Università degli Studi di Napoli "Federico II"

Corso di Dottorato in Scienze Economiche

XXVII ciclo

Latent factors in large economic datasets: forecasting and data analysis using factor models. An application to the insurance sector

Tutor
Prof. Marco Lippi
Coordinatore
Prof. Antonio Acconcia

Candidato
Donatella Albano

Anno Accademico 2015-2016

Contents

Acknowledgements vi

Introduction **1**

1 Dynamic Factor Models **5**

1.1 An overview 5

1.2 Examples of Large-Dimensional DFMs 10

1.2.1 Static Representation: SW, FHLR 11

1.2.2 Dynamic Representation, FHLZ 12

2 DFMs: Comparing forecasting performance using Euro Area data **15**

2.1 Forecasting using Dynamic Factor models 15

2.2 Data description 17

2.3 Calibration of the models 19

2.3.1 Calibration of SW 20

2.3.2 Calibration of FHLZ 22

2.3.3 Calibration of FHLR 24

2.4 Results 25

2.4.1 Euro Area Industrial Production and Inflation 25

2.4.2 Forecasting the whole dataset: focus on national results 28

2.5 Conclusions 29

3 DFMs: An application to the insurance sector **31**

3.1 Dynamic Factor models to gather informations from data . . . 31

3.2 Towards a new forecasting model for the insurance demand . . 32
3.2.1 The insurance sector: an overview 33
3.2.2 The determinants of the insurance demand 36
3.2.3 Data description 39
3.3 First results 40

A Appendix 43
A.1 Appendice I - Chapter 2 43
A.1.1 Dataset description 43
A.1.2 Tables 48
A.1.3 Figures 62
A.2 Appendice II - Chapter 3 73
A.2.1 Dataset description 73
A.2.2 Glossary of italian insurance terms 75
A.2.3 Figures 79

Bibliography 85

Aknoledgments

Ringrazio di cuore il mio tutor Marco Lippi, oltre ad essere un grande *professor*, anche un grande uomo dotato di enorme pazienza e capace di tener viva una scintilla di amore per la ricerca quando quel sogno sembrava ormai solo un vago ricordo.
Un ringraziamento speciale a Stefano Soccorsi e Alessandro Giovannelli per avermi introdotto e guidato nel mondo dei *factor models*.
Un ringraziamento all'ANIA, per la disponibilita' delle banche dati e informativa, in particolare a Silvia Salati, Annalaura Grasso e Edoardo Marullo Reedtz per le traduzioni e le revisioni 'tecniche' e a Mara De Marzi per il suo super super supporto a 360 gradi.
E un ringraziamento a mia sorella Carla per il *supporto linguistico* a circa 10.000 km di distanza...e anche al fuso orario.

Introduction

The fast increase, occurred in the last twenty years, of the amount of time series available to the econometricians for their analysis, has led to a rapid development of the literature related to the so-called *factor models*. The term *factor model* indicates a model that aim at extrapolating from large dataset a small number of *latent factors* which are able to summarize the properties of the entire panel and therefore explaining themselves the comovements of each dataset. Each variable in the dataset (x_{it}) can therefore be decomposed into the sum of a common (χ_{it}) and an idiosyncratic component (ξ_{it}).

Dynamic factor models (DFMs) exist in literature since the beginning of the past century. The first generations of factor models have been widely used in psychology and other disciplines of the social sciences but their success was moderate in economic analysis until recent years, perhaps because some assumptions on factors and errors did not match up well with economic data. In the late nineties the seminal works of Forni, Hallin, Lippi, and Reichlin 2000 (FHLR) and of Stock and Watson 2002b (SW) have proved that, as both N and T goes to infinity, factor models can be consistently estimated with the method of static or dynamic principal components even under the assumption that the correlation in the data is due also to other *non-pervasive* shocks. The main intuition of the so called *approximate* factor literature is that as the number of variables increases to infinity the common component survives to aggregation whereas the idiosyncratic component vanishes. Once the literature understood how to estimate them under general assumptions, they have become a standard tool in the macroeconomic literature. Forni, Hallin, Lippi, and Zaffaroni 2015 (FHLZ) improved their findings solving the problem of two-sideness of filter estimated by FHLR, allowing their models to be used for forecasting purposes.

The purpose of this thesis is to retrace the main steps that were taken in the evolution of the factor models and, in addition, to introduce two examples of how to apply the newest techniques developed in such fields to two different typologies of dataset, one *traditional*, meaning that it is composed mainly by macroeconomic and financial time series, and the other one including time series relevant to the Italian insurance sector and a set of macroeconomic and financial series related to them. The work is divided into three sections.

Chapter 1 goes back over the literary history of the Dynamic Factor Models (DFMs). The aim of this section is to give an overview of the intuition behind factor models, of their evolution over time and of the main class of large-dynamic factor models which will be used in the *Chapter 2* for forecasting purposes and model comparisons. Moreover it introduce the key 'critical points' to consider when approaching to *factor models*, namely, *model representation*, *estimation of the number of factor* and *estimation of the factor and of their loadings*.

Factor models can be used for different purposes and, in particular, to *i)* forecast o predict the variables of interest within the dataset; *ii)* extract information from data to analyze their behavior and properties; *iii)* analyze the effects of unexpected shock on the observed variables. In the following chapters the first two points are examined.

Chapter 2 presents an application of the factor models for forecasting purposes. More specifically, it compares the pseudo real-time forecast performances of three different factor models over a panel of macroeconomic and financial Euro Area variables. Until today, the literary works focused on the comparison between models hasn't reached a conclusion on which model has to be preferred, yet. Generally speaking, the results obtained so far leads to the acknowledgement that a better or worse performance of one model over another mainly depends on the structure of the data itself. The models herein examined are the following: *i)* the dynamic factor model recently proposed in Forni, Hallin, Lippi, and Zaffaroni 2015, FHLZ; *ii)* the model based on generalised principal components introduced in Forni, Hallin, Lippi, and Reichlin 2005, FHLR; *iii)* the factor model based on standard principal components proposed in Stock and Watson 2002a, SW.

These three models mainly differ for the methodology they use to estimate the latent factors and for their representation. SW suggests a model that estimates the factors through principal component analysis and assumes a finite dimension of the space spanned by the common components; FHLR 'extends' SW concept by moving the estimation of covariances, and therefore that of the factors, from the *time-domain* towards the *frequency-domain* (factors are estimated through the so called 'dynamic or generalised principal component approach'), obtaining then a factor model that is more general than the first one as it exploits the dynamic structure of the data, even though both have a static representation. FHLZ offers a more general approach that further 'extends' FHLR by allowing the space spanned by the common components to have infinite dimension and the common components themselves to have a rational spectral density. The estimators they provide for the loadings and the dynamic component solve the so called problem of two-sided filters encountered by FHLR.

Chapter 3, finally, aims at investigating the adaptability of factor models to a panel composed by the data relevant to the insurance industry (in particular, the Italian insurance sector) as well as by macroeconomic and financial variables that are supposed to be linked with it. Putting itself as a new addition of the econometric literature that works with this kind of models and, more generally, with this kind of approach, purpose of this chapter is to process the first analysis of the nature of the data available and of the structure of their dataset by using the factor model techniques. The idea presented in this Chapter has arisen with (and from) the professional experience started in March 2012, still ongoing, with the Economic and Finance Research Departments of the National Association of Insurance Companies (ANIA). It comes from the consideration that in the econometric academic environment there is still little knowledge of insurance sector data and their dynamics, especially if compared with the banking sector and, more generally, with the financial one. Neverthless, a progressive and continuous growth of the Insurance sector in the past few decades, has contributed to a remarkably development of the the econometric academic interest and that of institutions towards the sector in all countries. First results lead the door open to promising results after a further and more in-depth analysis.

Chapter 1

Dynamic Factor Models

1.1 An overview

The fast increase occurred in the last twenty years of the amount of time series available in several field of research, together with the sharp evolution of technology, has stimulated a considerable progress in the development of time series forecasting methods that exploit many predictors, leading to a rapid development of the literature related to the so-called *factor models*.

Roughly speaking, the term *factor model* indicates a model that aim at extrapolating from large dataset a small number of *latent factors* which are able to summarize the properties of the entire panel and therefore explaining themselves the comovements of the related dataset. Each variable in the dataset (x_{it}) can therefore be decomposed into the sum of a common (χ_{it}) and an idiosyncratic component (ξ_{it}). They are a powerful dimension reduction technique which is proven successfull in forecasting, in construction of business cycle indicators and inflation indexes[1], in structural analysis as well as in the analysis of financial markets (see Luciani 2014b for useful references).

[1] These are real-time application of factor models. Their goal is to extract the main 'signal' in the data while backing-out the 'noise'. Examples are *Eurocoin* developed by Altissimo et al. 2010 and the *core-inflation index* of Cristadoro, Forni, Reichlin, and Veronese 2005

The aim of this section is to give an overview of the intuition behind factor models, of their evolution over time and of the main class of large-dynamic factor models which will be used in the *Chapter 2* for forecasting purposes and model comparisons[2].

First of all, it is worth to notice that the use of the term *dynamic* can be sometimes confusing: *i)* it is 'informally' used to refer to the 'time dimension' present in the new generation of factor models (in the first generation they were traditionally developed mainly for cross-sectional data); *ii)* it refers to the methods for the *estimation of the factors* based on frequency-domain space (against those based on time-domain space and classical principal components, defined *statics methods*); finally, *iii)* it refers to the *representation of the model*, against the *static representation models*. In the following sections we will try to reduce terminology misinterpretations.

Dynamic factor models (DFMs) exist in literature since the beginning of the past century. Sargent, Sims, et al. 1977 and Geweke 1977 were amongst the first to apply the dynamic factor approach to macroeconomic analysis. The first proposed them as a time-series extension of factor models previously developed for cross-sectional data; the latter showed that two dynamic factors could explain a large fraction of the variance of important U.S. quarterly macroeconomic variables, including output, employment, and prices; after them, many other authors confirmed their findings.

The first generations of factor models have been widely used in psychology and other disciplines of the social sciences but their success was moderate in economic analysis until recent years, perhaps because some assumptions on factors and errors did not match up well with economic data. These models were estimable only on small databases as their estimation, for example, ruled out the possibility of sectorial or regional shocks driving the comovements, a common feature of large economic datasets. The assumption of *exact factor structure*, i.e. the hypothesis that the idiosyncratic components are cross-

[2]see lecture notes edited by M. Barigozzi, *Dynamic Factor Models*, December 2015 for a more comprehensive intuition and for the main analytical results.

sectionally uncorrelated, is unrealistic on large database where sectorial or regional shocks might affect groups of variables.

In the late nineties the seminal works of Forni, Hallin, Lippi, and Reichlin 2000 and Stock and Watson 2002b have proved that, as both N and T goes to infinity, factor models can be consistently estimated with the method of static or dynamic principal components even under the assumption that the correlation in the data is due also to other *non-pervasive* shocks. The main intuition of the so called *approximate* factor literature is that as the number of variables increases to infinity the *common component* survives to aggregation whereas the *idiosyncratic component* vanishes. Once the literature understood how to estimate them under general assumptions, they have become a standard tool in the macroeconomic literature. Forni, Hallin, Lippi, and Zaffaroni 2016 improved their findings solving the problem of two-sideness of filter estimated by FHLR, therefore allowing their models to be used for forecasting purposes.

The main econometric issue recent DFMs attempt to solve is the so called *curse of dimensionality problem*: if on the one hand one can benefit of a huge number of monthly or quarterly macroeconomic and financial time series, N, on the other hand the number of years for which data are reliable and relevant, call it T, is relatively small. Classical models are in fact usually not appropriate in these cases, as they imply the estimation of too many parameters. If the number of regressors is proportional to the sample size, in fact, the OLS forecasts are not first-order efficient, that is, they do not converge to the infeasible optimal forecast.

The key points to consider when approaching to *factor models* are the following:

1. Exact vs. approximate factor structure

A key aspect of many-predictor forecasting is *imposing enough structure* so that estimation error is controlled (is asymptotically negligible) yet useful information is still extracted. Said differently, the challenge of many-predictor forecasting is *to turn dimensionality from a curse into a blessing.* A first distinction is therefore between *exact* and *approximate representation.* In the *exact* model the idiosincratic component has no cross-sectional dependence, thus it has a diagonal covariance matrix, while in the *approximate* one it is allowed to have mild cross-sectional, thus a covariance matrix which is not necessarily diagonal. Chamberlain and Rothschild 1982 introduced a useful distinction between exact and approximate DFMs. The second case is more realistic but estimation of the model is possible only if a large cross-section is available, while the first case imposes a more restrictive assumption but estimation is possible even for few time series.

2. Static vs. dynamic representation models
Factor models are based on the idea that macroeconomic fluctuations are the result of a small number of macroeconomic or structural shocks, u_t, which affect the variables, and of a large number of sectorial or regional shocks, e_t that affect one or a few variables. Generally speaking, we can distinguish factor model depending on the *effect of the factors on the data*; a model is *static* or *dynamic* in this sense, according of whether the factors have only a contemporaneous effect on the data or through their lags too. Typically for the same time series the number of static factors r is bigger than the number of dynamical ones q.

3. Determining the number of factors
Once we choose our representation model, we need to determine the number of factors we wish to estimate. According to our model, several methods are available for estimating the number of static factors r or the number of dynamic factors q, for example:

- the number of static factors, r, can be determined by a combination of a-priori knowledge, visual inspection of a scree plot, and the use of information criteria developed by Bai and Ng 2002. They developed a family of estimators of r that are motivated by information criteria used in model selection. Information criteria trade off the benefit of including an additional factor (or, more generally, an additional parameter in a model) against the cost of increased sampling variability arising from estimating another parameter.

- the number of dynamic factors, q, can be estimated by:

  i) a frequency-domain procedure proposed by Hallin, Liska, et al. 2007 based on the observation that the rank of the spectrum of the common component of X_t is q;

  ii) a method proposed by Bai and Ng 2002 based on the observation that the innovation variance matrix in the population VAR has rank q

  iii) a method, proposed by Amenguel and WatsonâĂŹs (2007) based on the observation that, in a regression of X_t on past values of Ft, the residuals have a factor structure with rank q.

4. 'Static' vs. 'Dynamic' estimation of factors
Here the term 'static' refers to *time-domain* based estimation techniques (see section 1.2.1, using *static principal components*), whereas 'dynamic' refers to the *frequency-domain* ones (see section 1.2.2). Once the factor are estimated, the second step is to estimate the filters, i.e the *low of motion for the factors* and that for the idiosyncratic components[3]. FHLR prove the consistency, and provide rates of convergence, of the common component estimated by dynamic principal components. Their method for estimation of factors by means of dynamic principal components requires two-sided smoothing, so estimates of the factors at the end of the sample are not available. Forni, Lippi, and Reichlin 2004 estimate factors using a two step approach based on dynamical (generalized) principal components in which observations are weighted according to their signal to noise ratio and on the imposing con-

[3]The seminal work of Geweke (1977) and Sargent and Sims (1977) only used frequency domain methods to look for evidence of a dynamic factor structure and to estimate the importance of the factor (they do not estimate factors directly and thus could not be used for forecasting).

straints implied by the dynamic factors structure in the projection of the variable of interest on the common factors.

1.2 Examples of Large-Dimensional DFMs

An important contribution of Forni, Hallin, Lippi, and Reichlin 2000 and Forni, Hallin, Lippi, and Zaffaroni 2016 is the formulation of the following *general form* of DFMs:

$$x_{it} = \chi_{it} + \xi_{it} = \frac{c_{i1}(L)}{d_{i1}(L)} u_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)} u_{2t} + \cdots + \frac{c_{iq}(L)}{d_{iq}(L)} u_{qt} + \xi_{it}, \tag{1.1}$$

where L is the lag operator, $t = \in \mathbb{Z}$, $i \in \mathbb{N}$,

$$c_{if}(L) = c_{if,0} + c_{if,1}L + \ldots + c_{if,s_1}L^{s_1}, \quad d_{if}(L) = d_{if,0} + d_{if,1}L + \ldots + d_{if,s_2}L^{s_2},$$

and $\mathbf{u}_t = (u_{1t}\ u_{2t}\ \cdots\ u_{qt})'$ is a q-dimensional orthonormal white noise.
The processes χ_{it} reprsents the *common components* driven by the *common shocks* $\mathbf{u}_t$, also called the *dynamic (common) factors*, while the processes ξ_{it} represents the *idiosyncratic components*.
Assumptions underlying 1.1 are:

i) the polynomials $d_{if}(L)$ are stable so that χ_{it} is stationary and is co-stationary with χ_{jt} for all $i, j \in \mathbb{N}$.

ii) ξ_{it} is stationary and co-stationary with ξ_{jt} for all $i, j \in \mathbb{N}$.

iii) ξ_{it} and $\mathbf{u}_t$ are orthogonal for all $i \in \mathbb{N}$ so that ξ_{it} and χ_{jt} are orthogonal for all $i, j \in \mathbb{N}$.

The assumptions above trivially imply that the observable process x_{it} is stationary and costationary with x_{jt}, for all $i, j \in \mathbb{N}$.
Assumptions on the eigenvalues of the spectral density of the vector processes (see Forni, Hallin, Lippi, and Reichlin 2000, Forni, Hallin, Lippi, and

Zaffaroni 2015 for details) imply that the common shocks and the common components (and therefore the idiosyncratic components)

$$\boldsymbol{\chi}_{nt} = (\chi_{1t}\ \chi_{2t}\ \cdots\ \chi_{nt})', \quad \boldsymbol{\xi}_{nt} = (\xi_{1t}\ \xi_{2t}\ \cdots\ \xi_{nt})',$$

can be recovered as limits of linear combinations of the first n observables x_{it}, as n tends to infinity.

The power of this representation is also in the fact that, imposing some restriction on the assumption, it is possible to derive the static form representation and to estimate static factors.

1.2.1 Static Representation: SW, FHLR

Suppose that for a given $\bar{t}$ the common components

$$\chi_{i\bar{t}} = \frac{c_{i1}(L)}{d_{i1}(L)} u_{1\bar{t}} + \frac{c_{i2}(L)}{d_{i2}(L)} u_{2\bar{t}} + \cdots + \frac{c_{iq}(L)}{d_{iq}(L)} u_{q\bar{t}}, \qquad i \in \mathbb{N},$$

span a finite-dimensional vector space $S_{\bar{t}}$ and denote by r its dimension. Stationarity of the common and idiosyncratic components implies that the same occurs for all $t \in \mathbb{Z}$, that the dimension of S_t is independent of t and there exists a 'stationary basis'

$$\mathbf{F}_t = (F_{1t}\ F_{2t}\ \cdots\ F_{rt})$$

such that (1.1) can be rewritten in the so called *static form*

$$x_{it} = \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \cdots + \lambda_{ir}F_{rt} + \xi_{it}, \tag{1.2}$$

Moreover, $r \geq q$, i.e. the number of the so-called *static factors* F_{ht} is at least equal to the number of dynamic factors.

Under the finite-dimension assumption, the static factors F_{jt} and the loadings λ_{ij} can be estimated using the first r standard principal components, or generalized principal components as in Forni, Lippi, and Reichlin 2004, of the first n observables x_{it}. These two method for estimating factors differs basically in that, while standard principal components are based on the covariances of the observables, $\mathbf{\Gamma}_0^x$, in dynamical principal components the covariances of the common and idiosyncratic components $\mathbf{\Gamma}_0^\chi$ and $\mathbf{\Gamma}_0^\xi$ are employed to estimate a basis of the factor space by means of generalized principal components. This means that the latter involves the following procedures to estimate covariances:

I) Estimation of the spectral density matrix of the observables x's

$$\hat{\mathbf{\Sigma}}^x(\theta) = \frac{1}{2\pi} \sum_{k=-M}^{M} e^{-ik\theta} w_k \hat{\mathbf{\Gamma}}_k$$

where w_k are the weights of a kernel function;

II) Computation of the spectral density matrix of the common compoments, $\hat{\mathbf{\Sigma}}^\chi(\theta)$, by means of the first q frequency-domain principal components of $\hat{\mathbf{\Sigma}}^x(\theta)$;

III) Computation of the autocovariance matrices of the common components $\hat{\mathbf{\Gamma}}_k^\chi = \int_{-\pi}^{\pi} e^{ik\theta} \hat{\mathbf{\Sigma}}^\chi(\theta)\, d\theta$.

The estimated factors can then be used, in both cases, as predictors in forecasting the variables x_{it}. Predictions based on the standard PCA method (as in SW) are referred to as the *static method of forecasting* (see section 2.6 for details), whereas those based on the generalized (or dynamic) PCA method (as in FHLR), are referred to as *static method of forecasting with frequency-domain estimation of factors* (see section 2.10 for details).

1.2.2 Dynamic Representation, FHLZ

The main adding of (1.1) (studied in Forni, Hallin, Lippi, Zaffaroni, et al. 2011, Forni, Hallin, Lippi, and Zaffaroni 2015, Forni, Hallin, Lippi, and Zaffaroni 2016) to the literature is that it allows the common components space to have infinite dimension. Under this assumption, a finite number of common shocks drives the common components, though the common components themselves are allowed to span an infinite-dimensional space. The estimator based on dynamic principal components in FHLR cannot provide an estimator of the common dynamic factors and moreover it is likely to deliver

a common component built using two-sided filters. A common component estimated in this way does not allow to run forecast and impulse response analysis, i.e. the study of the impact of unexpected shocks on observed variables. FHLZ (2015,2016) shows how to obtain one-sided estimators without the finite-dimension assumption imposing the weaker condition that the common components have a *rational spectral density*, that is, filter in (1.1) are ratios of polynomials in L. It provides consistent estimators for the loadings $\frac{c_{if}(L)}{d_{if}(L)}$ and the dynamic factors u_{ft}.
The basic result used in FHLZ is that the vector

$$\boldsymbol{\chi}_t = (\chi_{1t}\ \chi_{2t}\ \cdots\ \chi_{nt},\ \cdots)',$$

which is an infinite (or large) dimensional vector driven by a finite (relatively small) number of shocks, has, under fairly general conditions, a blockwise autoregressive representation of the form

$$\begin{pmatrix} \mathbf{A}^1(L) & 0 & \cdots & 0 & \cdots \\ 0 & \mathbf{A}^2(L) & \cdots & 0 & \\ & & \ddots & & \\ 0 & 0 & \cdots & \mathbf{A}^k(L) & \\ \vdots & & & & \ddots \end{pmatrix} \boldsymbol{\chi}_t = \begin{pmatrix} \mathbf{R}^1 \\ \mathbf{R}^2 \\ \vdots \\ \mathbf{R}^k \\ \vdots \end{pmatrix} \mathbf{u}_t, \tag{1.3}$$

where $\mathbf{A}^k(L)$ is a $(q+1)\times(q+1)$ polynomial matrix *with finite degree* and $\mathbf{R}^k$ is $(q+1)\times q$. See Forni, Hallin, Lippi, and Zaffaroni 2015.
Denoting by $\mathbf{A}(L)$ and $\mathbf{R}$ the (infinite) matrices on the left- and right-hand sides of (1.3), using $\boldsymbol{\chi}_t = \mathbf{x}_t - \boldsymbol{\xi}_t$, and setting $\mathbf{Z}_t = \mathbf{A}(L)\mathbf{x}_t$, we get:

$$\mathbf{Z}_t = \mathbf{R}\mathbf{u}_t + \mathbf{A}(L)\boldsymbol{\xi}_t. \tag{1.4}$$

Instead of estimating a basis (of dimension r, the number of static factors) of the factor space by means of generalized principal components, the covariances $\hat{\boldsymbol{\Gamma}}_k^\chi$ are used to compute the VAR matrices $\hat{\mathbf{A}}^k(L)$ and, finally, the shocks $\hat{\mathbf{u}}_t$ and the matrices $\hat{\mathbf{R}}^k$ are obtained by means of standard principal components of the estimated variables $\hat{\mathbf{Z}}_t$.

