	ushbrm_o	housing started midwest (ar) vola	3
42	ushbrn_o	housing started northeast (ar) vola	3
43	ushbrs_o	housing started south (ar) vola	3
44	ushbrw_o	housing started west (ar) vola	3
45	ushous_o	new private housing units started (ar) vola	3
Hours and Earnings			
46	ushkim_o	avg wkly hours manufacturing vola	3
47	ushxpmano	avg overtime hours manufacturing vola	3
48	uswr23_b	avg hrly earn construction cura	4
49	uswrim_b	avg hrly earn manufacturing cura	4
Output and Income			
50	usipmbuqg	indl prod business equipment vola	3
51	usipmcogg	indl prod consumer goods vola	3
52	usipmducg	indl prod durable consumer goods vola	3
53	usipmfgsg	industrial production manufacturing (sic) vola	3
54	usipmfing	indl prod final products, total vola	3
55	usipmmatg	indl prod materials, total vola	3
56	usipmnocg	indl prod nondurable consumer goods vola	3
57	usipmprog	indl prod final products nonindustrial supplies vola	3
58	usiptot_g	industrial production total index vola	3
59	usiumfgsq	indl utilizationmanufacturing (sic) sadj	1
60	uspdispib	disposable personal income (ar) cura	3
61	uspersinb	personal income (ar) cura	3
Interest Rates			
62	uscrbbaa	corporate bond yield moody's baa, seasoned issues	2
63	uscrbyld	corporate bond yield moody's aaa, seasoned issues	2
64	ustrb3av	treasury bill secondary market rate on discount basis 3 month	2
65	ustrcn10	treasury yield adjusted to constant maturity 10 year	2
66	ustrcn1_	treasury yield adjusted to constant maturity 1 year	2
67	ustrcn5_	treasury yield adjusted to constant maturity 5 year	2
68	usytb6sm	treasury bill secondary market rate on discount basis 6 month	2
69	ussfycrbyld	spread uscrbyld federal funds	1
70	ussfycrbbaa	spread uscrbbaa federal funds	1
71	ussfytrb3av	spread ustrb3av federal funds	1

... continued on next page

Table 8 ... ***continued from previous page***

	Mnemonic	Description	Transformation
72	ussfyytb6sm	spread usytb6sm federal funds	1
73	ussfytrcn1_	spread ustrcn1_ federal funds	1
74	ussfytrcn10	spread ustrcn10 federal funds	1
75	ussfytrcn5_	spread ustrcn5_ federal funds	1
Other Time Series			
76	usm0_b	monetary base cura	4
77	usnbrrsab	nonborrowed reserves of depository institutions cura	3
78	uspmchin	chicago purchasingmanager diffusion indexinventories(sa)	1
79	uspmchlt	chicago purchasingmanager diffusion indexdeliveries(sa)	1
80	uspmchp_	chicago purchasingmanager diffusion indexprodn. (sa) sadj	1
81	ustotrsab	total reserves of depository institutions cura	3
82	usexpgdsb	exports f.a.s. cura	3
83	uscnfbusq	ism purchasing managers index (mfg survey) sadj	1

Bibliography

[1] Alessi L., M. Barigozzi and M. Capasso (2007): "A Robust Criterion for Determining the Number of Static Factors in Approximate Factor Models", *LEM Papers Series 2007/19*, Sant'Anna School of Advanced Studies, Pisa, Italy.

[2] Altissimo F., R. Cristadoro, M. Forni, M. Lippi and G. Veronese (2001): "Eurocoin: a Real Time Coincident Indicator of the Euro Area Business Cycle", *CEPR Discussion Papers*, **3108**.

[3] Altissimo F., R. Cristadoro, M. Forni, M. Lippi and G. Veronese (2010): "New Eurocoin: Tracking Economic Growth in Real Time", *The Review of Economics and Statistics*, **92**, 1024-1034.

[4] Amisano G. and O. Tristani (2010): "A DSGE Model of the Term Structure with Regime Shifts", *European Central Bank Working Paper Series*.

[5] An S. and F. Schorfheide (2007): "Bayesian Analysis of DSGE Models", *Econometric Review*, **26**, 113-172.

[6] Bai, J. and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models", *Econometrica* **70**, 191-221.