The estimated factors can then be used as predictors in forecasting the variables x_{it}. Predictions based on (1.1) are referred to as the *dynamic method of forecasting*. See section 2.3.2 for details.

Chapter 2

DFMs: Comparing forecasting performance using Euro Area data

2.1 Forecasting using Dynamic Factor models

As seen in *Chapter* 1 the past decade has seen considerable progress in the development of time series analysis and forecasting methods that exploit many predictors. We have also seen that methods based on dynamic factor models have gained huge importance thanks to their capacity to exploit the comovements among a large number of economic variables and to treat them as arising from a small number of unobserved sources, the *factors*. One of the main objectives of this class of model is forecasting: in a dynamic factor model, the factors estimated (which become increasingly precise as the number of series increases) can be used to forecast individual economic variables.

Section 1.2 shows that the peculiarity of Large-Dimensional Dynamic Factor Models (*DFMs*) is that they represent each variable in the dataset as decomposed into a *common component*, driven by a small (as compared to the number of series in the dataset) number of *common factors* and an *idiosyncratic component* assumed to be orthogonal across different variables or only weakly correlated so that the covariance of the variables is mostly accounted for by the common components).

The literature comparing model forecasting performances, either with simulated or experimental data, has reached mixed conclusions so far. The use of real data, for example, clearly stress the fact that strong variations in the covariance structure of the dataset can affect the relative performances of the models depending on their robustness in situation of instability.

In comparing SW and FHLR using US data, Boivin and Ng 2005 found that SW generally outperforms FHLR, whereas D Agostino and Giannone 2012 found the two methods to perform equally well in their sample even if different performances are found in subsamples, e.g. the dynamic method fares better during the Great Moderation. Schumacher 2007, using German data, finds that frequency-domain methods based on generalised principal components provides more accurate forecasts of the GDP.

A similar result is obtained in Reijer 2005 with Dutch macroeconomic data. Recently Forni, Giovannelli, Lippi, and Soccorsi 2016 (FGLS) extended the comparison in Boivin and Ng 2005 and D Agostino and Giannone 2012 to more recent US data and include the new FHLZ forecasting dynamic factor model (see section 1.2.2). They use a dataset of US macroeconomic and financial monthly time series spanning from January 1959 to August 2014 thus including the Great Moderation, the Great Recession and the subsequent recovery. FGLS has produced the first systematic comparison of FHLZ with SW and FHLR for US monthly data[1]

The aim of this section is to replicate FGLS using a dataset of macroeconomics and financial Euro Area monthly time series, grouped in 11 categories and spanning the period from January 1986 to October 2015. The period thus includes the Great Moderation, the Great Recession - more specifically its spillover in the Euro Area - and the more recent Euro Area Sovereign Debt

[1]Before this, Forni, Hallin, Lippi, and Zaffaroni 2016 compared forecasts obtained with SW and FHLZ using simulated data and quarterly macroeconomic US data.

crisis in 2012. Target variables are Euro Area (log of) Industrial Production and annual Inflation rate.
As in FGLS, the selected models are the following:

i) (SW) The standard (static) principal-component model introduced by Stock and Watson in (Stock and Watson 2002a).

ii) (FHLR) The (static) model based on generalized principal components introduced by Forni, Hallin, Lippi and Reichlin in Forni, Hallin, Lippi, and Reichlin 2000 as a variant of the previous model, in which the covariances of the common and idiosyncratic component are estimated using a frequency-domain method.

iii) (FHLZ) The new (dynamic) model recently proposed by Forni, Hallin, Lippi and Zaffaroni in Forni, Hallin, Lippi, and Zaffaroni 2015), based on frequency-domain method for the estimation of the covariances, like in FHLR, but in which, differently from the static representation methods i) and ii) the common components themselves are allowed to span an infinite-dimensional space. The dynamic relationship between the variables and the factors in this model is more general as compared to *i)* and *ii)*.

2.2 Data description

The dataset consists of 176 Euro Area macroeconomic and financial time series observed at monthly frequency between January 1985 and October 2015. Data therefore include the Great Moderation, the Great Recession originated from the 2007 financial crisis and its spillover effect in the Euro Area from the second quarter of 2008 to the second quarter of 2009. It also includes the so called Euro Area Sovereign Debt crisis in 2012. The series are grouped into 11 main categories and each of them mainly consists of Euro Area aggregate series and country-specific series related to each of the main EA countries (see Appendix, A.1 for details).

To achieve stationarity the series are transformed into first difference of the logarithm (mainly real variables and stock prices), first difference of yearly difference of the logarithm (prices) and monthly difference (interest rates, surveys) and, if needed, deseasonalized. No treatment for outliers is applied. Therefore, let

$$\mathbf{X_t} = (X_{1t},\ X_{2t},\ \ldots,\ X_{nt})'$$

be the raw dataset, and

$$\mathbf{Z}_t = (Z_{1t},\ Z_{2t},\ \ldots,\ Z_{nt})' \tag{2.1}$$

its stationary version after the transformations are applied; $\hat{Z}_{i,t+h|t}$ are its forecasts computed for $h = 1, 3, 6, 12, 24$ months ahead. Following FGLS, the target at time $t + h$ is therefore

$$W_{i,t+h|t} = Z_{i,t+1} + \cdots + Z_{i,t+h},$$

which for our main variable of interest[2] is:

$$\begin{cases} W_{IP,t+h|t} = \log IP_{t+h} - \log IP_t, & \text{for i=IP;} \\ W_{CPI,t+h|t} = (1 - L^{12}) \log CPI_{t+h} - (1 - L^{12}) \log CPI_t., & \text{for i=CPI.} \end{cases}$$

In both cases the forecast is then;

$$\hat{W}_{i,t+h|t} = \hat{Z}_{i,t+1|t} + \cdots + \hat{Z}_{i,t+h|t} \tag{2.2}$$

and the prediction error, normalized for the horizon's length, is:

$$FE_{1,t,h} = \frac{1}{h}(\hat{W}_{1,t+h|t} - W_{1,t+h}) = \frac{1}{h}\Big((\hat{Z}_{1,t+1|t} - Z_{1,t+1}) + \cdots + (\hat{Z}_{1,t+h|t} - Z_{1,t+h})\Big). \tag{2.3}$$

As in FGLS, the involved procedure is the following one:

[2]Industrial Production is transformed into first difference of the logarithm while Consumer Price into first difference of yearly difference of the logarithm.

I) The sample is split in a *calibration pre-sample* and a *sample proper* for the model comparison

II) All the models i), ii), iii) are estimated for each t of a *rolling ten-year window* $[t-119,\ t]$ and forecasts are computed.

III) The pre-sample period is used to decide which method should be used to determine the number of factors, the number of lags of the factors or of the variable to be predicted, etc.

IV) The selected specification of the parameters are then used in the sample proper to get forecasts and comparison using 2.3.

2.3 Calibration of the models

Following the forecasting exercise metholodogy of FGLS, the sample is split into a *calibration pre-sample* and the *sample proper* (I). In this exercise the pre-sample spans the period from February 1986 to December 2000, sample proper from January 2001 to October 2015. For all four methods we use a *rolling ten-year window* $[t-119,\ t]$, and the models are re-estimated for each t (II). For each predictive model, the pre-sample forecasting performance is evaluated by its mean square forecast error (MSFE), which is defined as follows:

$$\mathrm{M}SFE^{m}_{i,h} = \frac{1}{(T_1-h)-T_0+1}\sum_{\tau=T_0}^{T_1-h}\mathrm{F}E^2_{i,\tau,h}, \tag{2.4}$$

where (i) T_0 and T_1 denote the first and the last dates of the sample, (ii) the superscript m stands for the model used and ranges over SW, FHLR, FHLZ, AR. Replacing the limits of the summation in (2.4) with any time interval within the sample we can measure local forecasting performances.
To compare specifications m_1 and m_2 of method m at different horizons we compute the ratio between the $MSFEs$

$$\mathrm{RMSFE}^{m_1/m_2}_{i,h} = \frac{\mathrm{MSFE}^{m_1}_{i,h}}{\mathrm{MSFE}^{m_2}_{i,h}}. \tag{2.5}$$

When no specification prevails uniformly across different horizons, we choose according to the average of the ratio (2.5) over all five horizons (III). The calibration procedure is restricted to aggregate Euro Area industrial production, $\mathrm{IP}_t = X_{93,t}$, and consumer price, $\mathrm{CPI}_t = X_{69,t}$ (see Appendix, Table A.3).

2.3.1 Calibration of SW

As descripted in Section 1.2.1, given N, the number of series available, and T, the number of observations for each series, the factors are estimated by means of the standard Principal Components of the variables in the dataset.
Let Z be the dataset after transformation (see 2.1), the SW forecasting equation for z_{it} (Z_{it} after standardization[3] is obtained by projecting $z_{i,t+h}$ on the space spanned by the factor, their lags and the lagged value of the dependent variable:

$$\hat{\mathbf{F}}_t,\ \hat{\mathbf{F}}_{t-1},\ldots,\ \hat{\mathbf{F}}_{t-g_1};\ z_{i,t},\ z_{i,t-1},\ldots,\ z_{i,t-g_2}$$

where g_{i1} denotes the number of lags for the factor and g_{i2} is the number of lags for the dependent variable. The equation to be estimated is therefore:

$$z^{SW}_{i,t+h|t} = \boldsymbol{\alpha}_i(L)\hat{\mathbf{F}}_t + \beta_i(L)z_{i,t}, \tag{2.6}$$

where $\boldsymbol{\alpha}_i(L)$ is a $1\times r$ matrix polynomial of degree g_{i1} and $\beta_i(L)$ a scalar polynomial of degree g_{i2}[4].
Estimation of equation (2.6) requires the following steps:

1. determining for each t of the rolling window the number of static factors r.
2. estimating the covariance matrix of $\mathbf{z}_{nt} = (z_{1t}\ z_{2t}\ \cdots\ z_{nt})$, $\hat{\mathbf{\Gamma}}_n$
3. calculating the the first r principal components of $\mathbf{z}_{nt}$, defined as

$$\hat{\mathbf{F}}_t = \left(\widehat{F}_{1t},\ \widehat{F}_{2t},\ \ldots,\ \widehat{F}_{rt}\right) = \mathbf{P}^r\mathbf{z}_{nt}$$

where:

$$\hat{\mathbf{P}}_{nh}\mathbf{z}_{nt} = \mathbf{P}_{nh}(z_{1t}\ z_{2t}\ \cdots\ z_{nt})',$$

for $h = 1, 2, \ldots, r$, $\hat{\mathbf{P}}_{nh}$ is the eigenvector corresponding to the h-th eigenvalue (in decreasing order) of $\hat{\mathbf{\Gamma}}_n$.

[3]mean an standard deviation are added back after calculation

[4]The presence of the terms $z_{i,t-k}$ can be motivated as possibly capturing autocorrelation in the idiosyncratic component ξ_{it}

The parameter to calibrate are therefore:

(i) the number r of static factors,

(ii) the maximum lag g_{i1} for $\boldsymbol{\alpha}_i(L)$,

(iii) the maximum lag g_{i2} for $\beta_i(L)$.

The number r of static factors is estimated according to Bai and Ng's criterion IC_2 (Bai and Ng 2002) at every t (*Case 1*) or is selected between 1 and 8 and kept fixed as the window moves in the pre-sample (*Case 2*). In both cases the models are estimated through the following steps:

S1 - No lags allowed for the factors or the variable to be predicted: the prediction equation is (2.6) with $\boldsymbol{\alpha}_i(L)$ of degree zero and $\beta_i(L) = 0$.
Ratios $\mathrm{RMSFE}_{i,h}^{m_1/m_2}$ are computed, where: (1) m_2 is *Case 2* with r equal to 7, (2) m_1 is either *Case 1* or *Case 2* with $r = 1, \ldots, 8$, (3) $i = 93$ (IP) or $i = 69$ (CPI), (4) $h = 1, 3, 6,\ 12,\ 24$. The results are reported in Table A.5, *Panel SW:S1*. We see that the best models are: (I) *Case 2* with $r = 5$ for IP with $r = 7$ very close, (II) *Case 2* with $r = 7$ for CPI, with $r = 8$ very close. The two best models are denoted by $\mathrm{SW}_{IP}^{0,0}(5)$ and $\mathrm{SW}_{CPI}^{0,0}(7)$ respectively (the superscript indicates the number of lags for the predicted variable and the factors, respectively).

S2 - Lags allowed for the predicted variable, no lags allowed for the factors: the prediction equation is (2.6) with $\boldsymbol{\alpha}_i(L)$ of degree zero and the order of the polynomial $\beta_i(L)$ determined by the AIC or BIC criterion.
Prediction equation is run with $r = 5$, $r = 7$ for IP and CPI respectvely, augmented with lags for the predicted variable. The degree of $\beta_i(L)$ is determined both by the AIC and BIC criteria setting the maximum number of lags to 6, the benchmark being $\mathrm{SW}_{IP}^{0,0}(5)$ for IP and $\mathrm{SW}_{CPI}^{0,0}(7)$ for CPI. The results, reported in the *Panel SW: S2 of Table* A.5, show that for both IP and CPI the best result is obtained using the above specifications with no lags for both the predicted variable and the factors.

S3 - Lags allowed for the factors, no lags allowed for the predicted variable: the prediction equation is (2.6) with the degree of the vector polynomial $\boldsymbol{\alpha}_i(L)$ determined by the AIC and the BIC criteria and $\beta_i(L) = 0$.

The models ${\rm SW}_{IP}{}^{0,0}(5)$ for IP and ${\rm SW}_{CPI}{}^{0,0}(7)$ for CPI augmented with lags of the factors are run. The degree of $\boldsymbol{\alpha}_i(L)$ is determined by the AIC and the BIC criteria setting the maximum number of lags to 6. Again, ${\rm SW}_{IP}{}^{0,0}(5)$ for IP and ${\rm SW}_{CPI}{}^{0,0}(7)$ for CPI are confirmed to be the best choice (see Table A.6 Panel SW:*S3* for the results).

S4 - Lags allowed for both the factors and the variable to be predicted: the prediction equation is (2.6) with the degree of $\boldsymbol{\alpha}_i(L)$ and the order of the polynomial $\beta_i(L)$ determined by the AIC or BIC criterion.
The models ${\rm SW}^{0,0}_{IP}(5)$ and ${\rm SW}^{0,0}_{CPI}(7)$ are augmented with both lags of the factors and of the predicted variable. The results are very poor (*see Table* A.6 *Panel SW: S4*).

Like in FGLS, no evidence is found that lags in the factors and in the predicted variable, in addition to the factors at t, do help predicting ${\rm CPI}_{t+h}$ or ${\rm IP}_{t+h}$, for $h = 1, 3, 6,\ 12,\ 24$, in the pre-sample period.

In conclusion, our exploration of the space of possible SW specifications points to ${\rm SW}^{0,0}_{IP}(5)$ and ${\rm SW}^{0,0}_{CPI}(7)$ as good models for IP and CPI respectively.

2.3.2 Calibration of FHLZ

As seen in *Chapter 1*, the basic result underlying FHLZ is that the vector of the common components in equation 1.1 has, under fairly general conditions, a block-wise autoregressive representation of the form:

$$\mathbf{A}(L)\boldsymbol{\chi}_t = \mathbf{R}\mathbf{u}_t \tag{2.7}$$

After estimation of $\hat{\mathbf{A}}(L)$ (see 1.2.2), we invert it in (2.7):

$$\hat{\boldsymbol{\chi}}_t = \left[\hat{\mathbf{A}}(L)\right]^{-1}\hat{\mathbf{R}}\hat{\mathbf{u}}_t = \widehat{\mathbf{W}}(L)\hat{\mathbf{u}}_t = \widehat{\mathbf{W}}_0\hat{\mathbf{u}}_t + \widehat{\mathbf{W}}_1\hat{\mathbf{u}}_{t-1} + \cdots , \tag{2.8}$$

where $\hat{\boldsymbol{\chi}}_t$ is n-dimensional, and the matrices $\hat{\mathbf{A}}(L)$, $\hat{\mathbf{R}}$ and $\widehat{\mathbf{W}}(L)$ are $n \times n$, and the resulting prediction equation is:

$$z^{FHLZ}_{t+h|t} = \chi^{FHLZ}_{t+h|t} = \widehat{\mathbf{W}}_h\hat{\mathbf{u}}_t + \widehat{\mathbf{W}}_{h+1}\hat{\mathbf{u}}_{t-1} + \cdots \tag{2.9}$$

where:

$$\widehat{\mathbf{W}}(L) = [\hat{\mathbf{A}}(L)]^{-1}\hat{\mathbf{R}}$$

Estimation of equation (2.9) requires the following steps:

1. determining for each t of the rolling window the number of dynamic factors q.

2. estimating the covariance matrix of the observables in order to compute that of the common component χ and of the idiosyncratic one ξ,

3. estimating the matrix polynomials $\hat{\mathbf{A}}(L)$ of dimension $(q+1) \times (q+1)$,

4. computing $\mathbf{R}^k$, of dumension $(q+1) \times q$ and the common shocks u_t using (2.7).

The calibrated parameter are therefore:

(i) the kernel and the lag window for the estimation of the spectral density of the observable variables $\mathbf{\Sigma}^x(\theta)$,

(ii) the number q of dynamic factors,

(iii) the maximum lag K and the order selection criteria for the matrix polynomials $\mathbf{A}^k(L)$.

The predictor based on FHLZ depends on the order of the variables in the dataset, therefore several predictors are produced by reordering the dataset and the final predictor used is the average of them. The number of permutation for the reordering of the variables of the dataset is set to $N_{per} = 100$ (following results in FGLS) and the N_{per} permutations are produced using the matlab command *randsample* with a pre-defined random number generator. The model is then estimated in the following steps:

S1 - Selection of the lag order criterion for the $(q+1)$*-dimensional VAR's*: taking as benchmark the model using the BIC criterion, keeping fix the maximum lag order k=3, Gaussian Kernel and bandwidth w = 30 (the bandwidth corresponding to the ten-year window), and q determined at each t by means of the Hallin-Liška criterion, the AIC criterion shows some advantage for IP while the BIC criterion does it for CPI. See *Table A.7*, *Panel FHLZ: S1* for details. The two best models are denoted by $\text{FHLZ}_{IP}^{AIC,3}(Gauss, 30)$ and $\text{FHLZ}_{CPI}^{BIC,3}(Gauss, 30)$ respectively.

S2 - Different maxima for the maximum lag in the lag order criteria tried, from 3 *to* 7: the specification for the maximum lag order gives a flatness in the results. A further comparison, which is not shown in the Appendix, suggest the choice for K=6 for IP e K=3 for CPI. The selected model at this stage are, therefore, $\text{FHLZ}_{IP}^{AIC,6}(Gauss, 30)$ and $\text{FHLZ}_{CPI}^{BIC,3}(Gauss, 30)$.

S3 - Selection of the bandwidth for the estimation of the spectral density of the observable vector: different values for the bandwidth, 25, 35 and 40, are tried using as a benchmark the model selected at the previuos stage.(step), leading to a choice for w=25 (see *Table A.7, Panel FHLZ: S3*). The selected model at this stage are, therefore, $\text{FHLZ}_{IP}^{AIC,6}(Gauss, 25)$ and $\text{FHLZ}_{CPI}^{BIC,3}(Gauss, 25)$.

S4 - Selection of the kernel for the estimation of the spectral density of the observable vector: finally a comparison between the last selected model and the ones using Triangular Kernel (keeping fix all the other specifications) is run. The Gaussian Kernel is confirmed to be the best choice.

Selected models are: $\text{FHLZ}_{IP}^{AIC,6}(Gauss, 25)$ and $\text{FHLZ}_{CPI}^{BIC,3}(Gauss, 25)$.

2.3.3 Calibration of FHLR

Unlike FHLZ, FHLR assumes that the space spanned by the common components has finite dimension r (see Section 1.2.1) but unlike SW, instead of using the standard principal components which are based on the covariances $\mathbf{\Gamma}_0^x$, the covariances of the common and the idiosyncratic components Γ_0^χ and $\mathbf{\Gamma}_0^\xi$ are estimated using a frequency-domain method, that is, the estimated variance of the idiosyncratic is taken into account. Factors are estimated by means of *Generalized Principal Components*:

$$\widehat{\mathbf{G}}_t = \left(\widehat{G}_{1t},\ \widehat{G}_{2t},\ \ldots,\ \widehat{G}_{rt}\right) = \mathbf{P}^{G,r}\mathbf{x}_{nt}$$

where $\mathbf{P}^{G,r}$ is $n \times r$ and has the eigenvectors associated with the first r generalized eigenvalues of $\left(\mathbf{\Gamma}_0^\chi,\ \mathbf{\Gamma}_0^\xi\right)$ on the columns. The covariances $\mathbf{\Gamma}_h^\chi$ and $\mathbf{\Gamma}_h^\xi$ are then employed to project $\chi_{i,t+h}$ on the factors.
The prediction equation is:

$$z_{i,t+h|t}^{FHLR} = \chi_{i,t+h|t}^{FHLR} = \boldsymbol{\gamma}_h \widehat{\mathbf{G}}_t, \tag{2.10}$$

with

$$\boldsymbol{\gamma}_h = \widehat{\boldsymbol{\Gamma}}_{\boldsymbol{h}}^{\boldsymbol{\chi}} \widehat{\boldsymbol{z}}^{\boldsymbol{g}\prime} \left(\widehat{\boldsymbol{z}}^g \widehat{\boldsymbol{\Gamma}}_{\boldsymbol{0}} \widehat{\boldsymbol{z}}^{\boldsymbol{g}\prime} \right)^{-1}$$

Estimation therefore requires determining:

i) the number of dynamic factors q (like in FHLZ),

ii) kernel and lag window for the estimation of $\boldsymbol{\Sigma}^x(\theta)$ (like in FHLZ),

iii) the number r of static factors (like in SW).

The model is then calibrated in the following steps:
S0 - Selection of the number of static (r) and dynamic (q) factors: the number r of static factors is estimated according to Bai and Ng's criterion IC_2 (Bai and Ng 2002) at every t (*Case 1*), while the number q of dynamic factors is determined at each t by means of the Hallin-Liška criterion.

S1 - Selection of the kernel for the estimation of the spectral density of the observable vector: a comparison between the model using Triangular Kernel and Gaussian Kernel is run, fixing the bandwith at $W = 30$. The Gaussian Kernel is found to be the best choice, ie the selected models are: $\text{FHLR}_{IP}(Gauss, 30)$ and $\text{FHLR}_{CPI}(Gauss, 30)$. *See Table A.8, Panel FHLZ: S1*

S2 - Selection of the bandwidth for the estimation of the spectral density of the observable vector: different values for the bandwidth, 25, 35 and 40, are tried using as a benchmark the model using $W = 30$ and the Gaussian Kernel selected at the previous stage. Table A.8, Panel FHLR: *S2*) shows some advantages in using $W = 35$ for IP and $W = 25$ for CPI.

In conclusion, our exploration of the space of possible FHLR specifications points to $\text{FHLR}_{IP}(Gauss, 35)$ and $\text{FHLR}_{CPI}(Gauss, 25)$.

2.4 Results

2.4.1 Euro Area Industrial Production and Inflation

After the selection of the parameters in the pre-sample calibration exercise, the performances of the factor models over the proper-sample (form February 2001) are compared in the prediction of the target variables IP and CPI. The ten years from January 1991 to December 2001 are used to produce the first forecasts within the sample. Thus we start by predicting February 2001, April 2001, July 2001, January 2002, January 2003, for $h = 1,\ 3,\ 6,\ 12,\ 24$ respectively. The last forecast is October 2015 for all horizons.

As in the calibration exercise, for each predictive model, the proper-sample forecasting performance is evaluated by its mean square forecast error ($MSFE$) and results are compared using (2.5) for Euro Area and country-specific Industrial Production and Inflation and for disaggregate real and nominal variables. The common benchmark for the factor models is the univariate AR. *Table A.9 and A.10* report the performance for $h = 1$, 3, 6, 12, 24 , measured by the RMSFE(2.5), of the three factor models relative to AR for our main variables of interest, namely Euroa Area IP and CPI. We give results for the *Great Moderation*, or *pre-crisis*, from January 2001 to March 2008, the beginning of the Great Recession in the Euro Area, *Panel A*, and the *full sample period*, from January 2001 to October 2015, *Panel B*. All the p-values are reported in Table A.11. The reason for splitting the sample is that, like in FGLS, the forecast performance of all methods, absolute and relative to one another, changes dramatically during the Great Recession. This is clearly illustrated in the lower graph in Figure A.1, which shows the cumulated sum of the square forecast errors for CPI for all methods at horizon 3. The shaded areas correspond to recessionary periods according to the CEPR[5]. We observe a steady increase of the cumulated sums in the pre-crisis period, a dramatic jump during the Great Recession, followed by another period of steady increase after the crisis. The graphs for the other horizons and for IP show the same pattern (see Figures A.2, A.2). Further graphic evidence is provided in Panels A.6, A.7, A.8, A.9. The solid line is the graph of the difference between the Square Forecast Error with methods m_1 and m_2, FHLZ and SW for example, relative to IP and CPI, at each horizon, normalized by its estimated standard deviation and smoothed by a centered moving average of length $m = 61$, with the coefficients equal to 1/m. FGLS use it to test against the null of equal local performance of two forecasting methods. The zero horizontal line indicates equal performance, the dotted lines

[5]In selected recession dates CEPR follows the method used by FRED to compute NBER Recession Inndicators for the United States

indicate the 5% critical values, so that m_1 outperforms (underperforms) m_2 locally, at the 5% significance level, when the solid line is below (above) the lower (upper) dashed line.[6]

Specific results on our main variables are the following:

IP. We see that on average, and for all horizons, FHLZ outperforms the other three methods in the pre-crisis period, significantly with respect to SW and AR. SW is outperformed also by FHLR and AR. See Panel A in Table A.9. During the crisis, see Panels A.6 and A.7, SW and FHLR behave significantly better than SW and FHLZ, while AR is outperformed by all three models. With the end of the crisis the pattern stays almost the same and only in few cases the solid line head back to the pre-crisis pattern (in 2012). On average over the whole sample, FHLR outperforms FHLZ and SW at almost all horizons (all but h=1), FHLZ outperforms SW and AR at horizons 6, 12 and 24. All methods do better than AR, see Panel B in Table A.9.

CPI. In the pre-crisis period FHLZ outperforms FHLR and SW on average and AR at horizons 1 and 24, see Panel A in Table A.10. In this case the crisis has a positive effect on the performance of all three factor methods as compared to AR, as all their performances improve in relative terms, see Panels A.8 and A.9. On average over the full sample, the best methods are the two spectral density methods FHLZ and FHLR, with the exception of horizon 6 in comparison with SW, see Panel B in Table A.10 Like for IP, in general with the end of the crisis the solid line doesn't go back to the pre-crisis pattern until 2012, see Panel A.8 and A.9.

As pointed out in FGLS, the dramatic deterioration of the predictive performance of all methods corresponds to the sharp increase in the slop during the Great Recession. See Figure A.1 where we plot the sum of squares

$$\sum_{\tau=1}^{t}\sum_{i=1}^{176} z_{i\tau}^2,$$

where z_{it} is equal to Z_{it} after standardization, in the upper graph and the cumulated sum of square forecast error, 3-step ahead, in the lower one. On the other hand,

[6] The last 30 values of the moving averages are not graphed.

as soon as the crisis breaks out the covariance structure of the dataset changes abruptly. The sudden change in the covariance structure of the dataset may affect the forecasting performance of the factor and AR (See FGLS for a detailed argument on it.)

2.4.2 Forecasting the whole dataset: focus on national results

The pseudo real-time exercise is finally extended for each time series in the dataset. For real variables we use the specification adopted for IP, while for the nominal variables that adopted for CPI.

First, we compare the pseudo-real time forecasting performances of the three factor models and that of AR for IP and CPI for each of the main European countries in the dataset, namely Italy, Germany, France and Spain. Panel A.4 and A.5 show basically a similar pattern to the aggregate results, although we can notice some country-specif dissimilarity. *Tables A.9, Panel A, B* and *A.10, Panel A, B* report the mean $RMSEs$. We left to future research a deeper investigation on country-specific forecasts. Secondly, we compute the mean $RMSE$ within every group of variables (see A.15 and A.12 for details)[7]

The best performance is given in bold. We see that in the full sample FHLZ performs better than FHLR, the latter being the most accurate mainly for the Industrial production, Demand and Prices (including consumer prices and production prices) categories. In the pre-crisis sample FHLZ performs better than FHLR and SW almost for all categories and horizons. Considering median values rather than means we obtain similar results. Results for the distribution of the $RMSE$ of the models can be found in A.14 and A.17.

[7] We exclude from the evaluation the variables whose AR prediction is at 10 percent more accurate for at least one predictive horizon and for all the three factor models. In particular excluded variables belong to the money category (category 1 in Table A.2).

2.5 Conclusions

The main results in the forecasting comparison exercise involving SW, FHLR and FHLZ for Euro Area data are very similar in terms of performances to that obtained in FGLS with US data: most of time in the Great Moderation period (*pre-crisis period*) FHLZ outperforms both FHLR and SW. This pattern changes, like in FGLS, when considering the *full sample*, i.e. is affected by the sharp variation in the covariance structure caused by the crisis in 2008. Over the full sample, on average, FHLR outperforms SW and FHLZ for Industrial Production, while FHLZ and FHLR outperform SW for Inflation. In the Great Moderation period, i.e. FHLZ outperforms FHLR and SW both for Industrial Production and Inflation.

Chapter 3

DFMs: An application to the insurance sector

3.1 Dynamic Factor models to gather informations from data

As seen in *Chapter*1, the premise of dynamic factor models is that the covariation among economic time series variables at leads and lags can be captured by a few underlying unobserved series, the so called factors. In a large N and large T setting, factors can be consistently estimated by static or dynamic principal components. Hence, the first issue econometricians using DFMs incur is to estimate the factors (or, more specifically, *the space spanned by the factors*) and to ascertain how many factor they have. Once this information has been reliably collected, factors can be used for multiple purposes besides forecasting as, for example, investigation of the structure of the data.

Historically, the analysis of high-dimensional time series has attracted much interest in the area of macroeconomic time series. The same interest has not been addressed towards other fields, like, for example, the insurance sector. A first reason for this can be found in the lack of sufficient time series available, N, or in that of the number of their observations, T. Moreover, there is an innate complexity that rules the dynamics underlying the insurance sector, mainly due to the concept of risk and of human behavior driving it. There is a chance of traditional statistical analysis oversimplifying the nature of the relationships by accommodating only for a deterministic trend, presuming that the relationship is constant and invariable. One of the few contributions in factor model direction is Born et al. 2014; they empirically analyze cash flow risk management of insurance firms under a dynamic factor modeling framework in an attempt to capture the dynamic interactions between insurance company's activities in financing, investing, underwriting, and risk transferring.

The rapid development of insurance industry in the last decades, together with the consequent increase in the amount of data available, could enable researchers to explore the sector and its dynamics under a new perspective, for example by using factor models[1].

After these considerations, this section has the purpose to investigate the adaptability of factor models to a panel of data related to the insurance sector. In particular, the data analyzed refer to the Italian insurance market.

To begin, a detailed description of the dynamics of the insurance sector - along with the determinants of the insurance demand - is needed.

3.2 Towards a new forecasting model for the insurance demand

In the present academic environment, there is still little knowledge of the insurance sector data and their dynamics, especially if compared with the banking sector and, more generally, with the financial one. The econometric academic interest and that of institutions towards the sector has developed remarkably only in the past few decades, as a consequence of its progressive and continuous growth (for life, in particular, but also non-life) in all countries. Their main purpose was the identification of any possible connection and causal relationships between insurance and economic variables. The econometric insurance literature is mainly divided into two fields of research: one looking for the causal relationship between insurance growth and economic development, and the other one investigating the determinants of

[1]After 1 January 2016, with the enforcement of Solvency, the quantity and quality of data available is constantly improving.

the insurance demand (which is, in a nutshell, the revers causal relationship). We will focus on the latter.

The complexity we mentioned in the above paragraph, bears in itself a number of issues when we try to build an econometric forcasting model to analyze insurance demand. We can summarize the main ones in the following:

1. a wide knowledge of the dynamics of the insurance sector, as well as a proper identification of suitable proxies for the target variable, is necessary;

2. each specific life and non-life class requires the implementation of a specific model, as the variables affecting its development will definitely be different;

3. in identifying the variables which presumably influence the variable of interest, it must be considered that two are the large groups that enjoy the insurance services: households (for welfare, fund management, healthcare, assets protection) and companies (mainly for business protection), each one with its own economic behavior;

4. the effect produced by some economic variables may be delayed, as a consequence of multiple reasons ascribable to the peculiarities of the causal relationship between the variables analyzed, as well as to the technical characteristic featuring different classes.

3.2.1 The insurance sector: an overview

Insurance is an economic transaction in which one party (the insurance company) commits to pay a sum or to provide an uncertain service to the insured party upon payment of a certain amount of money.

One of the main variables used in insurance literature to estimate demand for policies is the *volume of premiums* (see Outreville 2011 for a useful survey), which roughly speaking, corresponds to the sum paid by the policyholder in exchange of the insurance coverage. The premiums volume is therefore a useful indicator of the market activity; it can be interpreted both as a measure of the performance of the sector itself (or of the specific class) and as the amount of the insurance demand in the market (or in in each sector of it). Looking at this variable (refer to the charts A.2.3), it appears that the Italian insurance sector had a fast growth rate in the last decades. In nominal terms, this rate is twenty times higher of that reached in 1983. The amount of premiums collected by the non-life sector (motor e non-motor) more than doubled, in real terms, from 1983 to 2013. Similarly, the total amount of premiums collected by the life sector has also increased âĂŞ still speaking in real terms âĂŞ by reaching, in 2013, values forty times higher than those registered in 1982. Panel A and B in A.2.3 shows, on one side, a strong relationship between the economic and financial framework and the insurance sector dynamics. On the other side they clearly show differences among sectors and classes (See also quarterly data Figures in A.14 for the Life classes and in A.15 for Non-life ones. Evidence of the differences between life and non-life sectors can also be found in the structure of the balance sheet itself, which reflects an extremely different setting of the business strategies. A distinctive feature between the two sectors, for example, is that in the non-life sector investments are done to 'cover known liabilities', while in the life sector, investments are mainly done to 'generate a profit'. Here below is a short description of the two sectors:

Life Insurance: life insurance is mainly about financial, longevity and mortality risks. It has increasingly become an important part of the financial sector over the past 30 years, providing a range of financial services for consumers other than classical insurance contracts. Today, life insurance policies offer two main services: income replacement for premature death and savings instruments. They also combine them in a single product. The second category of products (or the mixed ones) typically earns interests which are returned to the policyholders through capital on maturation of the policy, policy dividends etc. This class includes also products linked to some index or fund performance. See Appendix A.2.2 for a detailed description.
The drivers of the life premium growth are, therefore, different and of different nature. Among them we find disposable income, interest rates, financial market trends etc.

An important element that has remarkably affected the volumes registered in the life sector is the development, in the second half of the 90s, of the bancassurance as one of the distribution channels[2].

[2] The development of the bancassurance was mainly determined by i) the introduction of the Second Bank Directive (1989) and the consequent displacement of regulatory barriers

The insurance demand for certain types of life policy, particularly those referred to the so-called *ring fenced funds* is strictly linked to the type of asset in which these segregated management invest; in Italy, historically, those assets are mainly Government bonds. A decrease in spread, like the one that came with the end of the crisis of Italian government securities, for example, has raised the problem of how to obtain suitable profits that could be more competitive than other financial products offered on the market. An analysis of the development in trend of the premiums collected by the single classes shows that the major increase - since the first 90s - was registered on class III (*unit-linked* and *index-linked policies*). Class III premiums, though, drastically dropped during the financial crisis in years 2007-2008; class I premiums (the so called *traditional policies*), on the other hand, rose. Class V premiums, on another side, marked a significant increase between years 2003 and 2005. For an in-depth analysis see Focarelli D., Nicelli A.,2014[3]

Non-Life Insurance non-life classes offer coverages for the following risk typologies: property (damages to the insured assets caused by events such as fire, natural disasters, theft), casualty (mainly damages to third parties resulting from civil liability), accidents and sickness and others (credit, money loss, legal protection, assistance, etc.)(see appendix for details). The underlying dynamic of non-life insurance products is mainly related to the economic cycle, prices, structural factors linked to the habits besides than the personal income and, more recently, the fast development of new technologies. Figures A.15 gives evidence of this pattern. Data are very seasonal and seems do not appear to follow the economic and financial cycles as life data does.

The implications of the non-life insurance industry in times of recession or reduced economic activity are multiple (see Focarelli and Nicelli, 2014). For example, if, on one side, recessions can bring improvements to accidents in sectors like the

which used to impede the commingling of risks; ii) customers' tendency to search inside the same commercial space a comprehensive answer for their own financial and insurance needs; iii) the increasing interest of the savers towards their own saving management with the purpose of gaining guaranteed interest rates, higher than those coming from deposits

[3] D. Focarelli e A. Nicelli, *Il sistema assicurativo italiano: sfide e opportunitÃă di un mercato in forte evoluzione*, 2014, Economia dei Servizi, Anno IX, n. 2, maggio-agosto, pp. 139-160, ed. Il Mulino

motor liability (due to the fact that, driving less, the frequency of accidents is reduced), on the other side the failed growth of the incomes and of the insurable assets has a depressive effect on the insurance demand in the other non-life classes. It is not a case that, between 2007 and 2013, volume of premiums in non-life classes different from motor liability has decreased.

3.2.2 The determinants of the insurance demand

A number of studies have been trying to identify what drives insurance demand using either cross-sectional or panel data, and which are the signs of the causal effect, if it exists. Yaari (1965) was the first to develop a theoretical model to explain the demand for life insurance. Later on, Fortune (1973), for the first time, focused on the sensitive relationship between life insurance purchase and financial variables, and linked its implications to the monetary policy and capital markets. Beenstock et al.(1988) examined the relationship between property liability insurance premium sums and income; Many others followed. (see Outreville 2011nota: This study contributes to this body of research by providing an extensive literary review of empirical studies that have looked at both sides of the relationship, i.e. the demand side (economic growth is an explanatory variable among other factors that affect the demand) and the economic development side (insurance is a determinant of growth). and Petrova 2014).

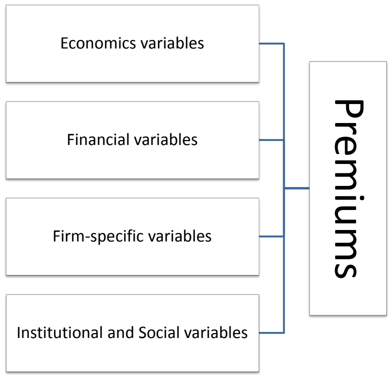
One of the first attempts to build a forecasting model for the Italian insurance market can be found in Zanghieri, 2005[4]. He provides a medium-term forecasting econometric model for life insurance premiums based on a simple theoretical model; Millo 2015 investigate the demand for Non-life Insurance in Italy.

In this project we investigate the adaptability of factor models to a panel of data related to the Italian insurance sector in order to verify the opportunity of applying dynamic factor modeling to capture the dynamic interactions between premiums volumes, economy-wide macro-variables and industry-wide business cycle variables.

We can summarize the determinants of the insurance demand in the following scheme 3.1

[4]P. Zanghieri, *Un modello trimestrale per la previsione dei premi del ramo vita*, 2005, Diritto ed economia dell'assicurazione, pp565-579, Giuffre' Editore ore

Figure 3.1: The determinants of insurance demand



Economics variables: (households and firms) income levels, prices, employement, unemmployement, exchange rates, national accounts variables (consumption, investments, imports, exports, etc) car and houses purchases.
The individual income level is clearly fundamental for investment choices (life sector) or for optional non-life coverages (non-life different from motor). Income level is obviously the corner stone of investment or coverage decisions (non-life different from motor) both for households and firms. The significant positive impact of level of income in the economy was found by all the researchers in the field. Also unemployment and inflation rate.

Financial variables: real interest rates, stock prices, etc.
Returns from insurance companies' investments or stock market performances have obvious consequences on the policyholders savings decisions. By definition, savings is what is not consumed; it is therefore allocated between different financial and real activities, functional to their relative income, generally referred to the structure of the interest rates. In a very low interest rates environment, for example, the insurance profit products attract consumer's interest, vice versa, if the guaranteed returns offered by the insurance company is low, policyholders are attracted by higher yield products. Products involving the payment of sums depending on the

performance of a specific index or fund (class III, see Appendix for details) are very sensitive with respect to financial variables.

Firm-specific variables: market actions, expansion of the distribution network, in particular of the *bancassurance*, have contributed to accelerate the development of the life sector. Recurring policies, for example, have been introduced on the market at the end of the 70s, but only in 1983, it is observed a turning point that leads to positive growth rates. However, the data regarding management actions are not easily available, while the ones referred to the distribution channels are not sufficiently long in time. Besides this, for what concerns the price of insurance, although virtually all theoretical work on insurance demand has identified price as an important factor, measuring the impact of price on the demand for insurance is difficult due to the problem of actually determining the price. The commercial price of life insurance is not observable. It is not possible, nevertheless, to estimate the effect of the tariff reduction on the market, because it involves an overall increase of the premiums, provided that the insurance demand is flexible (price sensitive), while there will be a decrease if this is not the case.

Institutional and Social variables: Political instability, regulation, life expectancy, dependency ratio, level of education and of financial education, consumer and business confidence levels, health expenditure.
Among institutional variables, changes in the regulatory framework can affect also the management actions of the insurance companies and, as a consequence, the consumer choices. The enforcement of Solvency II and the consequent introduction of 'risk-based' capital charges, for example, could lead to a process of 'de-risking' for certain types of products, with the gradual transfer of the risks at a policyholder level and therefore a change in trend of the insurance demand. Fiscal incentives for purchases of new houses, for example, or taxation changes over insurance premiums, may contribute to move insurance demand from one class to another[5]. Among social variables, the confidence in public welfare and healthcare, may lead the consumer to prefer private coverages to protect from risks. Ward and Zurbruegg 2000, Beck and Webb 2003 identify political and legal stability as im-

[5]For example, the increase of tax rate on insurance premiums occurred between 1983 and 1988 caused the policies deadlines to be moved forward on 31 December of the year prior the change of regime one, alterating therefore the trends of the insurance market

portant factors for a vibrant and growing life insurance market. The measurement of financial development, moreover, is very controversial, but two alternative proxies are usually employed. One is the ratio of quasi-money (M2-M1) to the broad definition of money (M2) as measure of the 'complexity' of the financial structure (higher ratio indicates higher level of financial development), another is the ratio of M2 to the nominal GDP. Furthermore, given that social security benefits come from taxes, which reduce available income to purchase life insurance, high social security expenditure is hypothesized to reduce the consumption of life insurance. Beenstock, Dickinson, and Khajuria (1986), Browne and Kim (1993), Skipper and Klein (2000), Ward and Zurbruegg (2002) and Beck and Webb (2002) showed that the need for life insurance purchase is reduced when government spending on social security is increased.