[7] Bai J. and S. Ng (2006): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions", *Econometrica* **74** 1133-1150.

[8] Bai J. and S. Ng (2007): "Determining the Number of Primitive Shocks in Factor Models", *Journal of Business and Economic Statistics*, **25**, 52-60.

[9] Banbura M., D. Giannone and L. Reichlin (2010): "Large Bayesian Vector AutoRegressions", *Journal of Applied Econometrics* **25**, 71-92.

[10] Bernanke B.S. and Boivin J. (2003): "Monetary Policy in a Data-Rich Environment", *Journal of Monetary Economics*, **50** 525-546.

[11] Boivin J. and M.P. Giannoni (2006): "DSGE Models in a Data-Rich Environment", *NBER Working Paper* n°12772.

[12] Boivin J. and M.P. Giannoni (2008): "Optimal Monetary Policy in a Data-Rich Environment", *NBER Working Paper*.

[13] Canova F. and L. Sala (2009): "Back to Square One: Identi.cation Issues in DSGE Models", *Journal of Monetary Economics* **56**, 431-449.

[14] Chamberlain G. (1983): "Funds, Factors, and Diversification in Arbitrage Pricing Models", *Econometrica* **51**, 1281-1304.

[15] Chamberlain G. and M. Rothschild (1983): "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets", *Econometrica* **51**, 1305-1324.

[16] Christoffel K., Coenen G. and A. Warne (2010): "Forecasting with DSGE Models", *European Central Bank Working Paper*, n°1185, may.

[17] Clark, T. and M. McCracken (2001): "Tests of Equal Forecast Accuracy and Encompassing for Nested Models", *Journal of Econometrics*, **105**, 85-110.

[18] Clark, T. and M. McCracken (2010): "Testing for Unconditional Predictive Ability", Forthcoming: *Oxford Handbook on Economic Forecasting.*

[19] De Jong P. and P. Shephard (1995): "The Simulation Smoother for Time Series" *Biometrika* **82**, 339-350.

[20] Del Negro M. and F. Schorfheide (2004): "Priors from General Equilibrium Models for VARs", *International Economic Review*, **45**, 643.673.

[21] Del Negro M. and F. Schorfheide (2008): "Forming priors of DSGE Models (and how it Affects the Assessment of Nominal Rigidities)", *Journal of Monetary Economics* **55** 1191-1208.

[22] Del Negro M. and F. Schorfheide (2009): "Monetary Policy Analysis with Potentially Misspecified Models", *American Economic Review* **99**, 1415-1450.

[23] Del Negro M., F. Schorfheide, F. Smets and R. Wouters (2007): "On the Fit and Forecasting Performance of New Keynesian Models"; *Journal of Business and Economic Statistics* **23**, 123-143.

[24] Diebold F. X. and R. S. Mariano (1995): "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics* **13**, 253-263.

[25] Edge R. M. and R. S. Gürkaynak (2011): "How Useful are Estimated DSGE Model Forecasts?", *Federal Reserve Bank of St Louis Working Paper.*

[26] Forni, M., M. Hallin, M. Lippi and L. Reichlin (2000): "The generalized dynamic factor model: identification and estimation", *The Review of Economics and Statistics* **82**, 540-554.

[27] Forni, M., M. Hallin, M. Lippi and L. Reichlin (2005): "The generalized factor model: one-sided estimation and forecasting", *Journal of the American Statistical Association* **100**, 830-840.

[28] Forni M., D. Giannone, M. Lippi and L. Reichlin (2009): "Opening the Black Box: Structural Factor Models with Large Cross-Sections", *Econometric Theory* **25**, 1319-1347.

[29] Forni M. and M. Lippi (2001): "The generalized dynamic factor model: representation theory", *Econometric Theory* **17**, 1113-41.

[30] Gerali A., Neri S., Sessa L. and F. Signoretti (2008): "Credit and Banking in a DSGE model", mimeo.

[31] Geweke J. (1977): "The dynamic factor analysis of economic time series", in D.J. Aigner and A.S. Goldberger, Eds., *Latent Variables in Socio-Economic Models.* Amsterdam: North Holland.