3.2.3 Data description

According to Luciani 2014b in factor analysis *the construction of the database is a crucial and practical problem for which there is no recipe.* How many, and which variables do we have to include in the analysis and whether there are variables that are worth excluding from the analysis have not an easy answer. A number of papers discuss whether when forecasting with factor models it is always useful to increase the size of the database. Boivin and Ng 2006 shows that, as the cross-correlation among the idiosyncratic errors increases, the estimation and forecasting performance of the model deteriorates, Luciani 2014a shows that tests and criteria for determining the number of factors are extremely unreliable when the database is poorly constructed, Onatski 2012 shows that if the explanatory power of the factors does not strongly dominate the explanatory power of the idiosyncratic terms, meaning that pervasive and nonpervasive shocks cannot be distinguished clearly, then the principal component estimator is inconsistent. In other words, when the importance of the idiosyncratic error is magnified, it will become more difficult to separate out the common from the idiosyncratic component in the data, and data with these characteristics cannot be ruled out in practice.

To summarize, in constructing the database, one should try to include enough variables to represent properly the economy he is analyzing, but not too many variables, which can jeopardize the success of the study itself. That is to say that only data that is truly informative about the factor structure should be used.

Starting from the considerations of the previous section 3.2.2, thus, some variables have been selected among the available ones. The dataset consists of 62 time series observed at quarterly frequency between January 1982 and June 2015, grouped in 8 main categories. Three of these categories consist of Italian insurance market data (market premium volumes classified by sector and by class) while the others consist of Italian macroeconomic and financial time series (prices, unemployement, interest rates, stock prices, etc). Some Euro Area time series are also included. See Appendix II for details. Data therefore include the reform on bancassurance in 1989, the Great Moderation, the Great Recession originated from the 2007 financial crisis and its spillover effect in the Euro Area from the second quarter of 2008 to the second quarter of 2009. It also includes the so called Euro Area sovereign debt crisis in 2012 and the following low interest rates environment.

For what concerns insurance data, the first 15 years of time series have been digitalized by using IVASS (former Isvap) reporting documents. A few adjustments have been done during the past years for what concerns reporting templates. In order to homogenize the series, therefore, a further classification has been executed on the pattern of the most recent one (both for life and non-life). The data referred to the distribution channels have not been included in the dataset because it has not been considered deep enough. As new life business data starts from 1988, the final dataset starts from 1988 in order, to include them data in the analysis.

Finally, to achieve stationarity, the series are transformed into first difference of the logarithm (mainly premiums, real variables and stock prices), first difference of yearly difference of the logarithm (prices) and monthly difference (interest rates, surveys), and, if needed, also deseasonalized. No treatment for outliers is applied. See appendix A.18 for details. Some series have been rejected because they values were not enough stationaty. Other variables, such as average prices, where not available.

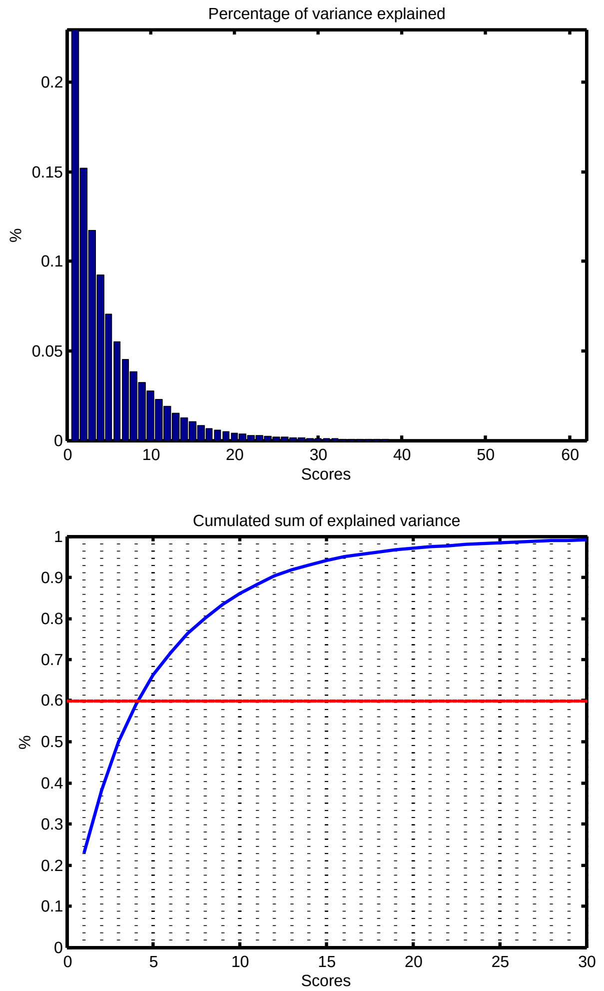
3.3 First results

The simplest statistic to describe comovements among series is the percentage of the variance of the panel accounted for by common factors estimated. If the series are characterized by strong comovements, then a small number of principal account for a relevant percentage of the overall panel variance while the remaining principal components have a small marginal contribution.

For this reason we started investigating the adaptability of factor models to insurance sector data by estimating the number of factors. The estimation method involved is the 'dynamic' one referred in section 1. Figure 3.2 show that a few number (4, for example) of dynamic principal components capture more than 60 percent of the variance of the panel. The same number of factors come out using Hallin-Liska criteria(see Chapter 1.

Keeping fix the selected number now, the second step is then the investigation of the amount of total variance explained by the common component for each series, in order to understand if our model can correctly work. Insurance series show relatively poor results with respect to that of macroeconomic or financial series; however, the relatively good findings together with the challenges linked to this new approach to insurance sector data, lead the door open to promising results after a further and more in-depth analysis.

Figure 3.2: Selecting the number of factors



Appendix A

Appendix

A.1 Appendice I - Chapter 2

A.1.1 Dataset description

In this section I give a description of the dataset, the transformation applied to each series and the category to which they belong. Table **??** refers to the number of series for each category. The dataset is an update and edited (in terms of categories and transformation) of the Eurocoin dataset.

Calling X_t a raw series, the transformations adopted are:

$$Z_t = \begin{cases} X_t & \text{if Tcode=1} \\ (1-L)X_t & \text{if Tcode=2} \\ (1-L)^2 X_t & \text{if Tcode=3} \\ \log X_t & \text{if Tcode=4} \\ (1-L)\log X_t & \text{if Tcode=5} \\ (1-L)^2 \log X_t & \text{if Tcode=6} \\ (1-L)(1-L^{12})\log X_t & \text{if Tcode=7} \end{cases}$$

Table A.1: List of series categories

CatCode	CatName	Italy	Germany	Spain	France	Euro Area	Total
1	Money	3	3	0	3	3	15
2	Import-export	0	2	2	2	0	12
3	Exchange rates	1	1	1	1	0	11
4	Prices	7	6	7	4	2	31
5	Unemployement	5	2	1	6	0	14
6	Wages	4	2	2	1	0	11
7	Industrial production	1	4	5	1	1	13
8	Demand	2	2	1	3	0	9
9	Surveys	1	2	1	5	5	24
10	Interest rates	1	1	1	1	5	11
11	Stock prices	3	3	3	3	13	25

Table A.2: List of the series

	Name	**Long Desc.**	**Tcode**	**Deseas**	**CatCode**
1	BDM1....A	BD MONEY SUPPLY - M1 - CURA	6	1	1
2	BDM2C...B	BD MONEY SUPPLY - M2 CURA	6	0	1
3	BDM3C...B	MONEY SUPPLY - M3 - CURA	6	0	1
4	FRM1....A	FR MONEY SUPPLY - M1 - CURN	6	1	1
5	FRM2....A	FR MONEY SUPPLY - M2 - CURN	6	1	1
6	FRM3....A	FR MONEY SUPPLY - M3 - CURN	6	1	1
7	ITM1....A	IT MONEY SUPPLY: M1 - CURN	6	1	1
8	ITM2....A	IT MONEY SUPPLY: M2 - CURN	6	1	1
9	ITM3....A	MONEY SUPPLY: M3 - CURN	6	1	1
10	NLM1....A	NL MONEY SUPPLY - M1 - CURN	6	1	1
11	NLM2....A	NL MONEY SUPPLY - M2 - CURN	6	1	1
12	NLM3....A	NL MONEY SUPPLY - M3 - CURN	6	1	1
13	EMECBM1.B	EM MONEY SUPPLY: M1 - CURA	6	0	1
14	EMM2....B	EM MONEY SUPPLY: M2 - CURA	6	0	1
15	EMECBM3.B	EM MONEY SUPPLY: M3 - CURA	6	0	1
16	NLIMPGDSA	NL IMPORTS - CIF - CURN	5	1	2
17	NLEXPGDSA	NL EXPORTS - FOB - CURN	5	1	2
18	FRIMPGDSB	FR IMPORTS FOB - CURA	5	1	2
19	FREXPGDSB	FR EXPORTS FOB - CURA	5	1	2
20	ESOXT003b	ES ITS EXPORTS F.O.B. TOTAL - CURA	5	1	2
21	ESOXT009b	ES ITS IMPORTS C.I.F. TOTAL - CURA	5	1	2
22	ESEXPGDSD	ES EXPORTS - CONA	5	1	2
23	ESIMPGDSD	ES IMPORTS - CONA	5	1	2
24	ESEXPPRCF	ES EXPORT UNIT VALUE INDEX - NADJ	5	1	2
25	ESIMPPRCF	ES IMPORT UNIT VALUE INDEX - NADJ	5	1	2
26	BDEXPGDSB	BD EXPORTS OF GOODS (FOB) - CURA	5	1	2
27	BDIMPGDSB	BD IMPORTS OF GOODS (CIF) - CURA	5	1	2
28	BDEXPPRCF	BD EXPORT PRICE INDEX - NADJ	7	1	4
29	BDIMPPRCF	BD IMPORT PRICE INDEX - NADJ	7	1	4
30	ITEXPPRCF	IT EXPORT UNIT VALUE INDEX - NADJ	7	1	4
31	BDOCC011	BD REAL EFFECTIVE EXCHANGE RATES - CPI BASED - NADJ	5	0	3
32	BGOCC011	BG REAL EFFECTIVE EXCHANGE RATES - CPI BASED - NADJ	5	0	3
33	ESOCC011	ES REAL EFFECTIVE EXCHANGE RATES - CPI BASED - NADJ	5	0	3
34	FNOCC011	FN REAL EFFECTIVE EXCHANGE RATES - CPI BASED - NADJ	5	0	3
35	FROCC011	FR REAL EFFECTIVE EXCHANGE RATES - CPI BASED -NADJ	5	0	3
36	GROCC011	GR REAL EFFECTIVE EXCHANGE RATES - CPI BASED - NADJ	5	0	3
37	IROCC011	IR REAL EFFECTIVE EXCHANGE RATES - CPI BASED - NADJ	5	0	3
38	ITOCC011	IT REAL EFFECTIVE EXCHANGE RATES - CPI BASED - NADJ	5	0	3
39	NLOCC011	NL REAL EFFECTIVE EXCHANGE RATES - CPI BASED - NADJ	5	0	3
40	OEOCC011	OE REAL EFFECTIVE EXCHANGE RATES - CPI BASED - NADJ	5	0	3
41	PTOCC011	PT REAL EFFECTIVE EXCHANGE RATES - CPI BASED - NADJ	5	0	3
42	BDESPPINF	BD PPI: MIG - NON-DURABLE CONSUMER GOODS - NADJ	7	0	4
43	BDPROPRCF	BD PPI: INDL. PRODUCTS, TOTAL, SOLD ON THE DOMESTIC MARKET -NADJ	7	0	4
44	BDESPPIEF	BD PPI: MIG - ENERGY - NADJ	7	0	4
45	FRESPPITF	FR PPI: MIG - INTERMEDIATE GOODS - NADJ	7	0	4
46	ITESPPINF	IT PPI: MIG - NON-DURABLE CONSUMER GOODS -NADJ	7	0	4
47	ITESPPIEF	IT PPI: MIG - ENERGY - NADJ	7	0	4
48	ESESPPITF	ES PPI: MIG - INTERMEDIATE GOODS - NADJ	7	0	4
49	ESESPPINF	ES PPI: MIG - NON-DURABLE CONSUMER GOODS - NADJ	7	0	4
50	ESPPDCNSF	ES PPI - CONSUMER GOODS, DURABLES - NADJ	7	0	4
51	ESPPINVSF	ES PPI - CAPITAL GOODS - NADJ	7	0	4
52	ESESPPIEF	ES PPI: MIG - ENERGY - NADJ	7	0	4
53	ESPROPRCF	ES PPI -NADJ	7	1	4
54	BGESPPITF	BG PPI: MIG - INTERMEDIATE GOODS - NADJ	7	0	4
55	BGESPPINF	BG PPI: MIG - NON-DURABLE CONSUMER GOODS - NADJ	7	1	4
56	BGESPPIIF	BG PPI: INDUSTRY - NADJ	7	0	4
57	NLESPPITF	NL PPI: MIG - INTERMEDIATE GOODS - NADJ	7	0	4
58	EKPROPRCF	EK PPI: INDUSTRY - NADJ	7	0	4
59	ITCPWORKF	IT CPI EXCLUDING TOBACCO (FOI) - NADJ	7	0	4
60	ITCP7500F	IT CPI (1975=100) - NADJ	7	0	4

Table A.3: List of the series - continued from previous page

	Name	Long Desc.	Tcode	Deseas	CatCode
61	ITRAWPRCF	IT RAW MATERIALS PRICE INDEX - NADJ	7	0	4
62	ITPROPRCF	IT PPI - NADJ	7	0	4
63	FRCONPRAF	FR CPI (LINKED & REBASED) - NADJ	7	0	4
64	FRAGPRC.F	FR AGRICULTURAL PRICE INDEX - NADJ	7	0	4
65	FRAGIIGSF	FR AGRICULTURAL INPUT PRICES - INVESTMENT GOODS & SERVICES - NADJ	7	1	4
66	BDCP7500F	BD CPI (1975=100) - NADJ	7	1	4
67	ESCONPRCF	ES CPI - NADJ	7	0	4
68	NLCONPRCF	NL CPI - NADJ	7	0	4
69	EMCONPRCF	EM CPI - NADJ	7	0	4
70	BDI..RELF	BD REAL EFFECTIVE FX RATE (REER) BASED ON UNIT LABOUR COSTS - NADJ	5	0	6
71	BDMWAGINF	BD WAGE&SALARY LEVEL,MTHLY BASIS - PRDG.SECT.(PAN BD M0191) NADJ	5	1	6
72	ESWAGES.F	ES WAGES: INCOME INDICATOR - VOLN	5	1	6
73	ESWAGES%F	ES WAGES: INCOME INDICATOR (%YOY) - VOLN	2	0	6
74	FRI..RELF	FR REAL EFFECTIVE FX RATE (REER) BASED ON UNIT LABOUR COSTS - NADJ	5	0	6
75	ITI..RELF	IT REAL EFFECTIVE FX RATE (REER) BASED ON UNIT LABOUR COSTS - NADJ	5	1	6
76	ITWAGES.F	IT CONTRACTUAL HOURLY WAGE: ALL WORKERS - NADJ	5	1	6
77	ITOLC007H	IT HOURLY WAGE RATE: INDUSTRY INCL. CONSTRUCTION - PROXY NADJ	5	1	6
78	ITWAGES%F	IT CONTRACTUAL HOURLY WAGE: ALL WORKERS (%YOY) - NADJ	5	0	6
79	NLI..RELF	NL REAL EFFECTIVE FX RATE (REER) BASED ON UNIT LABOUR COSTS - NADJ	5	0	6
80	NLOLC007H	NL HOURLY WAGE RATE: MFG - PROXY NADJ	5	1	6
81	BDIPTOT.G	BD INDUSTRIAL PRODUCTION INCLUDING CONSTRUCTION (CAL ADJ) - VOLA	5	0	7
82	BDESPISDH	BD IPI: MIG - DURABLE CONSUMER GOODS, VOLUME IOP (WDA) - VOLN	5	1	7
83	BDESPIESH	BD IPI: MIG-CAPITAL GOODS, VOLUME INDEX OF PRODUCTION (WDA) - VOLN	5	1	7
84	BDESPISNH	BD IPI: MIG - NON-DURABLE CONSUMER GOODS, VOLUME IOP (WDA) - VOLN	5	1	7
85	ESIPINTGH	ES INDUSTRIAL PRODUCTION - INTERMEDIATE GOODS - VOLN	5	1	7
86	ESIPINVSH	ES INDUSTRIAL PRODUCTION - CAPITAL GOODS - VOLN	5	1	7
87	ESESIBASG	ES IPI: MANUFACTURE OF BASIC METALS, VOLUME IOP (WDA) - VOLA	5	0	7
88	ESIPOMNPH	ES INDUSTRIAL PRODUCTION - OTHER NON-METAL MINERAL PRODUCTS - VOLN	5	1	7
89	ESIPTOT.G	ES INDUSTRIAL PRODUCTION (WDA) - VOLA	5	0	7
90	ITIPTOT.G	IT INDUSTRIAL PRODUCTION - VOLA	5	0	7
91	NLIPTOT.G	NL INDUSTRIAL PRODUCTION EXCLUDING CONSTRUCTION - VOLA	5	0	7
92	FRIPTOT.G	FR INDUSTRIAL PRODUCTION - VOLA	5	0	7
93	EU18	EK PRODUCTION - TOTAL INDUSTRY EXCL. CONSTRUCTION - VOLA	5	0	7
94	BDNEWORDE	BD MANUFACTURING ORDERS - SADJ	5	0	8
95	BDRVNCARP	BD NEW PASSENGER CAR REGISTRATIONS - VOLN	5	1	8
96	BGACECARP	BG NEW PASSENGER CAR REGISTRATIONS - VOLN	5	1	8
97	ESCAR...O	ES REGISTRATIONS: PASSENGER CAR - VOLA	5	0	8
98	FRCARREGO	FR NEW CAR REGISTRATIONS (CAL ADJ) -VOLA	5	0	8
99	FRHCONMFD	FR HOUSEHOLD CONSUMPTION - MANUFACTURED GOODS - CONA	5	0	8
100	FRHCONDGD	FR HOUSEHOLD CONSUMPTION - DURABLE GOODS - CONA	5	0	8
101	ITNEWORDF	IT NEW ORDERS - NADJ	5	1	8
102	ITRETTOTF	IT RETAIL SALES - NADJ	5	1	8
103	BDCNFCONQ	BD CONSUMER CONFIDENCE INDICATOR - GERMANY - SADJ	2	0	9
104	BGCNFCONQ	BG BNB CONS. SVY.: CONSUMER CONFIDENCE INDICATOR (EP) - SADJ	2	0	9
105	BGCNFBUSQ	BG BUSINESS INDICATOR SURVEY - ECONOMY - SADJ	2	0	9
106	BGEUSIOBQ	BG IND.: OVERALL - ORD BOOKS - SADJ	2	0	9
107	BG000183Q	BG BNB BUS. SVY. - MANUFACTURING - NOT SMOOTHED - SADJ	2	0	9
108	BG000186Q	BG BNB BUS. SVY. - BUILDING - NOT SMOOTHED - SADJ	2	0	9
109	BG000189Q	BG BNB BUS. SVY. - TRADE - NOT SMOOTHED - SADJ	2	0	9
110	BGSURECSQ	BG BNB CONS.SVY.: ECON.SITUATION- FCST. OVER NEXT 12 MONTHS - SADJ	2	0	9
111	BGSURPUHQ	BG BNB CONS.SVY.: MAJOR HH.PURCH-FCST.OVER NEXT 12 MONTHS(EP)	2	0	9
112	ESINT384R	ES PRODUCTION LEVEL - INDUSTRY - NADJ	2	0	9
113	FRINDSYNQ	FR SURVEY: MANUFACTURING - SYNTHETIC BUSINESS INDICATOR - SADJ	2	0	9
114	FRSURPMPQ	FR SURVEY: MANUFACTURING OUTPUT - RECENT OUTPUT TREND - SADJ	2	0	9
115	FRSURGMPQ	FR SURVEY: MANUFACTURING OUTPUT - ORDER BOOK & DEMAND - SADJ	2	0	9
116	FRSURGPDQ	FR SURVEY: MANUFACTURING OUTPUT LEVEL - GENERAL OUTLOOK - SADJ	2	0	9
117	FRSURTMPQ	FR SURVEY: MANUFACTURING OUTPUT - PERSONAL OUTLOOK - SADJ	2	0	9
118	ITHHFECSR	IT HOUSEHOLD CONFIDENCE SURVEY: FUTURE FINANCIAL POSITION - NADJ	2	0	9
119	ITCNFCONQ	IT HOUSEHOLD CONFIDENCE INDEX - SADJ	5	0	9
120	NLCNFBUSQ	NL CBS MFG. SVY.: PRODUCER CONFIDENCE INDEX - SADJ	2	0	9

Table A.4: List of the series - continues from previous page

	Name	Long Desc.	Tcode	Deseas	CatCode
121	NLEUSCPCR	NL CONSUMER SURVEY: MAJOR PURCH.OVER NEXT 12 MONTHS-NETHERLANDS	2	0	9
122	EKCNFBUSQ	EK INDUSTRIAL CONFIDENCE INDICATOR - EA - SADJ	2	0	9
123	EMEUSCCIQ	EM CONSUMER CONFIDENCE INDICATOR - EA - SADJ	2	0	9
124	EKEUSIPAQ	EK INDUSTRY SURVEY: PRODUCTION EXPECTATIONS (EA) - SADJ	2	0	9
125	EKEUBCI.R	EK BUSINESS CLIMATE INDICATOR-COMMON FACTOR IN IND. (EA) - NADJ	2	0	9
126	EKEUSESIG	EK ECONOMIC SENTIMENT INDICATOR (EA18) - VOLA	5	0	9
127	EMGBOND.	EM GOVERNMENT BOND YIELD - 10 YEAR	2	0	10
128	EMECB2Y.	EM GOVERNMENT BOND YIELD - 2 YEAR	2	0	10
129	EMECB3Y.	EM GOVERNMENT BOND YIELD - 3 YEAR	2	0	10
130	EMECB5Y.	EM GOVERNMENT BOND YIELD - 5 YEAR	2	0	10
131	EMECB7Y.	EM GOVERNMENT BOND YIELD - 7 YEAR	2	0	10
132	BDESSFUB	BD HARMONISED GOVERNMENT 10-YEAR BOND YIELD	2	0	10
133	FRESSFUB	FR HARMONISED GOVERNMENT 10-YEAR BOND YIELD	2	0	10
134	ESESSFUB	ES HARMONISED GOVERNMENT 10-YEAR BOND YIELD	2	0	10
135	BGESSFUB	BG HARMONISED GOVERNMENT 10-YEAR BOND YIELD	2	0	10
136	ITESSFUB	IT HARMONISED GOVERNMENT 10-YEAR BOND YIELD	2	0	10
137	ITINTER3	IT INTERBANK DEPOSIT RATE-AVERAGE ON 3-MONTHS DEPOSITS	2	0	10
138	MSEROP$ E	MSCI EUROPE U$ - PRICE INDEX	5	0	11
139	INDGSIT E	ITALY-DS Inds Gds & Svs - PRICE INDEX	5	0	11
140	INDGSBD E	GERMANY-DS Inds Gds & Svs - PRICE INDEX	5	0	11
141	INDGSFR E	FRANCE-DS Inds Gds & Svs - PRICE INDEX	5	0	11
142	INDUSBD E	GERMANY-DS Industrials - PRICE INDEX	5	0	11
143	INDUSFR E	FRANCE-DS Industrials - PRICE INDEX	5	0	11
144	INDUSIT E	ITALY-DS Industrials - PRICE INDEX	5	0	11
145	FINANFR E	FRANCE-DS Financials - PRICE INDEX	5	0	11
146	FINANBD E	GERMANY-DS Financials - PRICE INDEX	5	0	11
147	FINANIT E	ITALY-DS Financials - PRICE INDEX	5	0	11
148	CNSMGFR E	FRANCE-DS Consumer Gds - PRICE INDEX	5	0	11
149	CNSMGBD E	GERMANY-DS Consumer Gds - PRICE INDEX	5	0	11
150	CNSMGIT E	ITALY-DS Consumer Gds - PRICE INDEX	5	0	11
151	OILGSEM E	EMU-DS Oil & Gas - PRICE INDEX	5	0	11
152	BMATREM E	EMU-DS Basic Mats - PRICE INDEX	5	0	11
153	INDUSEM E	EMU-DS Industrials - PRICE INDEX	5	0	11
154	RITDVEM E	EMU-DS Divers. REITs - PRICE INDEX	5	0	11
155	CNSMGEM E	EMU-DS Consumer Gds - PRICE INDEX	5	0	11
156	HLTHCEM E	EMU-DS Health Care - PRICE INDEX	5	0	11
157	TELCMEM E	EMU-DS Telecom - PRICE INDEX	5	0	11
158	UTILSEM E	EMU-DS Utilities - PRICE INDEX	5	0	11
159	FINANEM E	EMU-DS Financials - PRICE INDEX	5	0	11
160	CNSMSEM E	EMU-DS Consumer Svs - PRICE INDEX	5	0	11
161	TECNOEM E	EMU-DS Technology - PRICE INDEX	5	0	11
162	EMSHRPRCF	EM DATASTREAM EURO SHARE PRICE INDEX (MONTHLY AVERAGE) - NADJ	5	0	11
163	BDMLM006Q	BD REGISTERED UNEMPLOYMENT: RATE (ALL PERSONS) - SADJ	2	0	5
164	BDMLM005O	BD REGISTERED UNEMPLOYMENT: LEVEL (ALL PERSONS) - VOLA	5	0	5
165	ITMLFT15O	IT HARMONISED UNEMPLOYMENT: LEVEL, ALL PERSONS (ALL AGES) - VOLA	5	0	5
166	ITMLRT16Q	IT HARMONISED UNEMPLOYMENT: RATE, ALL PERSONS (ALL AGES) - SADJ	2	0	5
167	ITMLRT14Q	IT HARMONISED UNEMPLOYMENT: RATE, ALL PERSONS (AGES 15-24) - SADJ	2	0	5
168	ITMLRF16Q	IT HARMONISED UNEMPLOYMENT: RATE, FEMMES (ALL AGES) - SADJ	2	0	5
169	ITMLRM16Q	IT HARMONISED UNEMPLOYMENT: RATE, HOMMES (ALL AGES) - SADJ	2	0	5
170	FRESTUNPO	FR UNEMPLOYMENT: TOTAL - TOTAL - VOLA	5	0	5
171	FRMLRT14Q	FR HARMONISED UNEMPLOYMENT: RATE, ALL PERSONS (AGES 15-24) - SADJ	2	0	5
172	FRMLRT15Q	FR HARMONISED UNEMPLMT.: RATE,ALL PERSONS(AGES 25 AND OVER) - SADJ	2	0	5
173	FRMLRT16Q	FR HARMONISED UNEMPLOYMENT: RATE, ALL PERSONS (ALL AGES) -SADJ	2	0	5
174	FRMLRF16Q	FR HARMONISED UNEMPLOYMENT: RATE, FEMMES (ALL AGES) - SADJ	2	0	5
175	FRMLRm16Q	FR HARMONISED UNEMPLOYMENT: RATE, HOMMES (ALL AGES) - SADJ	2	0	5
176	ESMLM005O	ES HARMONISED UNEMPLOYMENT: LEVEL, ALL PERSONS (ALL AGES) - VOLA	5	0	5

A.1.2 Tables

Table A.5: Calibration: SW

Panel SW: *S1* - number of static factors

h	IP(1)	IP(2)	IP(3)	IP(4)	IP(5)	IP(6)	IP(7)	IP(8)	IP(BN)
1	0.984	0.963	1.010	1.030	0.993	0.978	1.000	1.020	1.016
3	1.370	0.946	0.977	0.994	0.953	0.993	1.000	0.992	0.981
6	1.500	1.090	1.150	1.100	0.975	1.060	1.000	0.985	1.150
12	1.300	1.210	1.230	1.190	1.100	1.130	1.000	0.999	1.220
24	0.913	0.995	1.040	0.995	0.993	0.992	1.000	1.070	1.030
mean	1.210	1.040	1.080	1.060	**1.000**	1.030	**1.000**	**1.010**	1.080

h	CPI(1)	CPI(2)	CPI(3)	CPI(4)	CPI(5)	CPI(6)	CPI(7)	CPI(8)	CPI(BN)
1	1.120	1.030	0.956	0.965	0.962	0.980	1.000	1.020	1.010
3	0.952	1.010	1.010	1.040	1.030	1.050	1.000	0.985	1.020
6	0.980	1.040	1.050	1.070	1.070	1.110	1.000	1.040	1.050
12	1.130	1.120	1.130	1.140	1.130	1.130	1.000	0.980	1.120
24	1.050	0.999	1.020	1.020	1.030	1.050	1.000	0.985	0.995
mean	1.050	1.040	1.030	1.050	1.050	1.060	**1.000**	**1.001**	1.040

Panel SW: *S2* - target lag order $\beta_i(L)$

h	IP(1)	CPI(1)	IP(BIC)	CPI(BIC)	IP(AIC)	CPI(AIC)
1	1.000	1.000	1.060	0.999	1.090	0.999
3	1.000	1.000	1.100	1.060	1.130	1.100
6	1.000	1.000	1.070	1.080	1.090	1.140
12	1.000	1.000	1.070	1.040	1.060	1.080
24	1.000	1.000	0.978	1.010	1.020	1.010
mean	**1.000**	**1.000**	1.060	1.040	1.080	1.070

Table A.6: Calibration: SW

Panel SW: *S3* - factors lag order $\alpha_i(L)$

h	IP(0)	CPI(0)	IP(BIC)	CPI(BIC)	IP(AIC)	CPI(AIC)
1	1.000	1.000	0.920	1.090	0.899	1.270
3	1.000	1.000	1.120	1.190	1.150	1.270
6	1.000	1.000	1.040	1.190	1.180	1.310
12	1.000	1.000	1.230	1.100	1.300	1.140
24	1.000	1.000	1.080	1.060	1.250	1.060
mean	**1.000**	**1.000**	1.080	1.130	1.160	1.210

Panel SW: *S4* - target and factors lag order $\beta_i(L)$, $\alpha_i(L)$

h	IP(1,0)	CPI(1,0)	IP(BIC,BIC)	CPI(BIC,BIC)	IP(AIC,AIC)	CPI(AIC,AIC)
1	1.000	1.000	1.000	1.049	0.966	1.249
3	1.000	1.000	1.238	1.199	1.206	1.232
6	1.000	1.000	1.073	1.077	1.220	1.145
12	1.000	1.000	1.263	1.074	1.311	1.091
24	1.000	1.000	1.086	1.042	1.432	1.041
mean	**1.000**	**1.000**	1.132	1.088	1.227	1.152

Table A.7: Calibration: FHLZ

Panel FHLZ: *S1* - lag order criterion

h	IP(AIC)	CPI(AIC)	IP(BIC)	CPI(BIC)
1	**0.959**	1.031	1.000	**1.000**
3	**0.950**	1.061	1.000	**1.000**
6	**0.954**	1.045	1.000	**1.000**
12	**0.987**	1.028	1.000	**1.000**
24	**1.016**	1.011	1.000	**1.000**
mean	**0.973**	1.035	1.000	**1.000**

Panel FHLZ: *S2* - max lag order

h	IP(3)	CPI(3)	IP(4)	CPI(4)	IP(5)	CPI(5)	IP(6)	CPI(6)	IP(7)	CPI(7)
1	1.0	1.0	1.0009	1.0	1.0015	1.0	1.0015	1.0	1.0016	1.0
3	1.0	1.0	0.9998	1.0	1.0000	1.0	0.9997	1.0	0.9999	1.0
6	1.0	1.0	1.0000	1.0	0.9999	1.0	0.9990	1.0	0.9991	1.0
12	1.0	1.0	0.9995	1.0	0.9995	1.0	0.9991	1.0	0.9991	1.0
24	1.0	1.0	0.9998	1.0	0.9998	1.0	1.0000	1.0	1.0000	1.0
mean	1.0	1.0	1.0000	1.0	1.0001	1.0	0.9999	1.0	0.9999	1.0

Panel FHLZ: *S3* - bandwidth

h	IP(25)	CPI(25)	IP(30)	CPI(30)	IP(35)	CPI(35)	IP(40)	CPI(40)
1	1.008	0.998	1.0	1.0	1.018	1.003	1.019	1.006
3	0.982	0.998	1.0	1.0	1.007	1.002	1.012	1.003
6	0.982	0.998	1.0	1.0	1.008	1.001	1.012	1.001
12	0.990	0.997	1.0	1.0	1.002	1.000	1.002	1.001
24	0.999	0.999	1.0	1.0	1.000	1.000	0.996	1.000
mean	0.992	0,998	1.0	1.0	1.007	1.001	1.008	1.002

Panel FHLZ: *S4* - kernel

h	IP(triang)	CPI(triang)	IP(gauss)	CPI(gauss)
1	1.003	1.002	1.000	1.000
3	1.014	1.001	1.000	1.000
6	1.016	1.000	1.000	1.000
12	1.000	1.001	1.000	1.000
24	0.997	1.000	1.000	1.000
mean	1.007	1.001	1.000	1.000

Table A.8: Calibration: FHLR

FHLR: *S1* - kernel

h	IP(triang)	CPI(triang)	IP(gauss)	CPI(gauss)
1	1.000	1.000	**0.970**	**0.998**
3	1.000	1.000	**0.914**	**0.998**
6	1.000	1.000	**0.936**	**0.990**
12	**1.000**	1.000	1.019	**0.987**
24	**1.000**	1.000	1.051	**0.986**
mean	1.000	1.000	**0.978**	**0.992**

FHLR: *S2* - bandwidth

h	IP(25)	CPI(25)	IP(30)	CPI(30)	IP(35)	CPI(35)	IP(40)	CPI(40)
1	1.002	**0.995**	1.000	1.000	**0.947**	1.034	0.983	1.061
3	1.034	**0.991**	1.000	1.000	1.134	1.025	1.768	0.915
6	1.034	**0.993**	1.000	1.000	1.002	1.032	1.597	0.957
12	1.000	**0.922**	1.000	1.000	**0.946**	1.060	1.083	1.011
24	1.028	**1.000**	1.000	1.000	**0.922**	0.992	0.835	1.029
mean	1.019	**0.994**	1.000	1.000	**0.991**	1.029	1.253	0.995

Table A.9: Mean Square Forecast Error Relative to AR - IP

[a]

Panel A : Pre Crisis (2001 : 1 – 2008 : 3)

IP				
	FHLZ	FHLR	SW	AR
h=1	$0.989^{\dagger\dagger\dagger}$	$1.202^{\dagger\dagger}$	1.233	1.000
h=3	$0.763^{**\dagger}$	$0.763^{**\dagger}$	0.826	1.000
h=6	$0.765^{*\dagger}$	0.953	1.045	1.000
h=12	0.813^{**}	0.937	1.023	1.000
h=24	$0.924^{*\dagger}$	0.972	1.003	1.000

Main European countries IP (h-average)				
	FHLZ	FHLR	SW	AR
Italy	0.912	0.966	1.083	1.000
Germany	0.965	0.985	1.001	1.000
France	0.944	0.982	1.053	1.000
Spain	0.956	1.058	1.134	1.000

Panel B : Full Sample (2000 : 1 – 2015 : 10)

IP				
	FHLZ	FHLR	SW	AR
h=1	0.939	0.921	0.887	1.000
h=3	0.833	0.765	0.782	1.000
h=6	0.631	$0.622^{\dagger}$	0.651	1.000
h=12	0.758	0.757	0.765	1.000
h=24	$0.944^{*\dagger}$	$0.929^{*\dagger}$	0.982	1.000

Main European countries IP (h-average)				
	FHLZ	FHLR	SW	AR
Italy	0.861	0.836	0.856	1.000
Germany	0.811	0.787	0.801	1.000
France	0.868	0.855	0.902	1.000
Spain	0.876	0.871	0.904	1.000

[a]One, two or three asterisks indicate that the null of equal performance of the three factor models relative to AR is rejected at the 1%, 5%, 10% significance level, respectively, by the Diebold-Mariano test. One, two or three daggers indicate the for FHLZ or FHLR same with respect to SW. All the p-values are reported in Table A.11.

Table A.10: Mean Square Forecast Error Relative to AR - CPI

[a]

Panel A : Pre Crisis $(2001:1-2008:3)$

CPI				
	FHLZ	FHLR	SW	AR
h=1	0.959	0.974	0,970	1.000
h=3	1.059†	1.091	1.096	1.000
h=6	1.159†	1.201†	1.267	1.000
h=12	1.017†	1.083	1.240	1.000
h=24	0.841	0.805	0.871	1.000

Main European countries CPI (h-average)				
	FHLZ	FHLR	SW	AR
Italy	0.909	0.876	0.916	1.000
Germany	1.026	1.048	1.190	1.000
France	1.072	1.095	1.148	1.000
Spain	1.005	1.022	1.060	1.000

Panel B : Full Sample $(2000:1-2015:10)$

CPI				
	FHLZ	FHLR	SW	AR
h=1	0.871**	0.881**	0.899	1.000
h=3	0.815**†	0.858*	0.881	1.0
h=6	0.840	0.836	0.825	1.000
h=12	0.882	0.885*	0.910	1.000
h=24	0.989	0.942††	1.010	1.000

Main European countries CPI (h-average)				
	FHLZ	FHLR	SW	AR
Italy	0.953	1.005	1.197	1.000
Germany	0.948	0.963	1.027	1.000
France	0.925	0.928	0.947	1.000
Spain	0.946	0.933	0.938	1.000

[a] One, two or three asterisks indicate that the null of equal performance of the three factor models relative to AR is rejected at the 1%, 5%, 10% significance level, respectively, by the Diebold-Mariano test. One, two or three daggers indicate the for FHLZ or FHLR same with respect to SW. All the p-values are reported in Table A.11.

Table A.11: Diebold-Mariano test: p-values

Panel A : Pre Crisis $(2001:1-2008:3)$

IP						
	FHLZ vs SW	FHLR vs SW	FHLZ vs FHLR	FHLZ vs AR	FHLR vs AR	SW vs AR
h=1	0.001	0.179	0.001	0.476	0.958	0.970
h=3	0.216	0.057	0.499	0.013	0.025	0.088
h=6	0.031	0.256	0.000	0.087	0.364	0.609
h=12	0.500	0.118	0.500	0.017	0.500	0.500
h=24	0.067	0.500	0.000	0.060	0.500	0.497
CPI						
	FHLZ vs SW	FHLR vs SW	FHLZ vs FHLR	FHLZ vs AR	FHLR vs AR	SW vs AR
h=1	0.356	0.541	0.253	0.356	0.485	0.459
h=3	0.095	0.410	0.059	0.882	0.944	0.956
h=6	0.056	0.080	0.158	0.998	1.000	0.999
h=12	0.090	0.107	0.111	0.599	0.738	0.886
h=24	0.366	0.184	0.834	0.121	0.133	0.304

Panel B : Full Sample $(2000:1-2015:10)$

IP						
	FHLR vs SW	FHLZ vs FHLR	FHLZ vs AR	FHLR vs AR	SW vs AR	SW vs AR
h=1	0.762	0.889	0.576	0.193	0.223	0.130
h=3	0.680	0.328	0.822	0.241	0.132	0.137
h=6	0.192	0.004	0.661	0.166	0.17	0.188
h=12	0.447	0.385	0.514	0.126	0.144	0.172
h=24	0.022	0.014	0.749	0.081	0.077	0.500
CPI						
	FHLZ vs SW	FHLR vs SW	FHLZ vs FHLR	FHLZ vs AR	FHLR vs AR	SW vs AR
h=1	0.170	0.250	0.283	0.014	0.037	0.079
h=3	0.082	0.183	0.072	0.045	0.091	0.124
h=6	0.580	0.602	0.542	0.123	0.133	0.169
h=12	0.368	0.326	0.454	0.500	0.078	0.266
h=24	0.330	0.024	0.950	0.467	0.265	0.513

Table A.12: Mean RMSE by category - Full Sample (2001:1 - 2015:10)

FHLZ					
	h=1	h=3	h=6	h=12	h=24
Import-Export	**0.967**	0.916	**0.909**	**0.914**	**0.938**
Exchange rates	**0.994**	**0.991**	**0.995**	**0.976**	**0.934**
Prices (PPI, CPI)	0.992	0.968	0.948	0.943	1.040
Unemployement	1.025	**0.947**	**0.981**	**0.977**	0.951
Wages	**0.956**	**0.965**	**0.982**	**0.902**	**0.948**
Idustrial Production	**0.928**	0.855	0.758	0.837	0.928
Demand	**0.972**	0.920	0.860	0.877	0.949
Surveys	**0.966**	**0.980**	**0.960**	**0.945**	0.970
Interest rates	**0.822**	**0.840**	**0.860**	**0.869**	**0.862**
Stock prices	**0.940**	**0.946**	**0.939**	**0.933**	**0.957**

FHLR					
	h=1	h=3	h=6	h=12	h=24
Import-Export	1.067	**0.906**	0.924	0.933	0.939
Exchange rates	1.050	1.040	1.046	1.034	1.012
Prices (PPI, CPI)	**0.959**	**0.937**	0.906	**0.921**	**0.971**
Unemployement	1.066	0.978	0.997	**0.972**	0.943
Wages	0.987	1.011	1.039	0.924	0.958
Idustrial Production	0.932	**0.809**	**0.747**	**0.833**	**0.919**
Demand	0.991	**0.912**	**0.837**	**0.862**	**0.926**
Surveys	0.974	0.981	0.987	0.956	**0.928**
Interest rates	0.838	0.901	0.936	0.926	0.871
Stock prices	0.987	1.002	1.000	0.977	0.977

Table A.13: Mean RMSE by category - Full Sample (2001:1 - 2015:10)

	SW				
	h=1	h=3	h=6	h=12	h=24
Import-Export	1.063	0.913	0.943	0.972	1.037
Exchange rates	1.099	1.111	1.165	1.189	1.312
Prices (PPI, CPI)	0.965	**0.936**	**0.88**	0.926	1.068
Unemployement	1.092	1.004	1.04	1.014	**0.938**
Wages	1.071	1.124	1.234	1.21	1.363
Idustrial Production	0.939	0.831	0.782	0.838	0.965
Demand	1.061	0.963	0.878	0.876	0.967
Surveys	0.971	**0.979**	1.003	**0.948**	0.968
Interest rates	0.882	0.965	1.002	1.069	0.927
Stock prices	0.995	0.986	1.031	1.051	1.089

Table A.14: Distribution RMSE - Full Sample (2001:1 - 2015:10)

FHLZ					
Percentile:	0.05	0.25	0.50	0.75	0.95
h=1	0.869	0.926	0.961	0.998	1.076
h=3	0.813	0.918	0.954	0.993	1.065
h=6	0.755	0.91	0.947	0.989	1.047
h=12	0.793	0.902	0.943	0.975	1.033
h=24	0.864	0.926	0.960	1.007	1.102

FHLR					
Percentile:	0.05	0.25	0.50	0.75	0.95
h=1	0.829	0.939	0.985	1.030	1.128
h=3	0.789	0.903	0.987	1.027	1.073
h=6	0.743	0.899	0.971	1.025	1.089
h=12	0.805	0.890	0.954	0.998	1.082
h=24	0.844	0.918	0.952	0.997	1.073

SW					
Percentile:	0.05	0.25	0.50	0.75	0.95
h=1	0.858	0.945	1.004	1.058	1.18
h=3	0.785	0.91	0.984	1.059	1.134
h=6	0.73	0.877	1.005	1.084	1.222
h=12	0.757	0.903	0.987	1.108	1.286
h=24	0.855	0.966	1.042	1.133	1.378

Table A.15: Mean RMSE by category - Pre Crisis (2001:1 - 2008:3)

FHLZ					
	h=1	h=3	h=6	h=12	h=24
Import-Export	**0.948**	**0.927**	**0.943**	**0.898**	**0.805**
Exchange rates	**0.997**	1.013	1.021	**0.953**	**0.863**
Prices (PPI, CPI)	**0.982**	**0.972**	**0.952**	**0.941**	**0.806**
Unemployement	**0.963**	**0.840**	**0.822**	**0.770**	**0.812**
Wages	**0.933**	**0.931**	**0.967**	**0.872**	**0.881**
Idustrial Production	**1.010**	**0.877**	**0.847**	**0.863**	**0.811**
Demand	**0.986**	**0.971**	**0.919**	**0.946**	**0.8412**
Surveys	**0.955**	**0.931**	**0.892**	**0.879**	**0.706**
Interest rates	**0.929**	**0.904**	**0.929**	**1.007**	**1.011**
Stock prices	**0.954**	**0.939**	**0.942**	**0.952**	**0.926**

FHLR					
	h=1	h=3	h=6	h=12	h=24
Import-Export	1.016	0.988	1.080	0.959	**0.805**
Exchange rates	1.081	**1.008**	**0.990**	0.983	0.911
Prices (PPI, CPI)	0.984	1.005	1.020	1.023	**0.818**
Unemployement	1.019	0.950	0.973	0.922	0.945
Wages	0.965	0.944	0.982	0.910	0.928
Idustrial Production	1.169	0.900	0.993	0.918	**0.811**
Demand	1.078	1.090	1.081	1.057	0.852
Surveys	0.973	0.950	0.946	**0.871**	**0.707**
Interest rates	0.949	1.018	1.172	1.179	1.068
Stock prices	0.966	0.964	0.985	0.978	0.965