[32] Giacomini R. and H. White (2006): "Tests of conditional predictive ability", *Econometrica* **74**, 1545-1578.

[33] Giannoni D., Reichlin L. and L. Sala (2006): "Monetary Policy in Real Time", *NBER Macroeconomics Annual 2004*, 161-200.

[34] Gregory A., Head A. and J. Raynauld (1997): "Measuring World Business Cycles" *International Economic Review*, **38**, 677-701.

[35] Hallin M. and R. Liška (2007): "Determining the Number of Factors in the Generalized Dynamic Factor Model", *Journal of the American Statistical Association* **102**, 603-617.

[36] Inklaar R.J., Jacobs, J. and W. Romp (2003): "Business Cycle Indexes: does a Heap of Data Help?", *Working Paper*, University of Groningen.

[37] Kocherlakota N. (2007): "Model Fit and Model Selection", Federal Reserve Bank of St. Louis.

[38] Kryshko M. (2009): "Data-Rich DSGE and Dynamic Factor Models", mimeo.

[39] Lippi M. and D.L. Thornton (2004): "A Dynamic Factor Analysis of the Response of U.S. Interest Rates to News", *Federal Reserve Bank of St Louis Working Paper*.

[40] Ludvigson S.C. and S.Ng (2007): "The Empirical Risk-Return Relation: A Factor Analysis Approach", *Journal of Financial Economics* **83**, 171-222.

[41] Neal P. and Roberts, G. (2008): "Optimal Scaling for Random Walk Metropolis on Spherically Constrained Target Densities", *Methodology and Computing in Applied Probability*, **10**, 277-297.

[42] Onatski A. (2009): "Testing Hypotheses About the Number of Factors in Large Factor Models", *Econometrica* **77**, 1447-1479.

[43] Paccagnini A. (2011): "Model Evaluation in the DSGE Approach: A Comparison of Hybrid Models", *European University Institute EUI, Working Paper Series*, November.

[44] Roberts G. O., A. Gelman and W. R. Gilks (1997): "Weak Convergence and Optimal Scaling of Randam Walk Metropolis Algorithms", *Annals of Applied Probability*, **7**, 110-120.

[45] Rubaszek M. and P. Skrzypczynski (2008): "On the Forecasting Performance of a Small-Scale DSGE Model", *International Journal of Forecasting*, **24**, 498-512.

[46] Sargent T.J. and C.A. Sims (1977): "Business Cycle Modelling without Pretending to have too much a Priori Economic Theory", in C.A. Sims, Ed., *New Methods in Business Research*, Minneapolis: Federal Reserve Bank of Minneapolis.

[47] Schorfheide F. (2010): "Estimation and Evaluation of DSGE Models: Progress and Challenges", University of Pennsylvania.

[48] Schorfheide F., Sill K., and M. Kryshko (2010): "DSGE Model-Based Forecasting of Non-Modelled Variables" *International Journal of Forecasting*, **26**, 348-373.

[49] Sims C. (2002): "Solving Linear Rational Expectations Models", *Computational Economics* **20**, 1-20.

[50] Smets F. and R. Wouters (2003): An Estimated Stochastic Dynamic General Equilibrium Model for the Euro Area, *Journal of the European Economic Association* **1**, 1123-1175.

[51] Smets F. and R. Wouters (2007): Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach, *American Economic Review* **97**, 586-606.

[52] Stock J.H. and M.W. Watson (2002a): "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business and Economic Statistics* **20**, 147-162.

[53] Stock J.H. and M.W. Watson (2002b): "Forecasting Using Principal Components from a Large Number of Predictors", *Journal of the American Statistical Association* **97**, 1167-1179.

[54] Stock J.H. and M.W. Watson (2005): "Implications of Dynamic Factor Models for VAR Analysis", *NBER Working Papers* no. 11467.

[55] Stock J.H. and M.W. Watson (2010): "Dynamic Factor Models", *Oxford Handbook of Economic Forecasting* chapter 2.

[56] Wang M. (2009): "Comparing DSGE model with factor model: An out-of-sample forecasting experiment", *Journal of Forecasting* **28**. 167-182.

[57] West K. D. (1996): "Asymptotic Inference about Predictive Ability", *Econometrica* **64**. 1067-1084.