Table A.16: Mean RMSE by category - Pre Crisis (2001:1 - 2008:3)

	SW				
	h=1	h=3	h=6	h=12	h=24
Import-Export	1.042	1.03	1.208	1.117	0.953
Exchange rates	1.129	1.053	1.12	1.162	1.293
Prices (PPI, CPI)	0.991	1.028	1.098	1.147	1.117
Unemployement	1.051	1.012	1.101	1.107	1.008
Wages	1.051	1.045	1.16	1.17	1.551
Idustrial Production	1.221	0.984	1.133	1.036	0.989
Demand	1.171	1.143	1.235	1.161	0.942
Surveys	0.980	0.966	1.023	0.95	0.933
Interest rates	1.028	1.177	1.365	1.518	1.17
Stock prices	0.996	0.953	1.041	1.048	1.131

Table A.17: Distribution RMSE - Pre Crisis (2001:1 - 2008:3)

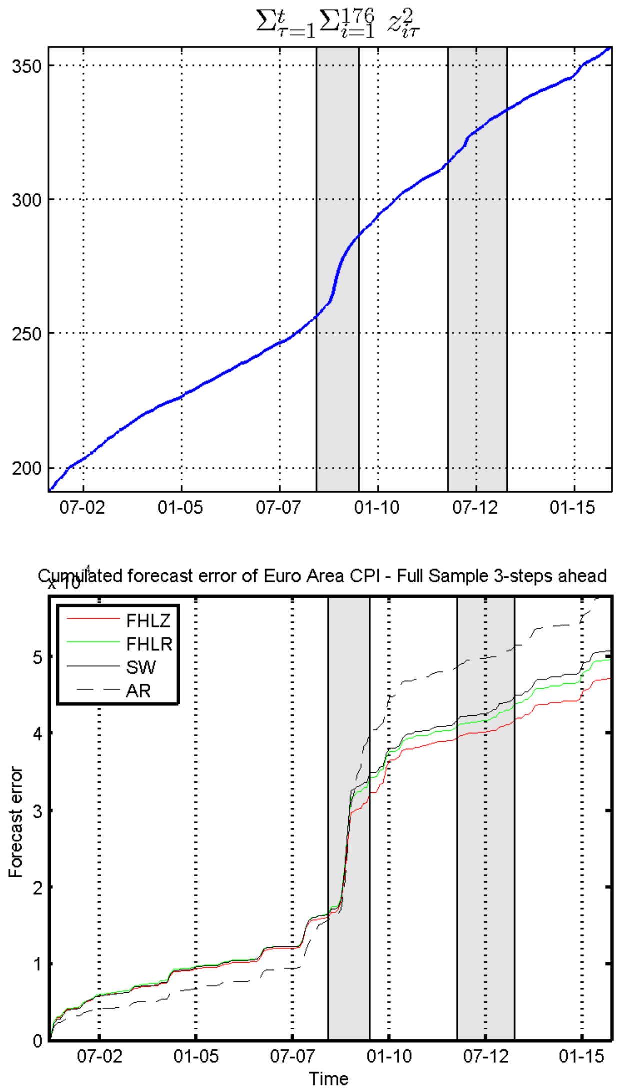
FHLZ					
Percentile:	0.05	0.25	0.50	0.75	0.95
h=1	0.882	0.925	0.954	1.005	1.054
h=3	0.780	0.910	0.935	0.981	1.076
h=6	0.774	0.891	0.942	0.974	1.050
h=12	0.672	0.887	0.935	0.981	1.054
h=24	0.606	0.763	0.879	0.954	1.024

FHLR					
Percentile:	0.05	0.25	0.50	0.75	0.95
h=1	0.870	0.939	0.986	1.041	1.161
h=3	0.861	0.926	0.988	1.020	1.117
h=6	0.890	0.934	0.994	1.071	1.200
h=12	0.800	0.914	0.969	1.040	1.201
h=24	0.604	0.786	0.893	0.994	1.107

SW					
Percentile:	0.05	0.25	0.50	0.75	0.95
h=1	0.871	0.960	1.017	1.073	1.228
h=3	0.870	0.936	1.014	1.085	1.182
h=6	0.926	1.001	1.080	1.190	1.419
h=12	0.874	0.999	1.059	1.185	1.537
h=24	0.844	0.921	1.055	1.208	1.659

A.1.3 Figures

Figure A.1: Graph of $\sum_{\tau=1}^{t} \sum_{i=1}^{176} z_{i\tau}^2$ and Cumulated Sum of Square Forecast Error, 3-step ahead, CPI



Shaded areas indicate CEPR recession dates, which follows the trough method used by FRED to compute NBER Recession Inndicators for the United States.

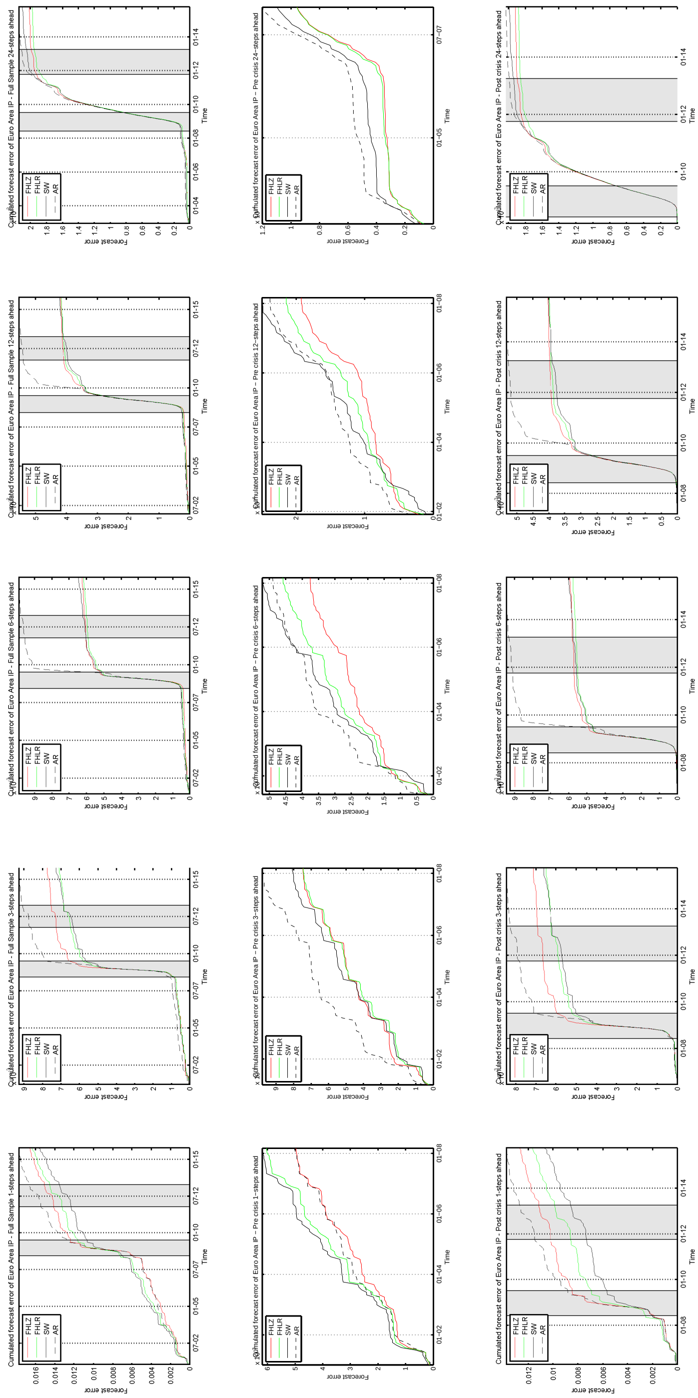


Figure A.2: Cumulated forecast error (IP)

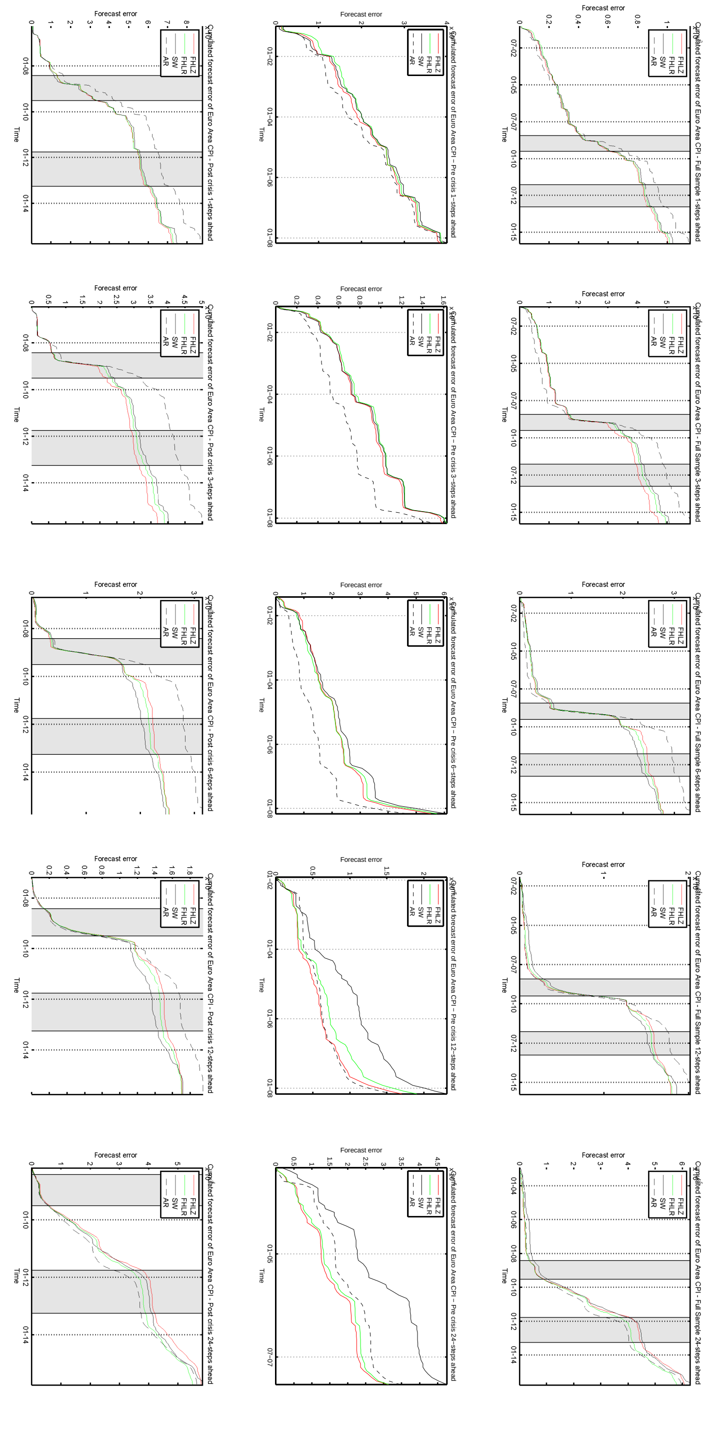


Figure A.3: Cumulated forecast error (CPI)

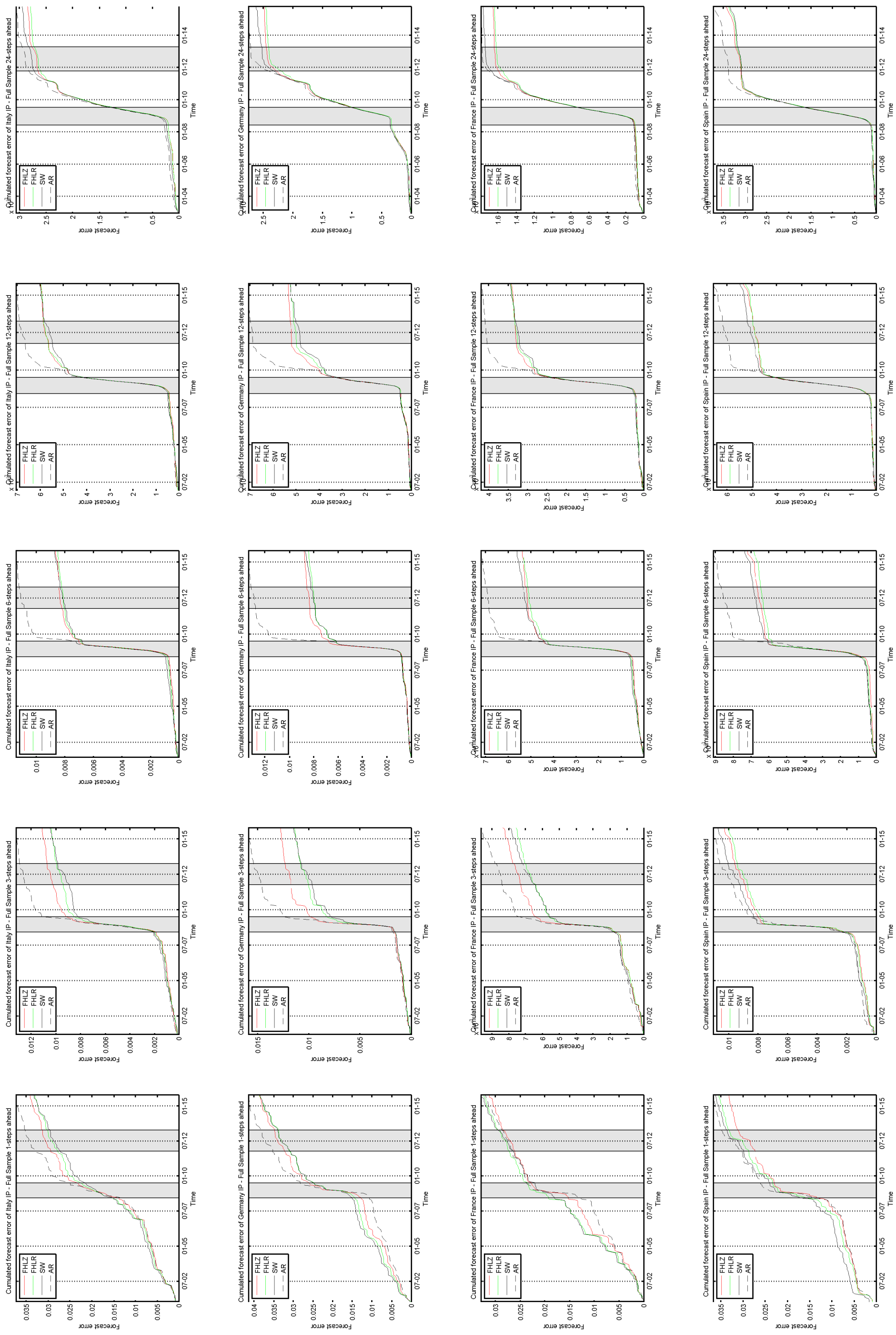


Figure A.4: Cumulated forecast error Main European countries (IP)

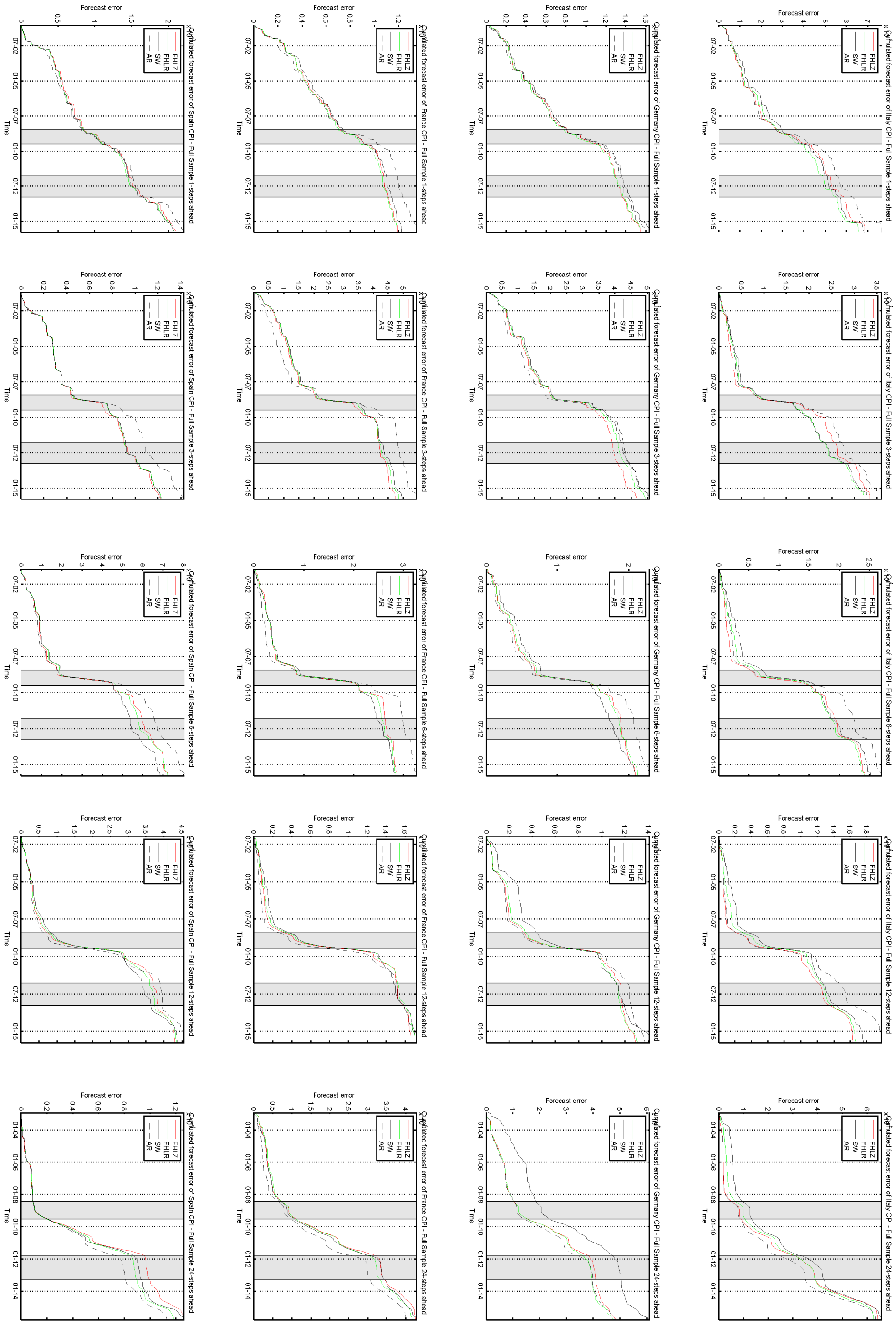
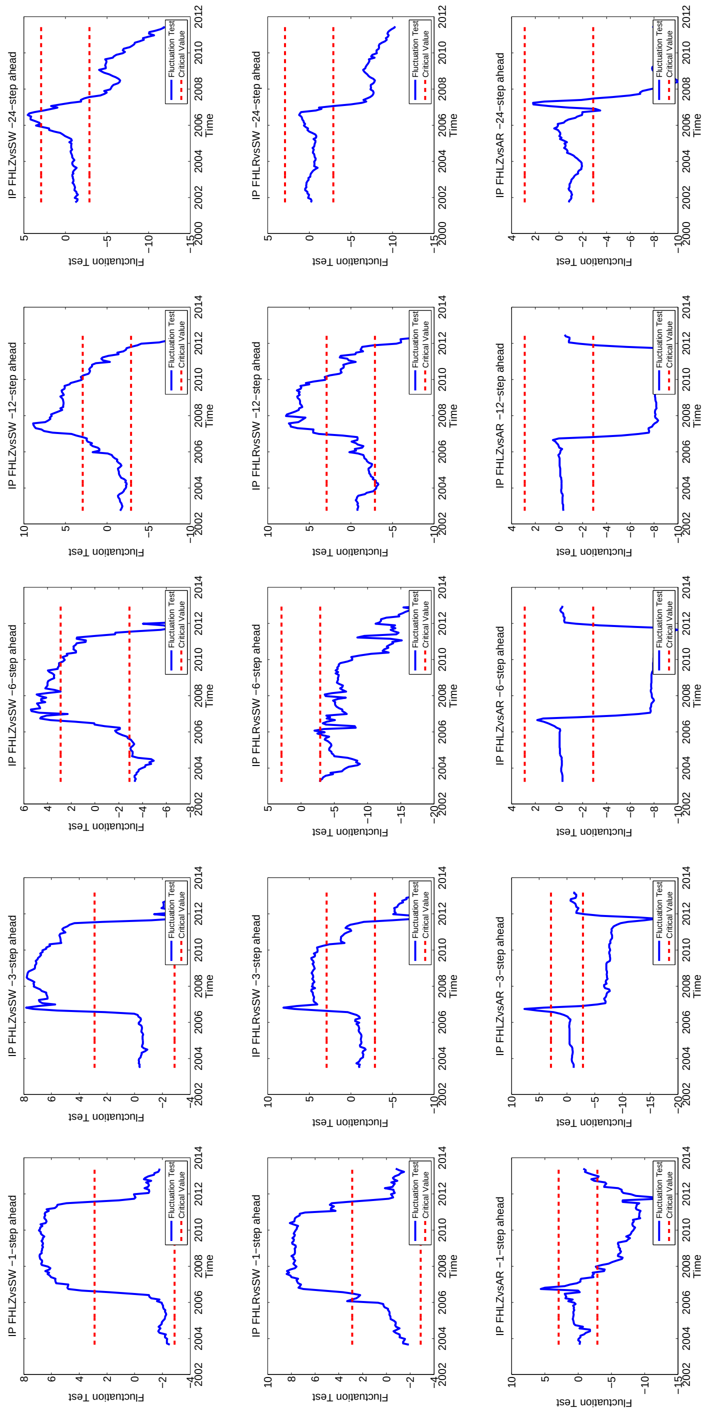


Figure A.5: Cumulated forecast error Main European countries (CPI)

Figure A.6: Fluctuation test (Euro Area IP)



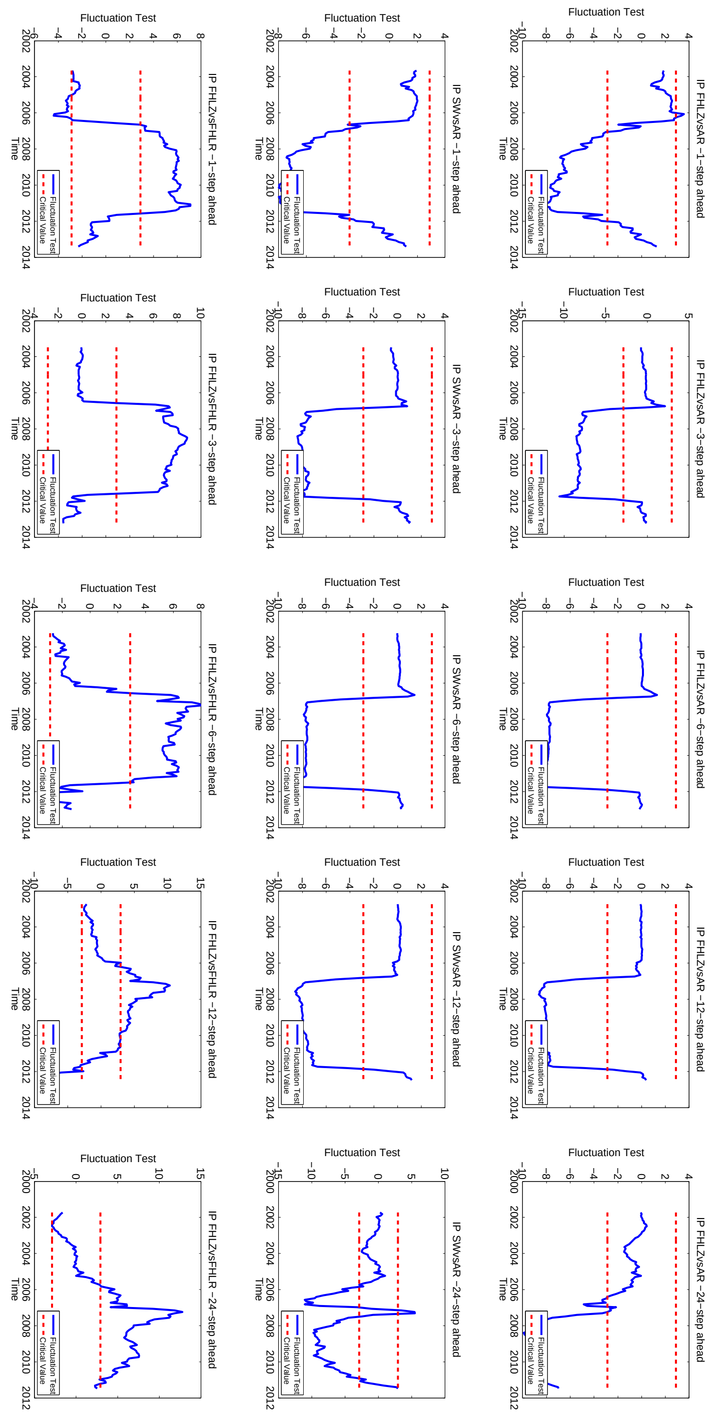
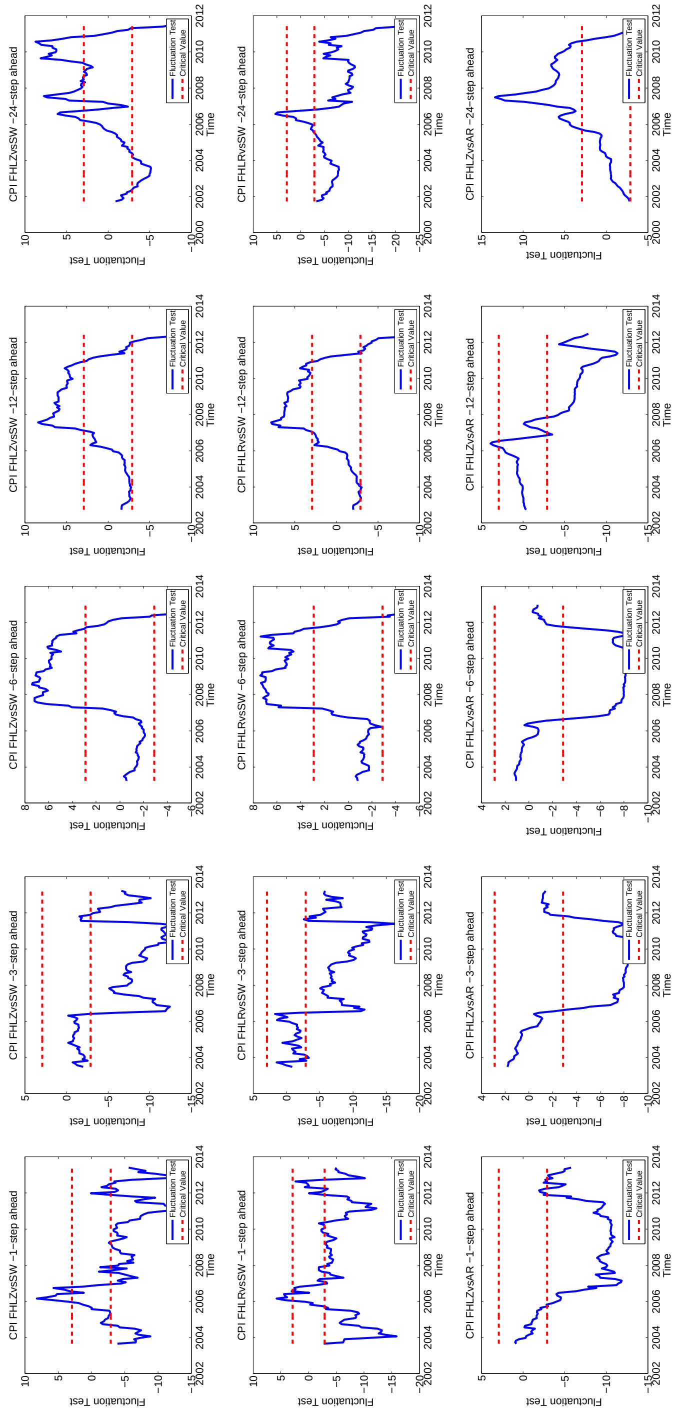


Figure A.7: Fluctuation test (Euro Area IP)

Figure A.8: Fluctuation test (Euro Area CPI)



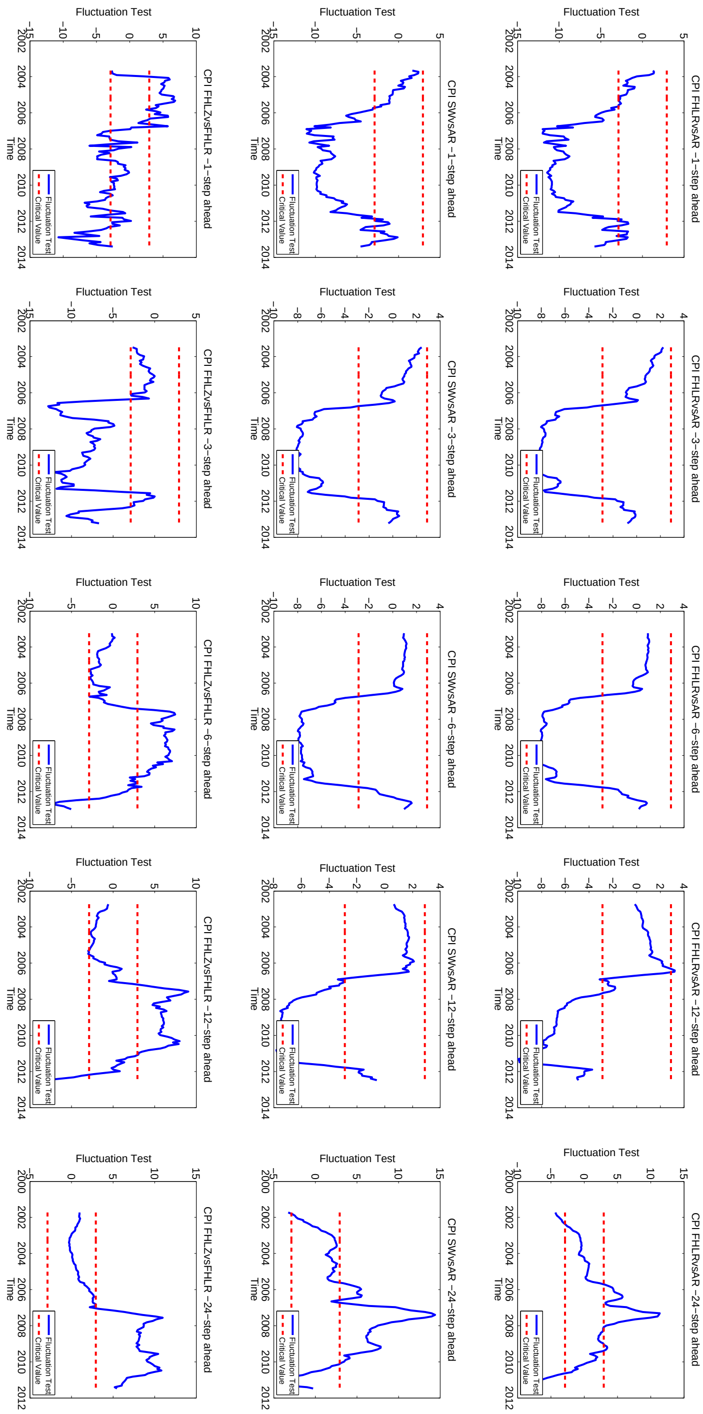
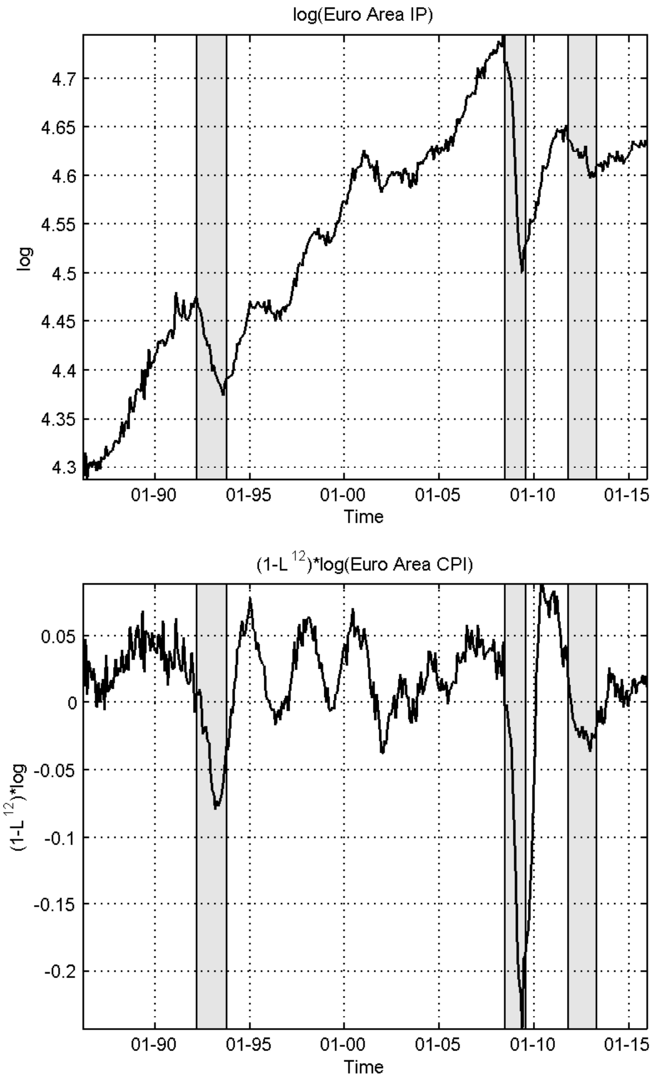


Figure A.9: Fluctuation test (Euro Area CPI)

Figure A.10: Target variables, log(IP) and $(1 - L^{12})\log(\text{CPI})$

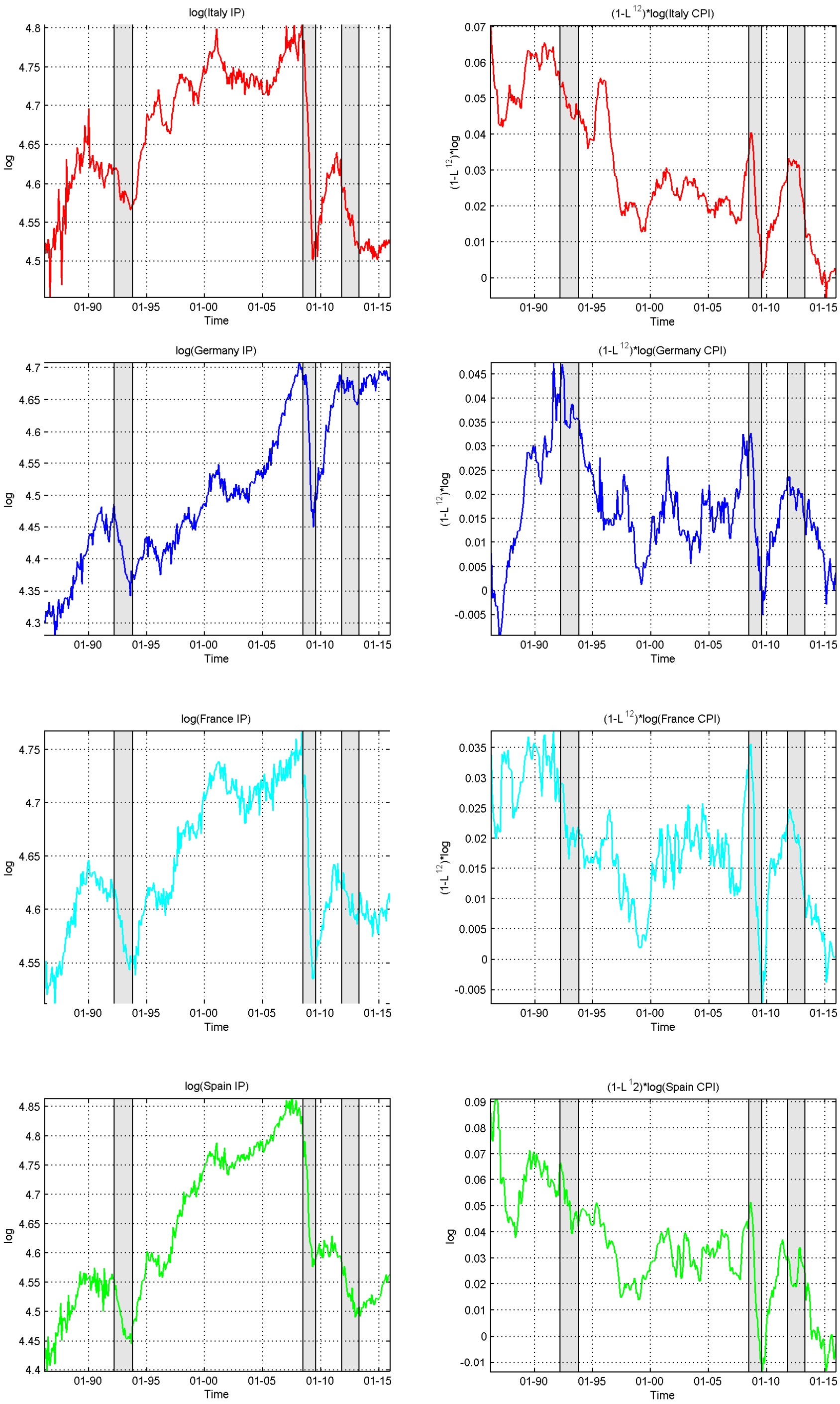
a



[a] Shaded areas indicate CEPR recession dates

Figure A.11: Main European countries target variables

a



[a]Shaded areas indicate CEPR recession dates

A.2 Appendice II - Chapter 3

A.2.1 Dataset description

The involved trasformation has the same trasformation code as in A.1.1

Table A.18: List of the series

	Name	Long Desc.	Tcode	Deseas	CatCode
1	TSPC1A00	Written Premiums - Class I - Life - Annual premiums	5	1	1
2	TSPC1U00	Written Premiums - Class I - Life - Single premiums	5	1	1
3	TSPC3A00	Written Premiums - Class III - Life - Annual premiums	5	1	1
4	TSPC3U00	Written Premiums - Class III - Life - Single premiums	5	1	1
5	TSPC5A00	Written Premiums - Class V - Life - Annual premiums	5	1	1
6	TSPC5U00	Written Premiums - Class I - Life - Single premiums	5	1	1
7	TSPCAA00	Written Premiums - Other - Life - Annual premiums	5	1	1
8	TSPDIM00	Written premiums - Accident and Sickness - Non Life	5	1	2
9	TSPDAR00	Written premiums - Other - Non Life	5	1	2
10	TSPDTR00	Written premiums - Transport - Non Life	5	1	2
11	TSPDRCG0	Written premiums - General Liability - Non Life	5	1	2
12	TSPDRAD0	Written premiums - Other Motor risks - Non Life	5	1	2
13	TSPDRCA0	Written premiums - Motor and vessels Liability - Non Life	5	1	2
14	TSNP1A00	New Business - Class I - Life - Annual premiums	5	1	3
15	TSNP1U00	New Business - Class I - Life - Single premiums	5	1	3
16	TSNP3A00	New Business - Class III - Life - Annual premiums	5	1	3
17	TSNP3U00	New Business - Class III - Life - Single premiums	5	1	3
18	TSNP5T00	New Business - Class V - Life - Total	5	1	3
19	TSNPTA00	New Business - Total - Life - Annual premiums	5	1	3
20	TSNPTU00	New Business - Total - Life - Single premiums	5	1	3

Based on IVASS quarterly statistics. Deseasonalisation and suitable transformation are involved in order to get stationarity.

Table A.19: List of the series

	Name	Long Description	Tcode	Deseas	CatCode
21	ITM1....A	IT MONEY SUPPLY: M1 - ITALIAN CONTRIBUTION TO THE EURO AREA CURN	6	1	4
22	ITM2....A	IT MONEY SUPPLY: M2 - ITALIAN CONTRIBUTION TO THE EURO AREA CURN	6	1	4
23	ITM3....A	IT MONEY SUPPLY: M3 - ITALIAN CONTRIBUTION TO THE EURO AREA CURN	6	1	4
24	ITMLRT14Q	IT HARMONISED UNEMPLOYMENT: RATE, ALL PERSONS (AGES 15-24) SADJ	2	0	5
25	ITMLRT15Q	IT HARMONISED UNEMPLMT.: RATE,ALL PERSONS(AGES 25 AND OVER) SADJ	2	0	5
26	ITMLRF16Q	IT HARMONISED UNEMPLOYMENT: RATE, FEMMES (ALL AGES) SADJ	2	0	5
27	ITMLRM16Q	IT HARMONISED UNEMPLOYMENT: RATE, HOMMES (ALL AGES) SADJ	2	0	5
28	ITESPPINF	IT PPI: MIG - NON-DURABLE CONSUMER GOODS NADJ	7	0	6
29	ITESPPIEF	IT PPI: MIG - ENERGY NADJ	7	0	6
30	ITCP7500F	IT CPI (1975=100) NADJ	7	0	6
31	ITRAWPRCF	IT RAW MATERIALS PRICE INDEX NADJ	7	0	6
32	ITESSFUB	IT HARMONISED GOVERNMENT 10-YEAR BOND YIELD	2	0	9
33	ITDISCRT	IT DISCOUNT RATE / SHORT TERM EURO REPO RATE (MTH.AVG.)	2	0	9
34	INDGSIT	ITALY-DS Inds Gds & Svs - PRICE INDEX	5	0	9
35	FINANIT	ITALY-DS Financials - PRICE INDEX	5	0	9
36	CNSMGIT	ITALY-DS Consumer Gds - PRICE INDEX	5	0	9
37	NLINSIT	ITALY-DS Nonlife Insur - PRICE INDEX	5	0	9
38	NLINSEM	EMU-DS Nonlife Insur - PRICE INDEX	5	0	9
39	PCINSIT	ITALY-DS Prop/Cas Insur - PRICE INDEX	5	0	9
40	PCINSEM	EMU-DS Prop/Cas Insur - PRICE INDEX	5	0	9
41	LFINSIT	ITALY-DS Life Insurance - PRICE INDEX	5	0	9
42	LFINSEM	EMU-DS Life Insurance - PRICE INDEX	5	0	9
43	ITCONPRCF	IT CPI INCLUDING TOBACCO (NIC) NADJ	2	0	6
44	ITCNFCONQ	IT HOUSEHOLD CONFIDENCE INDEX SADJ	2	0	8
45	ITECONOPR	IT BUS.SVY.: ECONOMY IN NEXT 3MOS- FAVOURABLES PLUS STABLES NADJ	2	0	8
46	ITCSSVPCR	IT CONSUMER SURVEY: SAVINGS - PRESENT CONVENIENCE (BALANCE) NADJ	2	0	8
47	ITCSSVFOR	IT CONSUMER SURVEY: FUTURE SAVINGS OPPORTUNITY (BALANCE) NADJ	2	0	8
48	ITYTHAR%R	IT ACTIVITY RATE: 15 TO 24 YEAR OLDS NADJ	2	1	5
49	ITEMPRT%R	IT EMPLOYMENT RATE NADJ	2	1	5
50	ITYTHEM%R	IT EMPLOYMENT RATE: MALE - 15 TO 24 YEAR OLDS NADJ	2	1	5
51	ITJBSSTHP	IT JOB SEEKERS - SOUTHERN ITALY VOLN	5	1	5
52	ITJBSCTRP	IT JOB SEEKERS - CENTRAL ITALY VOLN	5	1	5
53	ITJBSNRDP	IT JOB SEEKERS - NORTHERN ITALY VOLN	5	1	5
54	ITUNRSD%R	IT UNEMPLOYMENT RATE - SOUTHERN ITALY NADJ	2	1	5
55	ITUNRCT%R	IT UNEMPLOYMENT RATE - CENTRAL ITALY NADJ	2	1	5
56	ITUNRND%R	IT UNEMPLOYMENT RATE - NORTHERN ITALY NADJ	2	1	5
57	ITHHFECSR	IT HOUSEHOLD CONFIDENCE SURVEY: FUTURE FINANCIAL POSITION NADJ	2	0	8
58	ITCSENBAR	IT CONSUMER SURVEY: GEN.ECON.SITUATION EXPECTATIONS(BALANCE)	2	0	8
59	ITCSEYBAR	IT CONSUMER SURVEY: GENERAL ECONOMIC SITUATION (BALANCE) NADJ	2	0	8
60	ITCSPYBLR	IT CONSUMER SURVEY: PRICES (CPY) - BALANCE NADJ	2	0	8
61	ITCSPNBLR	IT CONSUMER SURVEY: PRICES IN NEXT 12 MTHS. - BALANCE NADJ	2	0	8
62	ITOCFILTR	IT LONG-TERM INTEREST RATE ON GOVERNMENT BONDS (AR) SADJ	2	0	9

Source: Datastream.

A.2.2 Glossary of italian insurance terms

[1]

Main definitions

- **Direct business**: premiums collected by a company net of those premiums coming from the active reinsurance business the company may make with other companies.
- **Gross written premiums**: they include all sums matured during pursuit of insurance business for insurance contracts, regardless of the fact that such sums have been collected or that they partially or totally refer to subsequent business; the amounts for the relative taxes and the contributions paid for compensations are excluded. They also include:

  a) premiums yet to be written, in case such premiums can be calculated only at year end;

  b) single premiums and sums destined to the purchase of a periodic annuity;

  c) in life insurance, single premiums coming from the provisions for participation in profits and rebates, to the extent that they must be considered as premiums on the basis of contracts;

  d) surcharges for premium splitting and complementary benefits of insureds aimed at covering the company's expenses;

  e) the companyâĂŹs premium shares acquired for co-insurance;

  f) reinsurance premiums coming from ceding and retroceding insurance companies.

- **New business**: premiums coming from the act of writing new policies
- **Non-EEA company offices**: branch offices of non-EU companies operating in Italy in Freedom of Establishment (FOE) or Freedom of Services (FOS).
- **Annual premiums**: sums matured for those contracts establishing that the contracting party must pay a generally constant amount at preset deadlines.
- **Single premiums**: sums matured for those contracts establishing that the contracting party must pay the premium in a single instalment at contract stipulation.

[1] For the huge classification and description results in this section I'm grateful to Angelo Silvaroli for his contribution to the digitalization (for the data spanning the period 1983:1988) and classification of the data during his intership in ANIA, Silvia Salati, ANIA Statistical Department, for the classification of Non-Life Classes and the Glossary and to Annalaura Grasso, ANIA International Relationship Department, for her support in translation. Every error is my responsiblity.

- **Recurring premiums**: sums matured for those contracts establishing that the contracting party must issue a series of 'single' payments generally established at contract stipulation and made at preset deadlines.

Life sector The Code of Private Insurance classifies Life Insurance in six classes:

- **Class I**: assurance on the length of human life classified according to the form of contract (caso di morte, caso vita, miste);

- **Class II**: marriage assurance, birth assurance (never activated);

- **Class III**: assurance referred to in classes I and II, whose main benefits are directly linked to the value of units of a UCITS (undertakings for collective investment in transferable securities) or the value of the assets in an internal fund (the so called *unit-linked policies* or else to an index or other reference values (price index, stock index, etc.)(the so called *index-linked*). As the monetary value of benefits depends on the value of the fund of the value of the index, the beneficiary bears a financial risk;

- **Class IV**: health insurance and insurance against the risk of dependency that are covered by permanent health insurance contracts not subject to cancellation, against the risk of serious disability resulting from accident or sickness or longevity;

- **Class V**: capital redemption operations, meaning operations mainly aimed at managing sums of money entrusted to the insurance company as manager. These are mainly financial operations as there is no connection to events linked to the length of human life, even though they include some aspects having insurance nature, such as financial risk cover (with the guarantee of a minimum yearly interest rate and the consolidation of the financial results).

- **Class VI**: management of group pension funds that effect payments on death or survival or in the event of discontinuance or curtailment of activity.

Non-Life sector The Code of Private Insurance classifies Non-Life Insurance in 18 classes, duly re-classified in 9 macro classes (see table **??** for details):

- **Accident**: insurance contracts aimed at covering possible ecomnomic damages arising from an accident, understood as a general reduction in the incapability of producing;

- **Sickness**: insurance contracts aimed at guaranteeing pecuniary benefits during hospitalisation in order to cover any residual loss or, in addition, also to cover expenses for treatment in a private hospital or nursing home;

- **Transport**: insurance contracts covering any damage undergone by sea, lake and river and canal vessels, railway rolling stock and aircrafts, any damage undergone by goods in transit or baggage, regardless of the type of the mean of transport and any liability deriving from the use of the aforesaid vessels, railway rolling stock and aircrafts, including carrier liability;
- **Credit**: insurance contracts relative to compensation for damage undergone by the creditor the debtor's payment unfulfillment;
- **Suretyship**: insurance contracts having the same juridical and economic function of a bond in money or, or of a bank guarantee that a subject may be obliged to stipulate in favour of the beneficiary in order to guarantee future obligations or for unfulfillment or as compensation for damages.
- **General Liability**: insurance contracts thanks to which the insurer is obliged to cover the insured for the risk that his/her capital is reduced as an economic consequence of the claims for compensations filed by third parties for the insuredâĂŹs alleged liability for facts or acts committed or by those subjects the insured must be held liable for in pursuing a specific activity described in the policy.
- **Motor and vessels Liability**: any liability coming from the use of land vehicles and sea, lake and river and canal vessels including the carrier's liability;
- **Other motor risks**: insurance contracts relative to the Motor class referring to risks different from those covered by MTPL (fire, theft, etc)
- **Other Non-Life classes**: other damages to property, pecuniary losses, legal expenses, assistance, fire and other natural forces.

Distribution channels

- **Insurance agencies**: insurance agents or subjects bearing the mandate of promoting contract stipulations on behalf of an insurance company (see art. 1742 Civil Code); these are independent collaborators of the main company and are different from the subject appointed by the company to manage the internal agencies.
- **Internal agencies for direct sale on the premises**: they are part of the company's internal organisation (a specific class or a branch office). This type of agent takes care of the agency and therefore he/she is not an autonomous collaborator but rather an instructor bound to the insurance company by means of a work relation based on the management of the agency (see artt. 2203-2208 Civil Code).
- **Bank counters**: intermediaries enrolled in section d) of the Single Register of Intermediaries that, besides banks, include post offices and can exclusively distribute

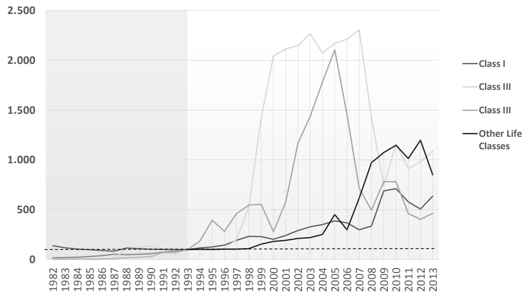
insurance products that contain preset, clauses and guarantees which cannot be modified by the subject entitled for distribution .

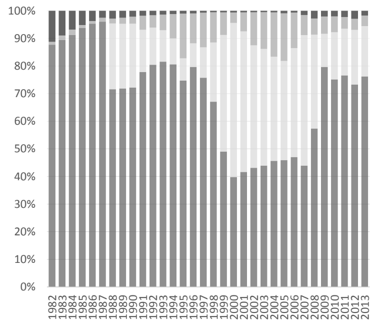
- **Financial Advisers**: usually employees or collaborators of brokerage firms, they are not directly part of the institutional insurance industry (their register is not managed by IVASS but by Consob, the Supervision Authority for financial markets and listed companies).
- **Brokers**: intermediaries operating upon mandate of the insured with no representation powers entrusted by insurance and reinsurance companies. The insurance mefiation activity must be carried out by a subject enrolled in section b) del Registro Unico elettronico degli Intermediari assicurativi e Riassicurativi

A.2.3 Figures

Annual gross written premiums - 1982-2013

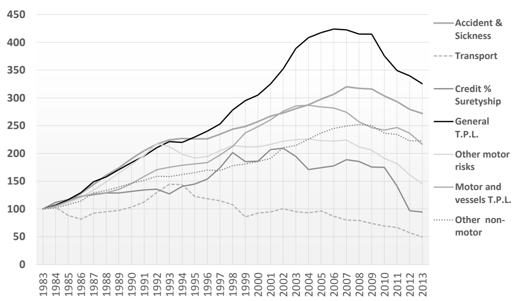
Figure A.12: Panel A - Life sector (direct business, national and Non-EEA company offices)

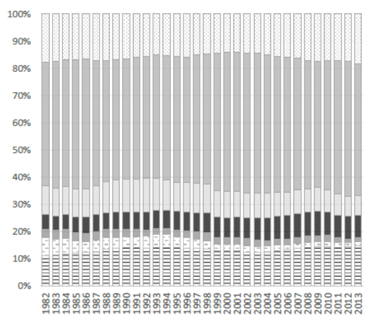




Left side: Aggregated annual premiums in real terms; 100=1993; Right side: annual decomposition by Life Classes

Figure A.13: Panel B - Non-Life sector (direct business, national and Non-EEA company offices)





Left side: Aggregated annual premiums in real terms; 100=1982; Right side: annual decomposition by Non-life Classes

Table A.20: Non-life sector

Accident and Sickness	From 1925 to 1939 this sector included also General and Motor Liability. Later has moved in a new class
Motor and vessels Liability	From 1925 to 1939 included in Accidents and Sickness. From 1940 to 1954 motor insurance was still not compulsory. From 1954 motor liability insurance become compulsory. This class comprises also vessels policies
Other motor risks	Start to be available since 1955. In 1998 was renamed in Land Vehicles.
Transport	This Class exists since 1925; from 1932 it includes Aviation risks, from 1998: Aircraft, Railway Rolling Stock, Ships, Goods in transit e Aircraft T.P.L.
General T.P.L.	From 1940 General T.P.L is excluded from Accident and Sickness but still include some Motor T.P.L. until 1955.
Credit and Suretyship	data for Suretyship are available starting from 1936
Other Non-Life classes	Exists since 1925. Several residual classes have been added on time: Dal 1974 si sono inclusi i rami Perdite Pecuniarie e Tutela Giudiziaria (ora Tutela Legale). From 1998 to 2010 includes: Other Damage to property, Miscellaneous Financial Loss, Legal Expenses e Assistance.

[H]

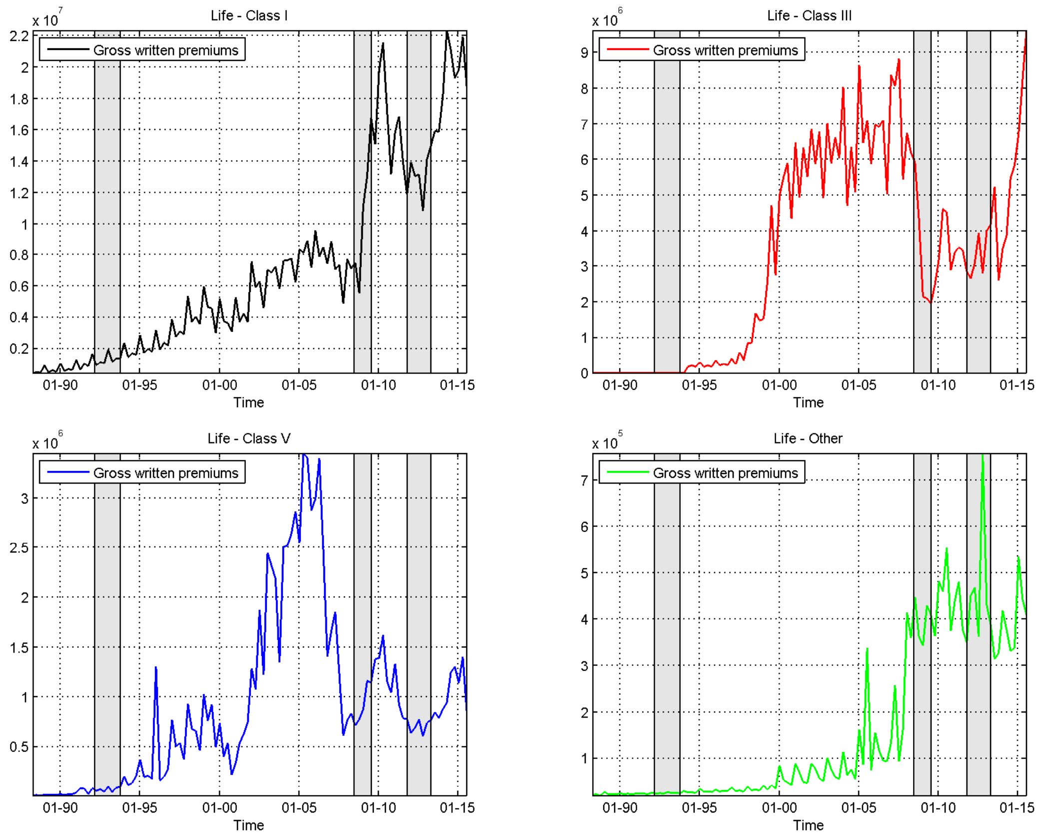
ANIA Statistic Department re-classification based on ANIA and IVASS statistics, IVASS reporting templates and Italian Code of Private Insurance classification

Table A.21: Life Sector

Class I	From 1982 to 1987 includes policies named *Ordinarie*, *Popolari* and *Collettive*. From 1988 to 1998 *Ageduabili*, *Indicizzate*, *Rivalutabili* and *Assicurazioni Collettive*. From 1999 Life Class 1.
Class III	From 1982 to 1987 the Class doesn't exist. From 1988 to 1998 inlcudes *polizze Connesse a fondi di investimento* and Others. From 1999 Life Class III.
CLass V	From 1982 to 1998 include the old class *Ramo Capitalizzazione*. From 1999 Life Class V.
Others Life Classes	From 1982 to 1987 includes the old class *polizze Popolari*. From 1988 to 1998 includes *Assicurazioni Popolari* and *Assicurazioni Complementari*. From 1999 Life Class IV and Life Class VI.

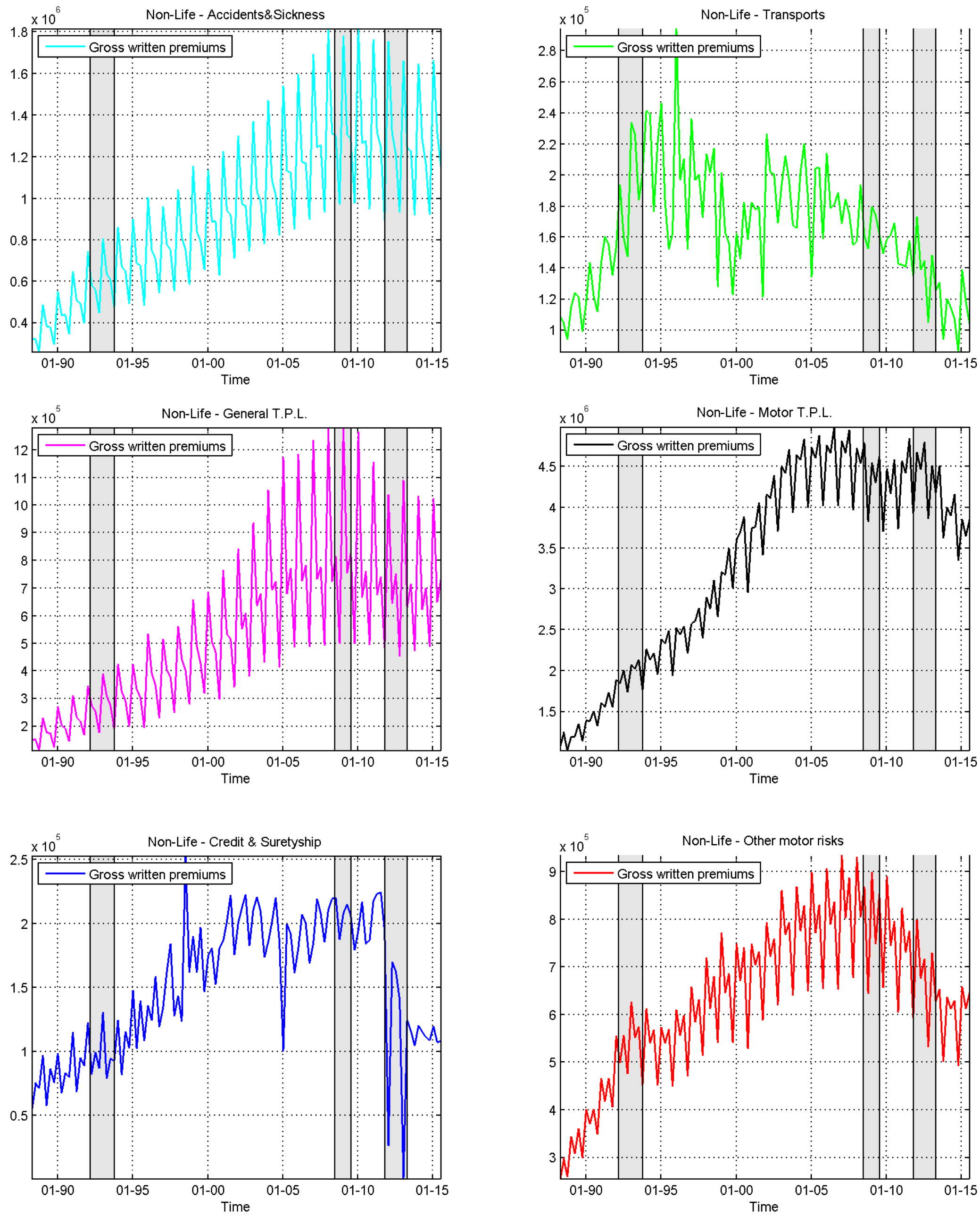
Author re-classification based on quarterly IVASS statistics and reporting templates and Italian Code of Private Insurance classification Revision of quarterly reporting templates refers to the following official IVASS communications: n.94, 24 June 1988; n. 253, 14 July 1995; n. 356, 1 March 1999.

Figure A.14: Life quarterly premiums volumes



Based on quarterly IVASS statistics; volumes; raw data, expressed in thousand euros. Direct business, national and Non-EEA company offices. Shaded areas indicate CEPR recession dates

Figure A.15: Non-life quarterly premiums - volumes



Based on quarterly IVASS statistics; volumes; raw data; expressed in thousand euros. Direct business, national and Non-EEA company offices.

Bibliography

Altissimo, Filippo et al. (2010). "New Eurocoin: Tracking economic growth in real time". In: *The review of economics and statistics* 92.4, pp. 1024–1034.

Bai, Jushan and Serena Ng (2002). "Determining the number of factors in approximate factor models". In: *Econometrica* 70.1, pp. 191–221.

Beck, Thorsten and Ian Webb (2003). "Economic, demographic, and institutional determinants of life insurance consumption across countries". In: *The World Bank Economic Review* 17.1, pp. 51–88.

Boivin, Jean and Serena Ng (2005). *Understanding and comparing factor-based forecasts.* Tech. rep. National Bureau of Economic Research.

— (2006). "Are more data always better for factor analysis?" In: *Journal of Econometrics* 132.1, pp. 169–194.

Born, Patricia H et al. (2014). "Cash Flow Risk Management in the Insurance Industry A Dynamic Factor Modeling Approach". In:

Chamberlain, Gary and Michael Rothschild (1982). *Arbitrage, factor structure, and mean-variance analysis on large asset markets.*

Cristadoro, Riccardo et al. (2005). "A core inflation indicator for the euro area". In: *Journal of Money, Credit and Banking*, pp. 539–560.

D Agostino, Antonello and Domenico Giannone (2012). "Comparing Alternative Predictors Based on Large-Panel Factor Models". In: *Oxford bulletin of economics and statistics* 74.2, pp. 306–326.

Forni, M. et al. (2016). *Dynamic Factor model with infinite dimensional factor space: forecasting.* Tech. rep. London: Centre for Economic Policy Research. URL: `http://www.cepr.org/active/publications/discussion_papers/dp.php?dpno=11161`.

Forni, Mario et al. (2000). "The generalized dynamic-factor model: Identification and estimation". In: *Review of Economics and statistics* 82.4, pp. 540–554.

— (2005). "The generalized dynamic factor model: One-sided estimation and forecasting". In: *Journal of the American Statistical Association.*

Forni, Mario et al. (2015). "Dynamic factor models with infinite-dimensional factor spaces: representation". In: *Journal of econometrics* 185.2, pp. 359–371.

— (2016). "Dynamic Factor model with infinite dimensional factor space: estimation". In: *Mimeo.*

Forni, Mario et al. (2011). "One-sided representations of generalized dynamic factor models". In: *ECARES Working Papers.*

Forni, Mario, Marco Lippi, and Lucrezia Reichlin (2004). *The generalized dynamic factor model, one-sided estimation and forecasting. 2003.*

Geweke, John (1977). *The dynamic factor analysis of economic time series.* Vol. 185, pp. 365–383.

Hallin, Marc, Roman Liska, et al. (2007). "The generalized dynamic factor model: determining the number of factors". In: *Journal of the American Statistical Association* 102.478, pp. 603–617.

Luciani, Matteo (2014a). "Forecasting with Approximate Dynamic Factor Models: the role of non-pervasive shocks". In: *International journal of forecasting* 30.1, pp. 20–29.

— (2014b). "Large-dimensional dynamic factor models in real-time: A survey". In: *Available at SSRN 2511872.*

Millo, Giovanni (2015). "The income elasticity of nonlife insurance: a reassessment". In: *Journal of Risk and Insurance.*

Onatski, Alexei (2012). "Asymptotics of the principal components estimator of large factor models with weakly influential factors". In: *Journal of Econometrics* 168.2, pp. 244–258.

Outreville, J François (2011). “The relationship between insurance growth and economic development: 80 empirical papers for a review of the literature”. In:

Petrova, Yana (2014). “The Relationship between Insurance Market Activity and Economic Growth”. In:

Reijer, Ard (2005). “Forecasting Dutch GDP using large scale factor models”. In:

Sargent, Thomas J, Christopher A Sims, et al. (1977). “Business cycle modeling without pretending to have too much a priori economic theory”. In: *New methods in business cycle research* 1, pp. 145–168.

Schumacher, Christian (2007). “Forecasting German GDP using alternative factor models based on large datasets”. In: *Journal of Forecasting* 26.4, pp. 271–302.

Stock, James H and Mark W Watson (2002a). “Forecasting using principal components from a large number of predictors”. In: *Journal of the American statistical association* 97.460, pp. 1167–1179.

— (2002b). “Macroeconomic forecasting using diffusion indexes”. In: *Journal of Business & Economic Statistics* 20.2, pp. 147–162.

Ward, Damian and Ralf Zurbruegg (2000). “Does insurance promote economic growth? Evidence from OECD countries”. In: *Journal of Risk and Insurance*, pp. 489–506.